

# DMS 2014



**Proceedings of the Twentieth  
International Conference on  
Distributed Multimedia Systems**

**Pittsburgh  
August 27-29, 2014**

**PROCEEDINGS**

**DMS 2014**

**The 20<sup>th</sup> International Conference on  
Distributed Multimedia Systems**

**Sponsored by**

Knowledge Systems Institute Graduate School, USA

**Technical Program**

**August 27 – 29, 2014**

**Wyndham Pittsburgh University Center, Pittsburgh, USA**

**Organized by**

Knowledge Systems Institute Graduate School, USA



Copyright © 2014 by Knowledge Systems Institute Graduate School

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written consent of the publisher.

ISBN: 1-891706-36-5

ISSN: 2326-3261 (print)

2326-3318 (online)

Additional copies can be order from:

Knowledge Systems Institute Graduate School

3420 Main Street

Skokie, IL 60076 USA

Tel: +1-847-679-3135

Fax: +1-847-679-3166

Email: [dms@ksiresearch.org](mailto:dms@ksiresearch.org)

Web: <http://www.ksi.edu>

Proceedings preparation, editing and printing are sponsored by Knowledge Systems Institute Graduate School, USA.

Printed by Knowledge Systems Institute Graduate School, USA.

# FOREWORD

It is a pleasure to welcome everyone to the 20th International Conference on Distributed Multimedia Systems (DMS) that this year takes place in Pittsburgh, PA, USA. At this international conference meeting we will consequently be able to celebrate the twentieth conference in this series and hopefully we will also be able to look forward to a number of further successful events in the years to come. DMS has established itself as an international forum on distributed multimedia systems with participants from academia, industry, and government agencies. Discussions on exchange of ideas, research results and experiences in the field will hopefully have a positive influence on the quality of the meeting. The DMS community has become an important source of ideas and innovations within the area and contributions to DMS have been supplied from researcher from many countries all over the world. On behalf of the Program Committee, Co-Chairs and the entire Program Committee, I would like to extend to you really warm welcome to DMS 2014.

We have received 33 submissions from about 10 countries this year. Proceedings will be published as a special issue of the journal of Visual Languages and Computing. The review process has for this reason been carried out in a somewhat different way than for traditional conferences. Eventually, three review rounds will be carried out. As for now just two have been finished and the third is yet to come after the conference. The first review round was carried out mainly in May following traditional guidelines. The outcome of this round gave as a result acceptance of 27 papers; that is an acceptance rate of about 81%. Among the accepted papers 12 were accepted as full papers and 17 as short papers. After the first review round the authors were asked to revise their papers and turn them in again for the second round. Furthermore, the revised versions of the papers will be available at the conference in a preliminary proceedings. The second review round took place in July and again the authors are asked to revise the papers in accordance with the reports from the second round. However, the second revision has to be finished on September 20. After which the third review round will start and carried out jointly by the JVLC EIC, the general co-chairs, PC chair and co- chairs basically for quality control.

The high quality of the DMS 2014 technical program would not have been possible without the tireless efforts and hard work by many individuals. First of all, I would like to express my sincere appreciation to all the authors whose technical contributions have made the final technical program possible. I am very grateful to all the Program Committee members whose expertise and dedication made my responsibility that much easier. My gratitude also goes to the keynote speaker and panelists who graciously agreed to share their insight on important research issues, to the conference organizing committee members for their superb work, and to the external reviewers for their contributions.

Finally, I would like to send special thanks to my old friend Dr. S.-K. Chang who, as Chair of the Steering Committee has been of great support and whose extraordinary experiences again has been of the greatest value. Furthermore, I must also acknowledge the important contributions made by the KSI staff members. Their timely and dependable support and assistance throughout the entire process have been truly remarkable. It has been a great pleasure to work with all of them. Finally, I hope you will enjoy both the scientific part of DMS 2014 by exciting exchange of ideas related to the various topics of multimedia and the visit to Pittsburgh, which is a very nice and interesting city with a fascinating history.

Erland Jungert,  
DMS 2014 Program Chair





# **DMS 2014**

## **The 20<sup>th</sup> International Conference on Distributed Multimedia Systems**

**August 27 – 29, 2014**

**Wyndham Pittsburgh University Center, Pittsburgh, USA**

### **Conference Organization**

#### **DMS'14 Conference Co-Chairs**

Paolo Nesi, University of Florence, Italy; conference co-chair

Kia Ng, University of Leeds, UK; conference co-chair

#### **DMS'14 Program Co-Chairs**

Erland Jungert, Linkoping University, Sweden; Program Chair

Wei Lu, Beijing Jiao Tung University, China; Program Co-Chair

#### **DMS'14 Steering Committee Chair**

Shi-Kuo Chang, University of Pittsburgh, USA

#### **DMS'14 Program Committee**

Arvind K. Bansal, Kent State University, USA

Andrew Blake, University of Brighton, UK

Augusto Celentano, Universita Ca Foscari di Venezia, Italy

Ing-Ray Chen, Virginia Tech (VPI&SU), USA

Shu-Ching Chen, Florida International University, USA

William Cheng-Chung Chu, Tunghai University, Taiwan

F. Colace, University of Salerno, Italy

Gennaro Costagliola, Univ of Salerno, Italy

Alfredo Cuzzocrea, ICAR-CNR and University of Calabria, Italy

Andrea De Lucia, Univ. of Salerno, Italy

Tiansi Dong, Bonn-Aachen International Center for Information Technology, Germany

David H. C. Du, Univ. of Minnesota, USA



Larbi Esmahi, National Research Council of Canada, Canada  
 Daniela Fogli, Università degli Studi di Brescia, Italy  
 Kaori Fujinami, Tokyo University of Agriculture and Technology, Japan  
 David Fuschi, Brunel University, UK  
 Ombretta Gaggi, Univ. of Padova, Italy  
 Nikolaos Gkalelis, Informatics & Telematic Institute, Greece  
 Angela Guercio, Kent State University, USA  
 Niklas Hallberg, FOI, Sweden  
 Carlos A. Iglesias, Intelligent Systems Group, Spain  
 Erland Jungert, Linköping University, Sweden  
 Yau-Hwang Kuo, National Cheng Kung University, Taiwan  
 Fuhua Lin, Athabasca University, Canada  
 Alan Liu, National Chung Cheng University, Taiwan  
 Jonathan Liu, University of Florida, USA  
 Sethuraman Panchanathan, Arizona State Univ., USA  
 Antonio Piccinno, Univ. of Bari, Italy  
 Giuseppe Polese, University of Salerno, Italy  
 Buntarou Shizuki, University of Tsukuba, Japan  
 Peter Stanchev, Kettering University, USA  
 Mahbubur Syed, Minnesota State University, USA  
 Christian Timmerer, Klagenfurt University, Austria  
 Genny Tortora, University of Salerno, Italy  
 Weiwei Xing, Beijing Jiao Tong University, China  
 Atsuo Yoshitaka, JAIST, Japan  
 Ing Tomas Zeman, Czech Technical University, Czech Republic  
 Kang Zhang, University of Texas at Dallas, USA  
 Roberto Zicari, Goethe University Frankfurt, Germany

### **DET'14 Workshop Co-Chairs**

Angela Guercio, Kent State University, USA; workshop co-chair  
 Tim Arndt, Cleveland State University, USA; workshop co-chair

### **DET'14 Program Chair**

Paolo Maresca, University Federico II, Napoli, Italy; Program Chair

### **DET'14 Program Committee**

Tim Arndt, Cleveland State University, USA  
 Maiga Chang, Athabasca University, Canada  
 Yuan-Sun Chu, National Chung Cheng University, Taiwan  
 Mauro Coccoli, University of Genova, Italy  
 Luigi Colazzo, University of Trento, Italy

Massimo De Santo, Università degli Studi di Salerno, Italy  
Rita Francese, University of Salerno, Italy  
Angelo Gargantini, University of Bergamo, Italy  
Angela Guercio, Kent State University, USA  
Pedro Isaias, Open University, Portugal  
Lian Li, HeFei University of Technology, China  
Hong Lin, University of Houston-Downtown, USA  
Paolo Maresca, University Federico II, Napoli, Italy  
Andrea Molinari, University of Trento, Trento, Italy  
Ignazio Passero, University of Salerno, Italy  
Elvinia Riccobene, University of Milano, Italy  
Michele Risi, University of Salerno, Italy  
Teresa Roselli, University of Bari, Italy  
Veronica Rossano, University of Bari, Italy  
Giuseppe Scanniello, University of Salerno, Italy  
Lidia Stanganelli, University of Napoli Federico II, Italy  
Mahbubur Syed, Minnesota State University, USA  
Kazuo Yana, Hosei University, Japan

### **VLC'14 Workshop Co-Chairs**

A. J. Delaney, University of Brighton, UK; workshop co-chair  
Jennifer Leopold, Missouri University of Science & Technology, USA; workshop co-chair

### **VLC'14 Program Chair**

Franklyn A Turbak, Wellesley College, USA; Program chair

### **VLC'14 Program Committee**

Danilo Avola, University of Rome, Italy  
Paolo Bottoni, Università Sapienza, Italy  
Paolo Buono, University of Bari, Italy  
Alfonso F. Cardenas, University of California, USA  
Kendra Cooper, University of Texas at Dallas, USA  
Gennaro Costagliola, University of Salerno, Italy  
Philip Cox, Dalhousie University, Canada  
Sergiu Dascalu, University of Nevada, USA  
Aidan Delaney, University of Brighton, UK  
Vincenzo Deufemia, University of Salerno, Italy  
Filomena Ferrucci, University of Salerno, Italy  
Andrew Fish, University of Brighton, UK  
Paul Fishwick, University of Florida, USA  
Manuel J. Fonseca, INESC-ID, Portugal



Vittorio Fuccella, University of Salerno, Italy  
Joaquim A Jorge, Instituto Suoerior Tecnico, Portugal  
Levent Burak Kara, Carnegie Mellon University, USA  
Jun Kong, North Dokota State University, USA  
Robert Laurini, University of Lyon, France  
Jennifer Leopold, Missouri University of Science & Technology, USA  
Sean McDirmid, Microsoft Research Asia, China  
Joseph J. Pfeiffer, Jr., New Mexico State University, USA  
Peter Rodgers, University of Kent, UK  
Monica Sebillo, University of Salerno, Italy  
Gem Stapleton, University of Brighton, UK  
Giuliana Vitiello, University of Salerno, Italy  
Kang Zhang, University of Texas at Dallas, USA

### **DMS/DET/VLC'14 Publicity Chair**

Lidia Stanganelli, University of Napoli Federico II, Italy; Publicity Chair

## **Keynote**

Future of Shapes: How does information and fabrication technologies change the way we design?

Levent Burak Kara  
Carnegie Mellon University, USA

**About the Speaker:** Levent Burak Kara, Ph.D. Levent Burak Kara is an associate professor in the Department of Mechanical Engineering at Carnegie Mellon University. He is the founder of Visual Design and Engineering Laboratory. His research interests include computer-aided design, computer graphics, natural user interfaces and pattern recognition, with applications in industrial product design, automotive design, engineering education and bio-medical engineering. He is the recipient of National Science Foundation Career award and American Society of Mechanical Engineers Design Automation Society Young Investigator Award. At CMU, he teaches the Mechanical Engineering Senior Capstone Design course and the graduate level AI and Machine Learning for Engineering Design course. Kara has a BS in Mechanical Engineering from the Middle East Technical University, and an MS and PhD in Mechanical Engineering from Carnegie Mellon University.



# Keynote

## Interactive Multimedia Systems and Multimodal Interfaces for Performing Arts

Kia Ng

University of Leeds, UK

**About the Speaker:** Dr. Kia Ng, PhD, FBCS, FRSA, FIoD, CEng, CSci, is Director of the Interdisciplinary Centre for Scientific Research in Music (ICSRiM) at the University of Leeds. Kia is involved in several research domains including interactive multimedia, computer vision, document imaging, gestural interfaces, multimodal analysis, and computer music, in collaboration with many European and international organisations and individuals in the field. Large-scale projects include i-Maestro on technology-enhanced learning (coordinator), CASPAR on digital preservation and AXMEDIS on cross-media. Kia has over 200 publications and presented keynotes and invited lectures in over 33 countries. See <http://www.kcng.org>

## Table of Contents

<b>Foreword</b> .....	<b>iii</b>
<b>Conference Organization</b> .....	<b>v</b>
<b>KeyNote 1: Future of Shapes: How does information and fabrication technologies change the way we design?</b>	
<b>Professor Levent Burak Kara</b> .....	<b>ix</b>
<b>KeyNote 2: Interactive Multimedia Systems and Multimodal Interfaces for Performing Arts</b>	
<b>Professor Kia Ng</b> .....	<b>x</b>
<hr/> <b>Proceedings for Distributed Multimedia Systems DMS</b> <hr/>	
Crowd Behaviors Analysis and Abnormal Detection based on Surveillance Data (S) .....	3
<i>Jing Cui, Weibin Liu and Weiwei Xing</i>	
Maintenance and Emergency Management with an Integrated indoor/outdoor Navigation Support .....	8
<i>Paolo Nesi and Pierfrancesco Bellini</i>	
A System Design for Surveillance Systems Protecting Critical Infrastructures (S) .....	17
<i>Erland Jungert, Niklas Hallberg and Niclas Wadströmer</i>	
Joint Fingerprinting and Encryption in Hybrid domains for Multimedia Sharing in Social Networks (S) .....	22
<i>Conghuan Ye, Zenggang Xiong, Yaoming Ding, Jiping Li, Guangwei Wang and Kaibing Zhang</i>	
An Improved Collaborative Movie Recommendation system using Computational Intelligence .....	30
<i>Zan Wang, Xue Yu, Nan Feng and Zhenhua Wang</i>	
An Original-Stream Based Solution for Smoothly Replaying High-definition Videos in Desktop Virtualization Systems .....	39
<i>Kui Su, Zonghui Wang, Xuequan Lu and Wenzhi Chen</i>	
Fostering Participation and Co-Evolution in Sentient Multimedia Systems .....	47
<i>Federico Cabitza, Daniela Fogli and Antonio Piccinno</i>	
Adaptive Difficulty Scales for Parkour Games .....	57
<i>Yi-Na Li, Chi Yao, Dong-Jin Li and Kang Zhang</i>	
Magic Mirror in my Hand, what is the Sentiment in the Lens?: an Action Unit based Approach for Mining Sentiments from Multimedia Contents .....	64
<i>Francesco Colace, Luca Casaburi, Massimo De Santo and Luca Greco</i>	
Ranking Highlight Level of Movie Clips: A Template Based Adaptive Kernel SVM Method (S) .....	72
<i>Zheng Wang, Gaojun Ren, Meijun Sun, Jinchang Ren and Jesse S. Jin</i>	
Relevance Measures for the Creation of Groups in an Annotation System (S) .....	79
<i>Amjad Hawash, Paolo Bottoni and Danilo Avola</i>	

Evolutionary Crowdsourcing Approach for Quality Control (S) .....	86
<i>Duncan Yung and Shi-Kuo Chang</i>	
Linked Open Graph: browsing multiple SPARQL entry points to build your own LOD views .....	94
<i>Paolo Nesi, Pierfrancesco Bellini and Alessandro Venturi</i>	
Creation and Use of Service-based Distributed Interactive Workspaces .....	104
<i>Carmelo Ardito, Paolo Bottoni, Maria Francesca Costabile, Giuseppe Desolda, Maristella Matera and Matteo Picozzi</i>	
A Framework for Bimanual Inter-Device Interactions (S) .....	113
<i>Ali Roudaki, Jun Kong and Gurisimran Walia</i>	
SWOWS and Dynamic Queries to build Browsing Applications on Linked Data (S) .....	121
<i>Paolo Bottoni and Miguel Ceriani</i>	
Hybrid Classification Engine for Cardiac Arrhythmia Cloud Service in Elderly Healthcare Management (S) .....	128
<i>Huan Chen, Bo-Chao Cheng, Guo-Tan Liao, Min-Sheng Chien and Ting-Chun Kuo</i>	
Gait Recognition Based on Joint Distribution of Motion Angles (S) .....	135
<i>Wei Lu, Wei Zong and Weiwei Xing</i>	
Exposing Image Forgery by Detecting Traces of Feather Operation .....	143
<i>Jiangbin Zheng, Tingge Zhu and Zhe Li</i>	
Dual Graph Partitioning For Bottom-Up BVH Construction .....	151
<i>Nathan Eloe, Joseph Steurer, Jennifer Leopold and Chaman Sabharwal</i>	
Path planning directed motion control of virtual humans in complex environments .....	160
<i>Song Song, Weibin Liu, Ruxiang Wei, Weiwei Xing and Cheng Ren</i>	
Athena: Capacity Enhancement of Reversible Data Hiding with Consideration of the Adaptive Embedding Level .....	168
<i>Ya-Chi Hsu, Bo-Chao Cheng, Huan Chen and Yuan-Sun Chu</i>	
User Features-aware Trust Measurement of Cloud Services via Evidence Synthesis for Potential Users (S) .....	175
<i>Hua Ma and Zhigang Hu</i>	
Efficiency of Hybrid Index Structures - Theoretical Analysis and a Practical Application (S) .....	182
<i>Richard Göbel, Carsten Kropf and Sven Müller</i>	
Creating Web3D Educational Stories from Crowdsourced Annotations (S) .....	189
<i>Ivano Gatto and Fabio Pittarello</i>	
Terminological Ontology Learning and Population using LDA (S) .....	196
<i>Francesco Colace, Massimo De Santo, Luca Greco, Vincenzo Moscato and Antonio Picariello</i>	
A Formal Model for Intellectual Relationships among Knowledge Workers and Knowledge Organizations (S) .....	204
<i>Mao-Lin Li and Shi-Kuo Chang</i>	

Ontology Bulding vs Data Harvesting and Cleaning for Smart-city Services .....	211
<i>Paolo Nesi, Pierfrancesco Bellini, Nadia Rauch, Riccardo Billero and Monica Benigni</i>	
Auto-encoder Based Bagging Architecture for Sentiment Analysis .....	221
<i>Wenge Rong, Yifan Nie, Yuanxin Ouyang, Baolin Peng and Zhang Xiong</i>	
Towards a Trust, Reputation and Recommendation Meta Model (S) .....	231
<i>Gennaro Costagliola, Vittorio Fuccella and Fernando Antonio Pascuccio</i>	
<hr/> <b>Proceedings for Visual Languages and Computing VLC</b> <hr/>	
Spatial Relations Between 3D Objects: The Association Between Natural Language, Topology, and Metrics .....	241
<i>Jennifer Leopold, Chaman Sabharwal and Katrina Ward</i>	
Nonoverlapped View Management for Augmented Reality by Tabletop Projection .....	250
<i>Makoto Sato and Kaori Fujinami</i>	
Markerless Hand Gesture Interface Based on LEAP Motion Controller (S) .....	260
<i>Danilo Avola, Luigi Cinque, Stefano Levialdi, Andrea Petracca, Giuseppe Placidi and Matteo Spezialetti</i>	
Intermodal Image-Based Recognition of Planar Kinematic Mechanisms .....	267
<i>Matthew Eicholtz and Levent Burak Kara</i>	
A Circular Visualization of People's Activities in Distributed Teams (S) .....	276
<i>Paolo Buono, Maria Francesca Costabile and Rosa Lanzilotti</i>	
A Tabu Search Based Approach for Graph Layout .....	283
<i>Fadi Dib and Peter Rodgers</i>	
eulerForce: Force-directed Layout for Euler Diagrams .....	292
<i>Luana Micallef and Peter Rodgers</i>	
A Normal Form for Spider Diagrams of Order .....	300
<i>Aidan Delaney, Gem Stapleton, John Taylor and Simon Thompson</i>	
PaL Diagrams: A Linear Diagram-Based Visual Language .....	310
<i>Peter Chapman, Gem Stapleton and Peter Rodgers</i>	
Local Context-based Recognition of Sketched diagrams (S) .....	321
<i>Gennaro Costagliola, Mattia De Rosa and Vittorio Fuccella</i>	
Towards a Lightweight Approach for Modding Serious Educational Games: Assisting Novice Designers (S) .....	329
<i>Jacob Dahleen, Alex Hunsberger, Ryan Weber, Dennis Brylow, Shaun Longstreet and Kendra Cooper</i>	
Visualization Techniques to Empower Communities of Volunteers in Emergency Management .....	335
<i>Sergio Herranz, Rosa Romero and Paloma Díaz</i>	
PetroAdvisor: A Volunteer-based Information System for Collecting and Rating Petroglyph Data .....	345
<i>Vincenzo Deufemia, Luca Paolino, Giuseppe Polese, Viviana Mascardi and Henry de Lumley</i>	

Combining personal diaries with territorial intelligence to empower diabetic patients.....	354
<i>Athula Ginige, Pasquale Di Giovanni, Monica Sebillo, Genny Tortora, Maurizio Tucci and Giuliana Vitiello</i>	
Predicting Traffic Congestion in Presence of Planned Special Events (S) .....	357
<i>Simon Kwoczek, Sergio Di Martino and Wolfgang Nejdl</i>	
Representing Uncertainty in Visual Integration (S) .....	365
<i>Bilal Berjawi, Elisabeth Chesneau, Fabien Duchateau, Franck Favetta, Claire Cunty, Maryvonne Miquel and Robert Laurini</i>	

---

### Proceedings for Distributed Education Technologies DET

---

Smarter Universities: a Vision for the Fast Changing Digital Era .....	375
<i>Mauro Coccoli, Angela Guercio, Paolo Maresca and Lidia Stanganelli</i>	
Interaction with Objects and Objects Annotation in the Semantic Web of Things.....	383
<i>Mauro Coccoli and Ilaria Torre</i>	
Computer Tutors Can Reduce Student Errors and Promote Solution Efficiency for Complex Engineering Problems (S).....	391
<i>Paul S. Steif, Matthew Eicholtz and Levent B. Kara</i>	
Digital Agenda and E-learning in Italian Public Administration .....	399
<i>Andrea Molinari, Milena Casagrande and Luigi Colazzo</i>	
Castor: Designing and Experimenting a Context-Aware Architecture for Creating Stories Outdoors (S) .....	405
<i>Fabio Pittarello and Luca Bertani</i>	

---

### Poster and Demo

---

Distance Learning Immersive Environments: Sense of Presence Exploration (P).....	415
<i>Max North</i>	
The Recovery System for Hadoop Cluster (P) .....	416
<i>Priya Deshpande and Darshan Bora</i>	
Improving App Inventor Usability via Conversion between Blocks and Text (P).....	421
<i>Karishma Chadha and Franklyn Turbak</i>	
A Decision Support System for Evidence Based Medicine (P) .....	423
<i>Giuseppe Polese</i>	

Author's Index .....	A-1
Program Committee's Index .....	A-5
External Reviews' Index .....	A-7

### Note:

(S) indicates a short paper.

(P) indicates a poster or demo, which is not a refereed paper.



# Proceedings for Distributed Multimedia Systems



# Crowd Behaviors Analysis and Abnormal Detection based on Surveillance Data

Jing Cui, Weibin Liu

Institute of Information Science  
Beijing Jiaotong University  
Beijing Key Laboratory of Advanced  
Information Science and Network Technology  
Beijing 100044, China  
wbliu@bjtu.edu.cn

WeiWei Xing

School of Software Engineering  
Beijing Jiaotong University  
Beijing 100044, China

**Abstract**—Crowd analysis and abnormal trajectories detection are the hot topics in computer vision and pattern recognition. As more and more video monitoring equipments are installed in public places for public safety and public management, researches become urgent to learn the crowd behavior patterns through the trajectories obtained by the intelligent video surveillance technology. In this paper, the FCM (Fuzzy c-means) algorithm is adopted to cluster the source points and sink points of trajectories that are deemed as critical points into several groups. Naturally, the trajectory clusters can be acquired. After refining them, the feature information statistical histogram for each one which contains the motion information will be built after refining the trajectory clusters with Hausdorff distances. Eventually, the local motion coherence between test trajectories and refined trajectory clusters will be used to judge whether they are abnormal.

**Keywords:** Crowd analysis, abnormal trajectories detection, FCM, feature information statistical histogram

## I. INTRODUCTION

As more and more video monitoring equipments are installed in public places for public safety and public management, researchers can learn the motion patterns of crowds and do further studies by analyzing the observed data. As the traditional methods can not be applied in the analysis of unstructured situations, to overcome this problem, we propose a new approach. The approach adopted in our research focuses on the motion patterns learning and abnormal trajectories detecting. Our approach has several advantages: Firstly, FCM[8] is used to cluster the source and sink points, and the hidden unstructure information of the unstructured scene will be learned. Secondly, according to the hidden unstructure information, we will get the training trajectory clusters. And the parallel coordinates which can represent data in high-dimension are used to describe the motion patterns of crowd. Thirdly, we can judge which trajectory cluster the test trajectory most possibly belongs to with the hidden unstructured information and our training trajectory clusters learned before. Then we just need to make a compared between the test trajectory and the cluster which it most possibly belongs to, instead of the whole

clusters. As a result, the computational efficiency is improved greatly.

## II. RELATED WORK

The data observed by monitoring equipments in a scene usually can not be studied directly. Researchers, like Sugimura et. al. [2], proposed a method for tracking persons in the crowd. After transforming the observed data into trajectories of tracking objects, the crowd behaviors can be analyzed. Crowd behavior analysis has three major aspects: motion patterns learning, abnormal behaviors detection and behaviors prediction. Next, we will briefly describe some of the achievements on them.

Generally speaking, motion patterns learning is the primary step in the related studies. It practices the regular motion trajectories, namely, motion patterns, by using the observed data. For instance, Fatih Porikli et.al[3] learned the trajectory patterns by computing affinity matrices and applying eigenvector decomposition. Few years later, an improved DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method was used to divide the motion flows into different patterns[4].

Abnormal behaviors detection aims at identifying the movement behaviors which are obviously different from other motion tracks or have low probabilities of occurrence by using the motion patterns discovered before. Claudio Rosito Jung et.al.[1] proposed a approach that used 4-D histogram to make abnormal detection. Stauffer et.al. [5] modeled each pixel as a mixture of Gaussians, used an on-line approximation to update the model at the same time.

Behaviors prediction is a subject which attracts much attention. Researchers desire to forecast the next moving region or semantic behavior based on the priori knowledge and motion patterns of moving objects. Josh Jia-Ching Ying et.al[6] combined the geographic features and the semantic features of users' trajectories together, and then it evaluated the next location of a mobile user based on the frequent behaviors of similar users in the same cluster.

In addition, other aspects of crowds also appeal to scholars. Jan Šochman et.al [7] proposed an automatic on-line inference of social groups based on the Social Force Model in crowded scenarios.

---

This research is partially supported by National Natural Science Foundation of China (61370127, 61100143), Program for New Century Excellent Talents in University (NCET-13-0659), Fundamental Research Funds for the Central Universities (2014JBZ004), Beijing Higher Education Young Elite Teacher Project (YETP0583). {Corresponding author: Weibin Liu, wbliu@bjtu.edu.cn}

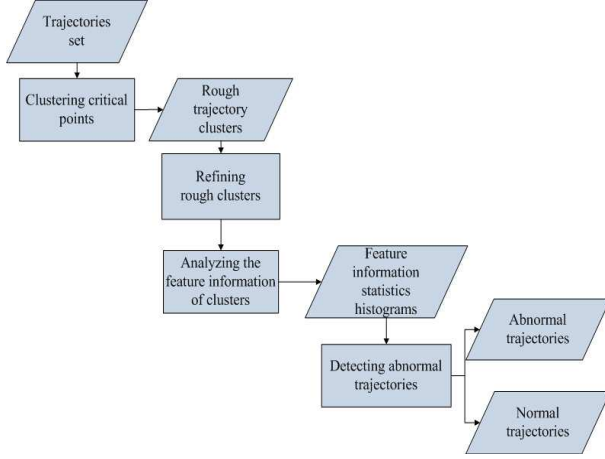


Figure 1: Framework of crowd analysis and abnormal detection

### III. OVERVIEW OF OUR FRAMEWORK

The framework of our approach roughly includes five stages: 1) trajectory preprocessing; 2) critical points clustering; 3) rough trajectory clusters refining; 4) analysis of the behaviors of crowd; 5) abnormal trajectories detection, as shown in Figure 1.

### IV. CROWD MOTION PATTERNS LEARNING

#### A. Trajectory Preprocessing

In most researches, each trajectory will be represented by a sequence of flow vectors.

$$F_n = \{f_1, f_2, \dots, f_n\} \quad (1)$$

Where  $n$  is the length of the trajectory. The flow vector in Equation (1) is formed by a tetrad as follow:

$$f_n = \langle x(t_n), y(t_n), \theta(t_n), v(t_n) \rangle \quad (2)$$

It contains the spatial, direction and velocity information of the trajectory at time.

As the persons pass the scene from different regions and the speeds of them are not uniform, the trajectories of moving objects need a pre-process to get a unity length. In the preprocessing stage, we use linear interpolation to get normalization. In that case, we can perform the following study more conveniently. The trajectory of each object has its unique features. The detailed information of it usually can not be studied directly. Hence the researchers always make some preprocessing to match their needs. For example, if persons who pass the scene through different regions or with variant velocities, the length of their trajectories tracked by the camera may have a different value. For next experiments, we should let each trajectories be represented by the same number of flow vectors. That is to say, the number of sample points within each trajectory should be unified so that we can assess the similarity between trajectories.

The common trajectory normalization methods contain resample and smoothing. As the trajectory of each object is described by a flow vectors set in chronological order, there we use the resample based on one dimension linear interpolation to

make all the trajectories described by the flow vector set with the same number. Eventually, the resample points computed by our algorithm can help to access the similarities between trajectories in the next processes easily.

#### B. Critical Points Clustering

In order to analyze the behaviors of crowd, we'd better to cluster the similar trajectories into one same group for further research. In our research, we extract the critical points of all tracks: source point and sink point of each trajectory, which usually appear at the edge regions of the scene. As so far, researchers have developed a lot of clustering methods, it is pointed out that the standard FCM algorithm is robust to the scaling transformation of the dataset, while others are sensitive to such transformation.

In our research, the FCM algorithm is performed for critical points clustering. And then we can obtain  $N$  points groups. Obviously, we can get  $N * N$  motion patterns of crowd trajectories roughly.

#### C. Build Feature Information Statistical Histograms for Refined Clusters

After the previous works, we have gotten the rough similar trajectories. A group of similar trajectories means a pattern of crowd behaviors. In order to learn the behavior rules better, we should find the center trajectory whose sum of the Hausdorff distances to other trajectories in its cluster is minimum. And then those trajectories that have unreasonable Hausdorff distances to the center trajectory of its cluster should be removed. As a result we achieve refined trajectories clusters and then build the feature information statistical histograms for them.

After the previous works, we have gotten the rough similar trajectories. In order to learn the behavior rules better, we should remove the trajectories that have unreasonable Hausdorff distances to the center trajectory of its cluster to refine trajectory clusters, and transform sample  $(x, y)$  into  $(x, y, \theta, v)$  to build the feature information statistical histograms which can describe the probability distributions of trajectories in the scene.

In order to get the histograms. First of all, we should discretize each trajectory as a sequence of  $(x_i^d(t), y_i^d(t), \theta_i^d(t), v_i^d(t))$ . The feature information statistical histogram  $H_k$  related to the  $k$ th cluster is built by spreading the according to a kernel  $g$

$$H_k(x^d, y^d, \theta^d, v^d) = \sum_{i=1}^{N_k} \sum_{t=1}^{N_f(i)} g(x^d - x_i^d(t), y^d - y_i^d(t), \theta^d - \theta_i^d(t), v^d - v_i^d(t)) \quad (3)$$

Where  $N_k$  is the number of the trajectories in  $k$ th cluster and  $N_f(i)$  is the length of the  $i$ th trajectory in cluster  $k$  and  $g(x, y, \theta, v)$  is the spreading kernel.

$$g(x, y, \theta, v) = g_1(x, y)g_2(\theta)g_3(v) \quad (4)$$

In order to build the feature information statistical histogram for each cluster, three steps should be followed:

1) The first step is to discretize the information of each point in  $i$ th trajectory. we change the location of the moving object  $(x, y)$  to a discrete value  $(x^d, y^d)$ , where  $x^d \in \{0, 1, \dots, N_x-1\}$ ,  $y^d \in \{0, 1, \dots, N_y-1\}$  and  $N_x$  is the number of lines of the image,  $N_y$  is the number of columns of the image.

2) Then we transform the local direction vector  $\theta$  into a discrete value  $\theta_d$ , which has  $N_\theta$  levels, and each level comprises a circular sector with internal angle:

$$\theta^d \in \{\theta_0, \theta_1, \dots, \theta_{N_\theta-1}\}, \theta_j = \frac{j}{N_\theta} 2\pi \quad (5)$$

At the same time, the velocity  $v$  is turned into a discrete value  $v^d$  within  $N_v$  levels (low, middle and high)

$$v^d = \begin{cases} 0(\text{low speed}) & , \text{if } 0 \leq v < v_t \\ 1(\text{mediu speed}) & , \text{if } v_t \leq v < v_h \\ 2(\text{high speed}) & , \text{if } v_h \leq v \end{cases} \quad (6)$$

$$v_t = \mu_v - k\sigma_v, v_h = v_t + \mu_v + k\sigma_v \quad (7)$$

where  $k$  control the velocity, set  $k=2$ . In each cluster, we will obtain an array about the velocity of trajectories,  $\mu_v$  is the mean of the array, and  $\sigma_v$  is its variance.

3) The third step is to compute the histograms with Equation (3) and Equation (4). For  $g(x, y)$  in Equation (4), a discrete Gaussian function is the best choice. And  $\sigma$  is the standard deviation of function. The normal spatial distribution in  $n$ -dimensional space is shown as in Formula (8).

$$g(r) = \frac{1}{\sqrt{2\pi\sigma^2}^N} e^{-\frac{r^2}{2\sigma^2}} \quad (8)$$

As the research works on a 4-dimensional space, we can transform the above equation into Formula (9).

$$g_1(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}} \quad (9)$$

Then we compute the  $g_2(\theta)$ , the second part of Equation (4), as in the following equation,

$$g_2(\theta) = \max\{0, 1 - \frac{|\theta|_{ang}}{\Delta\theta_{N_\theta}}\} \quad (10)$$

Where  $\theta \in \{\theta_0, \dots, \theta_{N_\theta-1}\}$ ,  $\theta_j$  is a discrete direction value,  $\Delta\theta_{N_\theta} = 2\pi/N_\theta$  and  $|\theta|_{ang} = \min\{|\theta|, 2\pi - |\theta|\}$ .

The last item of the spreading kernel is  $g_3(v) = \delta(v)$  (11),  $\delta(v)$  is the discrete Dirac Delta(unit impulse) function. Eventually, we can get a histogram with dimensions  $N_x \times N_y \times N_\theta \times N_v$  for each cluster. It should be noticed that a feature information statistical histogram is computed for each cluster and it represents a motion pattern.

## V. ABNORMAL TRAJECTORY DETECTION

According to the result of previous experiments, the refined clusters and their histograms have been obtained to help us judge whether the test trajectories are abnormal or not. We will use the following formula to calculate the local consistency between the given trajectory and the  $k$ th cluster histogram  $H_k$ .

$$d_k(t) = H_k(x^d(t), y^d(t), \theta^d(t), v^d(t)) \quad (12)$$

If they are local consistent at time  $t$ , the value of  $d_k(t)$  will be large. Otherwise the value will be small. These abnormal trajectories are that most sample points of them have low level values of  $d_k(t)$ .

We should choose a threshold value  $T_k$  for each cluster. If  $d_k < T_k$  (13), we will regard the trajectory as local abnormal at time  $t$ . Each trajectory of  $k$ th cluster should be calculated the value of  $d_k^i(t)$ , then we can obtain a set  $D_k$  of  $k$ th cluster as in Equation (14)

$$D_k = \bigcup_{i=1}^{N_k} \bigcup_{t=1}^{N_f(i)} \{d_k^i(t)\} \quad (14)$$

where  $T_k$  is  $r$ -quantile of the distribution of  $D_k$ .

## VI. EXPERIMENT RESULTS

This section will demonstrate the results of experiment based on the previous work. All the experiment data are from the pedestrian trajectory database of Edinburgh University. In our research, we draw 20000 pedestrian trajectories randomly to analyze and study the motion patterns of them.

### A. Critical Points Clustering

Firstly, we extract the source and sink points of all the trajectories in our database. According to Figure 2(a)-(b), we can see that the entry regions and exit regions are almost bi-directional; So we merge source set and sink set to a whole point set for further learning. After that, we should group all the points into several clusters by FCM. In this study, the number of clusters that can lead the best result will be 12. All point clusters are shown in Figure 2(c). Through grouping the critical point set into several clusters some structure information of the giving scene can be learnt at this stage.

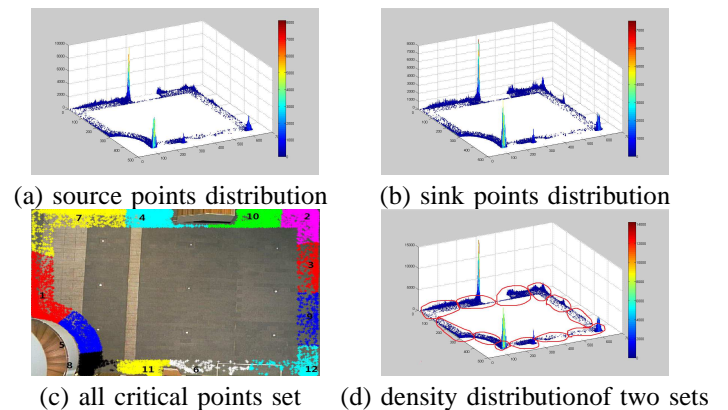


Figure 2: Point clusters learning

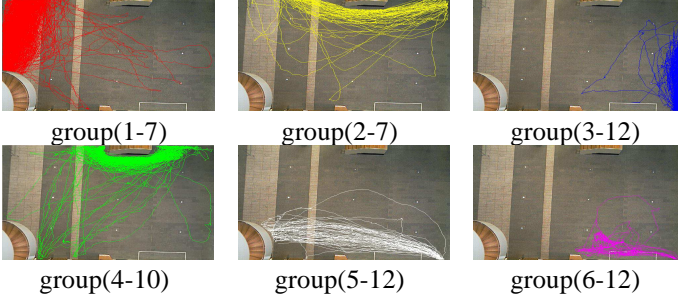


Figure 3: Several important rough trajectory clusters

### B. Trajectory Motion Patterns

In the first step, 12 point clusters can be found. Then, we regard the trajectories which appeared at region A and disappeared at region B as a rough trajectory cluster. That is to say, we have achieved 144 rough trajectory clusters. Several prominent trajectory clusters are shown in Figure 3.

Secondly, with the purpose of learning crowd motion regularity easily and accurately, we need to refine the trajectory clusters obtained before. A group of similar trajectories means a pattern of crowd behaviors. In order to learn the behavior rules better, we should find the center trajectory whose sum of the Hausdorff distances to other trajectories in its cluster is minimum. And then those trajectories that have unreasonable Hausdorff distances to the center trajectory of its cluster should be removed. As a result we achieve refined trajectories clusters and then build the feature information statistical histograms for them.

Moreover, building feature information statistical histograms for refined clusters can help us learn the probability distribution of trajectories. In order to describe all the feature information about the clusters, we adopt the parallel coordinates as shown in Figure 4.

Parallel coordinates is a common method for the high dimension data visualization. A high dimension data point  $(x^d, y^d, \theta^d, v^d, H_k(x^d, y^d, \theta^d, v^d))$  can be expressed as a broken line. The inflection points of it are located at each parallel axis and they can show the value of corresponding dimension. What's more, the value of  $H_k(x^d, y^d, \theta^d, v^d)$  shows that the sum of the probabilities of each location sample point in cluster  $k$  belongs to the statistic class  $(x^d, y^d, \theta^d, v^d)$ .

### C. Abnormal Detection

Eventually, we extract 5000 new trajectories to make abnormal detections. The critical points of these test trajectories are

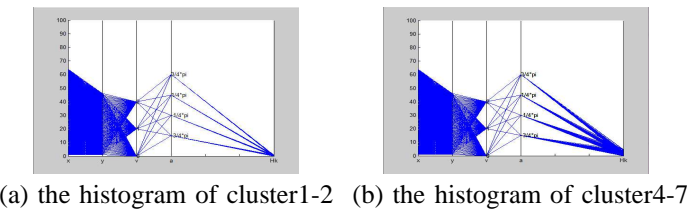


Figure 4: Feature information statistical histograms of clusters

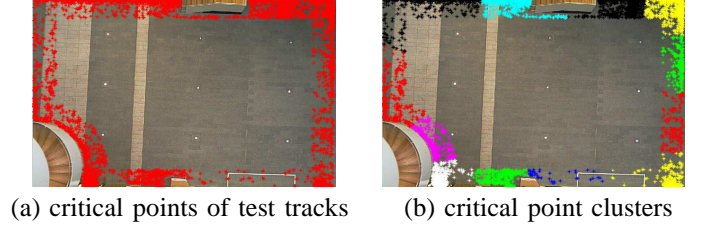


Figure 5: Status of the critical points of test trajectories

divided into 12 groups according to the regions they belonged to as shown in Figure 5.

Moreover the status of test trajectories is displayed in TABLE I. Each element in TABLE I represents the number of test trajectories that most possible belong to each training cluster learned before. Then the later work can determine the correctness of our preliminary judgments.

TABLE I: ENTRY-EXIT REGION TRANSFERRING MATRIX OF TEST TRAJECTORIES

		Exit Region											
		1	2	3	4	5	6	7	8	9	10	11	12
Entrance Region	1	61	6	0	14	17	1	44	4	3	9	1	1
	2	1	74	16	190	12	7	22	121	4	63	17	53
	3	0	6	15	22	0	0	19	12	13	4	0	57
	4	12	216	22	45	64	2	44	644	7	67	54	115
	5	18	6	1	65	6	0	18	14	0	34	6	8
	6	0	2	1	4	2	3	1	0	1	3	0	8
	7	54	26	14	26	19	1	62	93	0	16	25	22
	8	26	111	12	586	9	1	71	7	0	242	69	34
	9	0	11	21	0	0	1	1	0	37	3	0	24
	10	1	92	13	60	8	1	22	109	1	71	16	6
	11	5	13	1	34	3	3	24	82	0	14	13	21
	12	2	54	50	120	8	8	31	41	13	12	24	18

Next, we detect the local motion coherence between test trajectories and histogram set of refined trajectory clusters, the value of  $r$  mentioned in section V is set to 20. In Figure 6, figure (a) shows most of trajectory 42305(the index of this trajectory in test set is 42305) are local coherence with the histogram of cluster2-12. Then in (b), most of the test trajectory 39206 the index of it is 39206 are not local coherence with the same histogram. So we regard trajectory 39206 as an abnormal one.

The trajectory, most parts of which are coherence with the histogram will be regarded as normal. Otherwise it will be reckoned as an abnormal trajectory. Several prominent detection results are shown in Figure 7. The refined trajectory

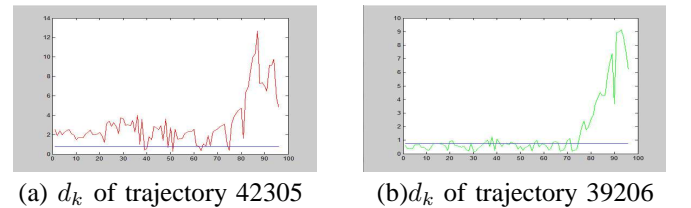


Figure 6: Local motion coherence of two trajectories between histogram of cluster2-12



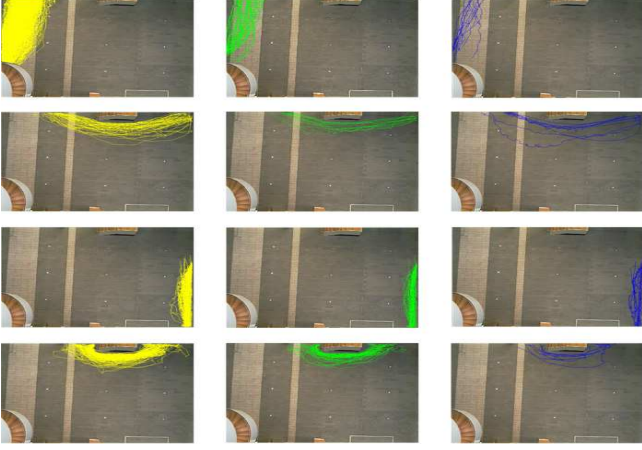


Figure 7: Test trajectories detection

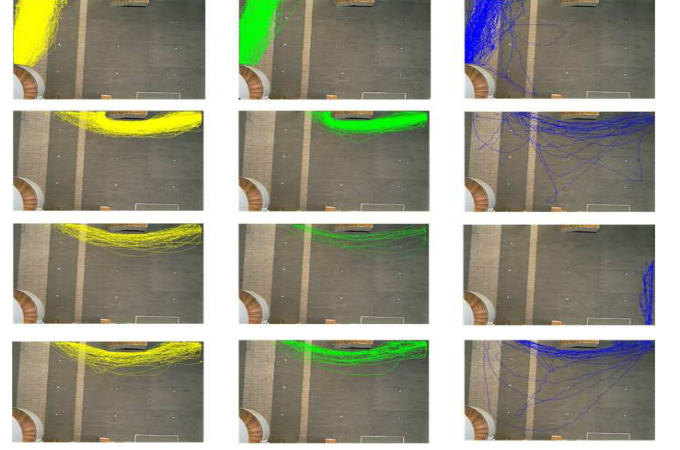


Figure 8: Abnormal detection for test set2

clusters on the left, the normal trajectories for them on the middle and the abnormal trajectories on the right.

In order to describe the effect of our abnormal detection step better, we also prepare another test set with 10000 trajectories. The status of it is displayed in TABLE II. And several detection results are show in Figure 8. It proves the effect of our approach again.

TABLE II: ENTRY-EXIT REGION TRANSFERRING MATRIX OF TEST SET2

		Exit Region											
		1	2	3	4	5	6	7	8	9	10	11	12
Entrance Region	1	71	12	0	62	32	0	314	45	3	13	34	38
	2	5	107	18	245	18	3	41	134	5	95	47	105
	3	0	13	22	28	0	0	22	10	17	5	2	94
	4	56	414	27	38	310	4	79	980	15	223	108	288
	5	33	11	4	255	12	0	77	45	0	66	27	20
	6	0	3	0	2	0	3	0	3	1	1	5	5
	7	350	58	32	47	80	2	159	217	2	39	84	103
	8	54	116	11	843	28	1	228	22	0	380	105	99
	9	3	13	22	1	0	1	2	0	37	9	0	29
	10	2	113	19	101	50	0	29	199	1	72	28	15
	11	30	49	5	65	24	2	75	118	0	33	26	64
	12	34	134	104	292	17	4	130	112	15	13	53	21

We can see that if the pass regions of a test trajectory are different from most members in its corresponding cluster, it will be found out as an abnormal one, just as most in Figure 7 and 8. Even though some trajectories seem similar with other members in corresponding cluster in spatial aspect, the motion direction or moving velocity may have distinct differences. They are also deemed to have another motion tendency and also be regarded as abnormal, like the examples in the third line of Figure 7.

## VII. CONCLUSION

In this paper, we have done some progressive work in the crowd behaviors analysis and abnormal trajectory detection: Experiment result indicated that the output (normal trajectories) produced accurately by our method was mostly coherent with the test cluster. The advantage of our approach is that the hidden unstructure information of the unstructured scene

are learned on the training step, so we can judge which trajectory cluster the test trajectory most possibly belongs to with it. Then on the detecting step, the test trajectory just needs to compare with the trajectory cluster which has a high motion consistency with it, instead of the whole clusters. As a result, the computational efficiency is improved greatly.

In the future, the following work can be carried out as improvements of the method: the more optimal cluster algorithm for critical points clustering should be implemented for learning the cluster number automatically, and the behavior prediction can use the motion patterns obtained in section IV to predict the next moving region and semantic behavior.

## REFERENCES

- [1] C.R.Jung, L.Hennemann and S.R.Musse. Event Detection using Trajectory Clustering and 4-D Histograms. IEEE Transactions on Circuits and Systems for Video Technology. Vol. 18, November 2008, pages 1565-1575,
- [2] D.Sugimura, K.M.Kitani, T.Okabe, Y.Sato and A.Sugimoto. Using Individuality to Track Individuals: Clustering Individual Trajectories in Crowds using Local Appearance and Frequency Trait. 12th IEEE International Conference on Computer Vision, Kyoto, September 2009, pages 1467-1474;
- [3] F.Porikli. Learning Object Trajectory Patterns by Spectral Clustering. IEEE International Conference on Multimedia and Expo. vol.2, Taipei, June 2004(ICME), pages 1171-1174;
- [4] W.He and Z.Liu. Motion Pattern Analysis in Crowded Scenes by using Density based Clustering. 9th International Conference on Fuzzy Systems and Knowledge Discovery, Sichuan, May 2012(FSKD2012), pages 1855-1858;
- [5] C.Stauffer and W.E.L.Crimson. Learning Patterns of Activity Using Real-Time Tracking. IEEE Transactions on Pattern Analysis And Machine Intelligence. Vol.22, August 2000, pages 747-757.
- [6] J.J.Ying, W.C.Lee, T.C.Weng and V.S.Tseng. Semantic Trajectory Mining for Location Prediction. ACM SIGSPATIAL GIS11, Chicago, November 2011, pages 34-43;
- [7] J.Šochman and D.C.Hogg. Who Knows Who Inverting the Social Force Model for Finding Groups. IEEE International Conference on Computer Vision Workshops, Barcelona, November 2011, pages 830-837;
- [8] J.C.Bezdek, B.Ehrlich, W.Full. FCM: the fuzzy c-means clustering algorithm. Comput Geosciences, Vol.10, October 1984, pages 191-203.

# Maintenance and Emergency Management with an Integrated indoor/outdoor Navigation Support

Pierfrancesco Bellini, Paolo Nesi, Matteo Simoncini, Alessandro Tibo

Distributed Systems and Internet Technology Lab, DISIT Lab, Department of Information Engineering  
University of Florence, Florence, Italy, tel: +39-055-4796523, fax: +39-055-4797363  
<http://www.disit.dinfo.unifi.it>, [paolo.nesi@unifi.it](mailto:paolo.nesi@unifi.it), [pierfrancesco.bellini@unifi.it](mailto:pierfrancesco.bellini@unifi.it)

**Abstract** -- Large and complex infrastructures as industry plants and hospitals are vulnerable to natural, man-made disasters, and causality events. In this paper, we present a solution addressing the guiding personnel during maintenance and/or emergency conditions. The aim is to reduce the time needed to react and to cope with organization and maintenance support, while facilitating communication, and indoor / outdoor navigation. The solution is based on the formalization of protocol, the modelling of knowledge for navigation, the algorithms and the development of a mobile application and corresponding server device for integrated indoor/outdoor navigation. The navigation algorithms are based on low costs mobile sensors and Adaptive Extended Kalman Filter. The solution has been validated and tried out within a large medical infrastructure, thus demonstrating the validity of the identified modalities and procedures, measuring the advantage from both qualitative and quantitative aspects. The indoor navigation solution has been compared with other former solutions based on classical Kalman and dead reckoning.

**Keywords** *maintenance and emergency management, mobile emergency, indoor / outdoor navigation.*

## I. INTRODUCTION

Reactions with respect to incidents represent one of the greatest challenges in maintenance and emergency management. In most cases, the accessible information on the nature of the incidents is inaccurate as the needs to solve them; thus the personnel is inefficiently coordinated, informed neither on real conditions, nor on available resources. The logistics aspects related to the intervention and to the movement of personnel and patients are very relevant. Information and communication technologies play a vital role in coordinating crisis response between services and the emergency departments [1]. In [2], the emergency plan has been transformed into a multimedia software environment to combine information coming from different sources. Mobile devices have the potential to improve the response to incident and emergency. In that case, the organization of the central station has been revised in order to reduce the possible bottleneck and facilitating the organization of the information. Involved personnel need to have access at updated information and knowledge in the emergency and maintenance conditions. The knowledge supports personnel in a continuously changing situation, where what is needed is the adoption of local standard intervention protocols, complex dosages, checklists, etc. [3]. Therefore, mobile devices are mandatory tools for information access and to help sometimes in taking decisions. On such grounds, the solution has to guarantee the access to any right and updated information in the needed time [4], [5]. In addition, the identification of the positions of critical points,

emergency facts, of the personnel involved in the emergency scenarios, are very important aspects to be addressed. In fact, the main purposes of managers are to provide support to (i) rescue/ maintenance teams to obtain all the relevant information about a problem state and to know how to reach it, (ii) personnel involved in the critical area in getting the closest updated emergency exits, (iii) rescue/maintenance teams to reach points of interest, POIs, to solve the critical issues and to collaborate each other. To cope with the above aspects, an integrated indoor/outdoor position and navigation solution is fundamental.

Outdoor navigation systems are accessible from almost all smartphones. On the other hand, low cost precise indoor navigation systems are still problematic. And, the integration of indoor-outdoor navigation presents open problems. The simple integration of mapping indoor and outdoor maps is the first step [6]. Moreover, also the precise detection of the indoor/outdoor condition is a complex problem [7]. The condition detection can be based on light intensity, on GPS (Global Positioning System) accuracy, Wi-Fi power, magnetic field, etc. In/Out detectors can be used to switch from different navigation algorithms and to reduce the energy consumption. In the literature, many indoor navigation systems for robotics systems have been proposed using different techniques and solutions based on inertial, sensors, and markers as well as a mixture of them, taken from: dead reckoning, WiFi, Augmented reality, RFID (Radio-frequency identification), QR (Quick Response), etc.

In this paper, an integrated and cheap solution called Mobile Emergency Pro for supporting personnel in large infrastructures as factories and hospitals for maintenance and emergency management is presented. This paper is focussed on presenting the technical aspects of the system functionally of the new version of an early solution called Mobile Emergency presented in [5] and developed for hospital emergency management. The main improvements added with respect to the former version are related to: (i) the insertion of an improved solution for indoor-outdoor navigation based on adaptive Kalman filtering, (ii) the corresponding improvement of the mapping and point of interest (POI) modelling. The solutions proposed have been compared with the state of the art solutions based on dead reckoning (as the former Mobile Emergency tool [5]) and classical Kalman filtering.

The paper is organized as follows. In section II, the overview of the scenarios addressed by the solutions is presented. Section III describes the architecture of the solutions, and details about the mobile application called Mobile Emergency Pro. The provide description of the architecture



presents some details regarding the protocol for data interchange from server to the mobile applications and the facilities for collaboration among personnel. In addition, details regarding the implementation of the aspects related to the integrated indoor/outdoor navigation are also reported. Section IV presents the state of the art and the integrated inertial indoor-outdoor navigation model and algorithms (based on adaptive Kalman filtering) on which most of the app functionalities are based. In Section V, a comparison of the proposed solution with respect to the state of the art solution is reported together with some experimental results. Conclusions are drawn in Section VI.

## II. OVERVIEW OF SCENARIOS AND REQUIREMENTS

Mobile Emergency Pro solution aims at (i) managing communications among personnel during both maintenance and emergency management interventions, (ii) provide support with information and navigation (to reach the area of interest or to escape from them). The application scenario is focussed on managing these problems in large infrastructures based on several buildings, with thousands of personnel and visiting people (e.g., parks, industries, hospitals, etc.). In these large and complex scenarios, several interventions/events may occur per week, and sometimes per day. They may range from simple maintenance problems to serious fire outbursts. During emergency situations, there are many additional constraints. Connections can be discontinuous (even in the event of multiple networks and protocols: Wi-Fi, UMTS, GPRS,...). The infrastructures are supported by: a capillary positioning of plates, operative manuals, information on the walls to provide positions of stairs, escapes, extinguishers, phone numbers, while control cabinets are sensitive information for security reasons.

The internal personnel is the most credible in informing the Central Station about the inception of problems. The calls are performed via voice call, SMS and/or web based tools, where images and videos depicting the event could be useful to compose the scenario of the emergency/intervention.

In the context of emergency, the personnel may be involved in patient assistance. Traditional emergency guidelines and protocols do not offer support for team creation to cope with such kinds of problems. Specific additional collecting areas for each emergency/triage level have to be set up, to start treating patients in the area of disaster. These activities may be accelerated by recalling medical personnel from other areas of the hospital. Mobile device could be used to facilitate the aggregation and coordination of collaborative teams [5]. Moreover, the rescue team and personnel need to reach the emergency position and may be not fully aware about the precise location of each department, building and room where he/she has to come to. If the position of each person is known in real time, the Central Station may better coordinate the reactions to the events, the formation of teams, etc. Our aim was to create a solution including a Central Station (server) and a mobile application (namely: Mobile Emergency Pro application) to improve the readiness of personnel during the events, facilitate communication, assure positioning, provide information and knowledge, help teams and services in: reaching the event locations, taking decisions, and thus allowing more efficient rescue operations for the victims. The

main idea is to support the emergency management with the aim of: reducing time of intervention and coordination; facilitate and improve the coordination among personnel and structures involved; facilitate the understanding and the activities of the personnel; and provide support in absence of communication support.

## III. ARCHITECTURE OVERVIEW

According to the above described scenario a Central Station and the Mobile Emergency Pro mobile application have been designed and developed. As depicted in Figure 1, the main architecture of Mobile Emergency Solution is made of three main elements: the Central Station, the Mobile Emergency Pro application and the Mobile Medicine Server (<http://mobmed.axmedis.org>), which is a best practice network on medical procedures.

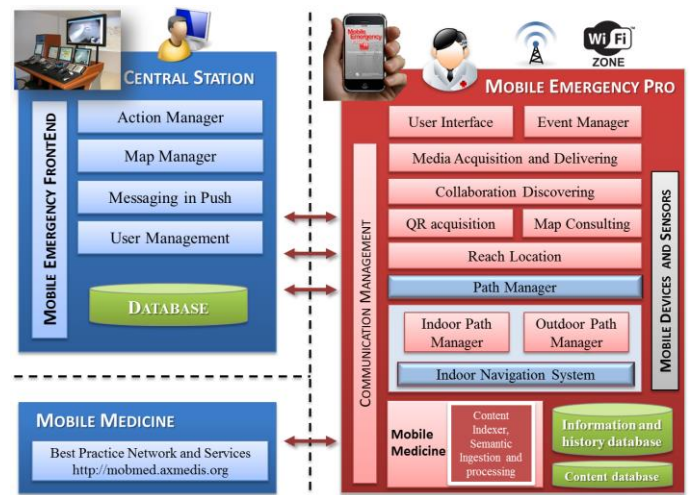


Figure 1. Architecture of the Mobile Emergency Pro solution

The **Central Station, CS**, provides services for the mobile devices, collecting alarms, supporting the personnel during the emergency/maintenance, providing support for:

- receiving alarms for intervention and emergency (classification and grouping of them in aggregated events),
- supporting people involved in the event/ emergency. They may need to (i) know the effective event/emergency status (location and area, severity, collecting areas, video and images, responsible, etc.), (ii) get the most viable and closest exit according to the context, (iii) reach a collecting area, (iv) reach a POI, (v) have support for moving a patient, (vi) establish collaboration among personnel.
- sending messages to the mobile devices in push by using the Apple Push Notification Service, APN. This activity is performed to: (i) provide them with the information about the next intervention, (ii) recall personnel, (iii) keep informed the personnel.

The **Map Manager** includes support for indoor map management on the CS. It provides fresh maps and related information to mobile devices via HTTP. The information associated with each map enables the integrated indoor-outdoor

navigation: scale, orientation, position of the exits, position and colour of collecting areas, POI with their name, ID, type, etc. The scale and the orientation are fundamental parameters for the proper functioning of the indoor navigation system. The scale indicates the ratio between centimetres and pixels. The orientation allows to present the map in a coherent manner to the user during the navigation, keeping the device magnetic Nord and the map magnetic Nord aligned according to the device movements. Once defined the scale and the orientation, the Map Manager operator can start the mapping phase of POI, exits, collecting areas, fire extinguishers, medical kits, stairs... For each of them a QR code is automatically generated and made available for the internal personnel to be placed, integrated in official plates.

### III.A Mobile Emergency Pro

The Mobile Emergency Pro mobile application is available on iPhone to provide support for personnel. The main functionalities are: possibilities of communicating events (maintenance and/or emergency), monitoring events/emergency getting information from the server, collaborating with other colleagues via mobile communication, navigating in/out getting information from the Mobile Medicine Best Practice Network [3]. The hospital area is covered by mobile network with GPRS, EDGE, HSDPA, 3G, and by local WiFi networks. The Mobile Emergency Pro application has to provide support for: authenticating and communicating with the CS to get fresh information about event / emergency status, enforcing the procedures, taking the positions by exploiting mobile devices sensors and QR, managing the maps and related information points and exits, discovering other users and collaborating, accessing to mobile medicine procedures and tools, and finally for navigating from the current position to the identified target (exit, colleagues, collecting area, points of interest, POIs, and tools) by using the integrated indoor / outdoor solution. In addition, the mobile application kept stored continuously updated information regarding maps, procedures, and the log of the user actions. This information can be used to reconstruct actions whenever a legal analysis of facts is requested. In respect of privacy policies, the user is informed about these aspects when the mobile device is registered to the CS. The system configuration includes the protocol and classification according to the maintenance/emergency manual. This information is enforced into the system during the set up and configuration by using XML files.

According to Figure 1 of the general architecture, the **Event Manager** allows the user to: (i) formalize and send the alarms and follow the procedure manual adopted, attaching a media (typically a video or some images, collected by using the **Media Acquisition and Delivering module**), (ii) receive direct calls from the CS in push (as suggestions, actions to be performed, tasks and the emergency sheet to coordinate the teams, assignments to move, to join a team, to become the responsible of a team, to move a different area and room), (iii) monitor the status of the active events/emergencies from the CS. In Figure 2, we can see the UML sequence diagram describing the methods sequence required to download the list of emergencies in progress. By creating a new NSURLConnection instance, the system makes an HTTP request to the Central Station. The

system gets all the information about the emergencies in progress and codes it by using the XML format. Once downloaded the XML file, the application parses the data through the class DownloadXMLParser and saves them in the device database allowing the user to consult them also in offline mode.

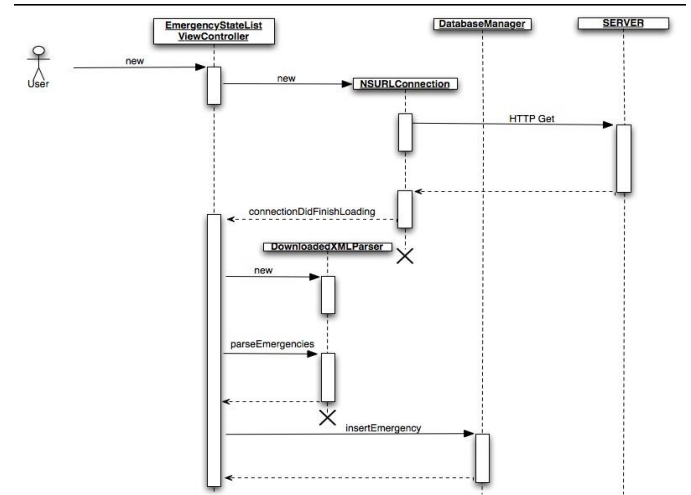


Figure 2 - UML Diagram sequence: recovery of emergencies status

The emergency status is provided via an XML containing the information related to the active emergencies. The mobile application parses the information to store it into the local database and provide it to the user. In the following, an example of the XML file is reported. The example shows a list of two emergencies which are located into the hospital.

```

<?xml version="1.0" encoding="UTF-8"?>
<emergencies>
  <emergency>
    <id>9</id>
    <date>2013-10-29 23:23:45</date>
    <type> gas leak</type>
    <dimension>Wide</dimension>
    <patientsNumber>none</patientsNumber>
    <colour>Green</colour>
    <description>30 Km distant</description>
    <patientsState>alive</patientsState>
    <street>VIA DI CAREGGI</street>
    <streetNumber>1</streetNumber>
    <building>001</building>
    <annex>I Obstetrics </annex>
    <floor>1</floor>
    <department>Maternity</department>
    <room>DEG1</room>
  </emergency>
  <emergency>
    <id>1</id>
    <date>2013-10-29 09:41:17</date>
    <type>fire</type>
    <dimension>contained</dimension>
    <patientsNumber>10</patientsNumber>
    <colour>yello</colour>
    <description>under control, needed
    intervention</description>
    <patientsState>stable</patientsState>
    <street>VIA dei Santi Benedetti</street>
    <streetNumber>76</streetNumber>
    <building>001</building>
    <annex>I General Anatomy</annex>
    <floor>1</floor>
    <department>Surgery</department>
    <room>AMB1</room>
  </emergency>
</emergencies>
  
```

```

</emergency>
</emergencies>

```

The **Media acquisition and delivering** module is exploited by the Event Manager when images and/or video regarding events/emergency have to be sent to the CS to enrich the event understanding.

The **Collaboration Discovering** module allows establishing direct communication with neighbourhoods' colleagues that are connected with the same network, to exploit the device connection to the local wireless network, and P2P solution. Thus, the operators inside a building are facilitated to communicate each other in order to cooperate and coordinate the actions. The operators may exchange messages both in broadcast or private mode. During an emergency situation, operators inside a building may need to communicate with each other in order to cooperate and coordinate the rescue actions. For this purpose, we introduced in Mobile Emergency Pro a functionality called Discovery Mode that allows to find all the operators inside the building, exploiting the device connection to the local wireless network. The Game Kit framework provides classes to create an ad-hoc wireless network among devices. Exploiting the P2P Connectivity 2 important functionalities have been designed and implemented:

- **Discovery:** the user can discover all the other operators inside the building and know their last position reported.
- **Instant Messaging:** the user can communicate with other operators connected by exchanging text messages both in broadcast or private mode.

In Figure 3, the classes modeling the collaboration and discovering aspects are reported. They implement the functionalities described according to the Model View control patter of the programming model for iOS applications. Relevant classes are those related to DiscoveryView and to BroadcastChatView that are used for managing P2P discovering phases and broadcast messaging, respectively.

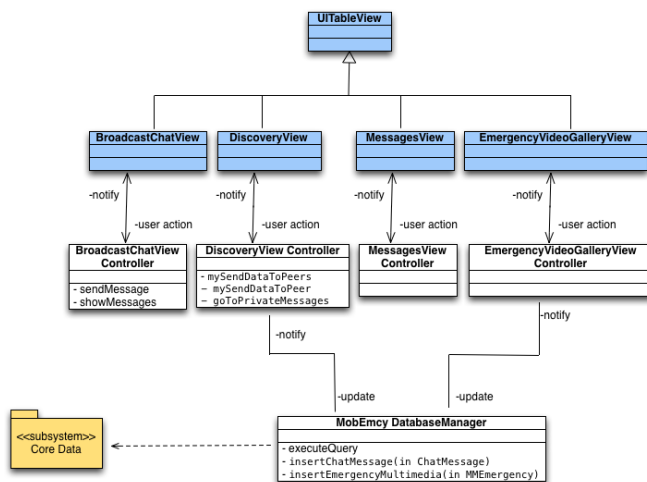


Figure 3 - UML Diagram regarding collaboration discovering aspects

The **QR acquisition** allows getting position corresponding to programmed QR codes placed in several positions in the infrastructure (door, cabinet, etc.). QR codes are very cheap and

can be printed aside of each wall map and hot point at low cost. The QR codes have been coded by using 30% of redundancy, and the string is defined as: <serverURL>ID<PositionID><checkdigit>. The user can grab QR code with the camera to take the position by using the Mobile Emergency App. Once taken the QR string the device creates a connection to the CS taking information about the corresponding position, which is buffered into the mobile device to avoid information preload. If the User is registered and authenticated, the CS implicitly deduces the position/location of the person. The access to QR URL by means of the Mobile Emergency App. implies the access to additional information used by the internal navigation system: building code; department code; currently updated image URL of the map; room code; spatial coordinates of the QR position on the map; spatial coordinates of the nearest exits; spatial coordinates of the nearest collecting areas, the list of point of interests in the taken map. The Mobile Emergency App exploits the additional information together with the maps downloaded from the server to allow the users to navigate in both indoor and outdoor in an integrated manner, to reach the nearest exit, the collecting area and/or any specific position as described in the following. On the other hand, the coding of a location code in a QR as an URL allowed the Central Station server to provide different information according to different QR readers, and user agents. Thus, if the QR URL is not called by the Mobile Emergency App, a simple map with the current position, exits and collecting areas is provided. This allows any user to exploit the information associated with QR place in the hospital by any simple QR based applications, even if with limited capabilities – e.g., no navigation, no emergency status, no networking, no communication, etc.



Figure 4 - (a) the Indoor Path Manager displays the path suggested to reach an emergency exit. (b) the Outdoor Path Manager shows the suggested path to reach an outdoor collecting area.

Once the maps are obtained or preloaded, the **Map Consulting** module may be used to visualize them. The user can visualize the building GPS position by consulting a geographical map and can also consult all its floor maps, passing from outdoor to indoor consultation. On each map, the



user can see the position of all the POI (e.g., stairs, fire extinguishers, medical kits, tech. cabinets, exits, ...).

In the **Mobile Emergency Pro** tool, the **Path Manager** module allows to visualize the path necessary to reach a chosen destination (exits, collecting areas, stairs, other operators) from the current position. This module is divided in two sub-modules and algorithms: the Indoor Path Manager and the Outdoor Path Manager. The **Indoor Path Manager** is responsible for guiding the user within a building. It exploits the floor map with information about the user and the destination positions. In order to help user to easily reach the destinations, the Indoor Path Manager exploits the **Indoor Navigation System** (see Figure 4a) that performs reasoning about the map information/descriptor, taking into account the position and the movements of the users. It estimates the user current position by using the sensors of the mobile device to perform adjustments with respect to the position set using a QR code. This module is used, for example, to reach an emergency exit, a collecting area or another medical operator detected with the discovery mode. On the other hand, the **Outdoor Path Manager** (see Figure 4b) guides the user towards a Point of Interest or buildings located outside a building displaying the path required to reach the destination on an outdoor iOS Map. This sub-module determines the path on the basis of the user position, obtained from the smartphone GPS sensor, and the knowledge of the POI geographical coordinates. At every movement, the system updates the user position on the map.

The **Reach Location** module provides support to teams helping them to reach the event/emergency location, especially even when they do not know the location details and how to reach it. The adopted strategy takes into account available information about the emergency status, maps, possible paths, and takes into account indoor and outdoor paths. To this purpose, the solution is based on an algorithm that work on the internal data base of the mobile phones and when needed ask for additional information to the central station.

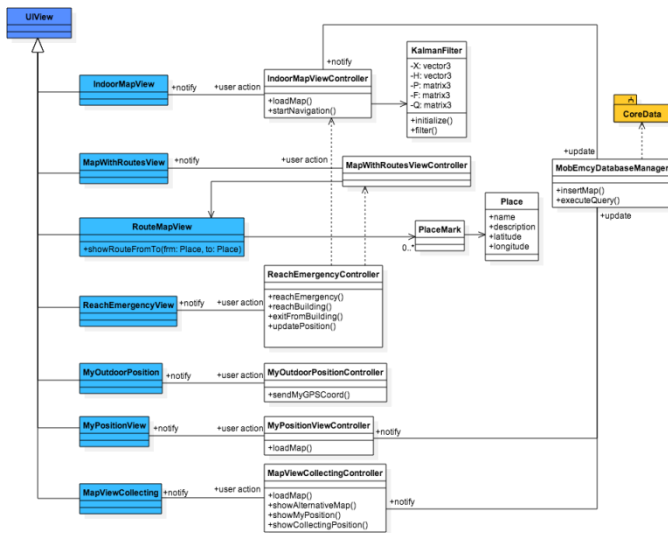


Figure 5 - UML Diagram regarding integrated navigation aspects.

As regards **software model for the navigation** aspects: indoor/outdoor, path manager, indoor navigation system, they have been mapped on several classes into the Mobile Emergency Pro applications. The programming paradigm of the Apple iOS constrains to map functional aspects connected to the views according to the Model-View-Control pattern. To this end, in Figure 5, the subset of classes connected to the navigation aspect of the applications are depicted. Among them, class IndoorMapView is the class that manages the indoor navigation activating the Kalman filtering (and former algorithms). Other interesting classes are: MapViewCollecting that allow to load maps and show positions in browsing modality, MyOutdoorPosition that permits to exploit the outdoor navigators connected with the operating system application and to collected the GPS coordinates, ReachMergencyView that enforce the algorithm for the strategies to reach the buildings, etc. These classes are also connected to the QR code acquisition when needed. More details about the algorithms for navigation are reported in the following section.

#### IV. INDOOR/OUTDOOR NAVIGATION

On the main purpose of Mobile Emergency Pro is to provide support to (i) teams to get details about reaching the event location, (ii) involved personnel in getting the closest and updated exit, (iii) registered users in reaching POI. In order to support the users reaching the target locations, an integrated indoor/outdoor navigation system support is needed. To this end, the teams/users to be moved can be located indoor/outdoor and may receive info to reach a different location in the minimum time. In the case of large infrastructures (industries, parks, hospitals) dedicated solutions are needed to provide updated and integrated indoor/outdoor information with all the POI, exits, cabinets, collecting areas, positions of colleagues, etc. For these reasons, the integrated indoor/outdoor navigation system has been designed to work with the CS.

##### IV.A State of the Art of Indoor Navigation Systems

In the literature, many indoor navigation systems have been proposed [8], [9]. In [10], a method of personal positioning for a wearable Augmented Reality System based on dead reckoning has been proposed. In that system, the user is equipped with a communication device endowed of built-in sensors, a wearable camera, and an inertial head tracker and display. Other solutions adopted triangulation of Wi-Fi hotspots or other source of signals (or laser light). In [11], a pedestrian location system has been proposed by combining a foot-mounted inertial unit and Wi-Fi, thus achieving a location accuracy of 0.73m in the 95% of cases. Alternative solutions have proposed the combination of inertial solutions based on accelerometers and RFID tags [12]. Most of these solutions are not viable in large and complex infrastructures since they may require to: (i) distribute specific devices in the area to mark the zone; or (ii) navigate by carrying on and using specific expensive devices; or (iii) combine both cases. In large public infrastructures, the solution has to be light and cheap enough to be placed in the hands of all personnel, which should carry them every day. In [13], a simple prototype for indoor navigation based on QR codes indicating

direction for impaired has been presented. QR code can be grabbed by a mobile phone camera, and the information is provided via Wi-Fi. A very primitive navigation system has been also provided, allowing to record the person's ID, time stamping the visited position, elapsed time after leaving the last position and expected arrival time to the next position, in order to activate the support team or family members in case of anomalies. In [14] an infrastructure-less solution for navigation in extreme environment (indoor and climbing) has been proposed. This solution has been based on an electronic compass, and for the indoor navigation case "You-Are-Here" (YAH-map) maps have been adopted to take the initial references. The successive positions are estimated by assuming a stable and regular walk step for the human. A low-cost indoor navigation system running on off-the shelf camera phones is presented in [14]. In this case, the proposed system uses smartphone cameras to determine user location by detecting fiduciary markers, called SignPost, which provide location information. The scanning of SignPosts has to be performed by the person with the phone camera (for example 37 markers in an area of 100x200 meters). The required infrastructure is limited to paper markers (square markers or frame markers) and static digital maps. Moreover, the movements from one SignPost to the next is not supported, they are just markers to be discovered. In [16], a solution for indoor pedestrian navigation based on QR and accelerometers, compass, camera and Internet connectivity of modern smartphones has been proposed. The user gets the initial position and map information from a QR code on a YAH-map. The system deduces the initial position of the user and his orientation on the basis of the distance and the angle from the scanned 2D datamatrix. Then, the system estimates the movement by calculating the number of user's steps from the starting point using the accelerometers and the direction using the compass, and accumulating large errors due to the assumption about the regular step length, in order of 4% after 40 steps.

#### IV.B Integrated Indoor/Outdoor Navigation

The integration of indoor/outdoor with high precision is a complex problem to be solved [7]. A detector of in/out passages can exploit almost all accessible sensors of the mobile device, among them: GPS accuracy, presence of WiFi, cell connection power, magnetic field, etc. When the navigation system starts as a first step, it has to understand if the device is indoor or outdoor.

In Mobile Emergency Pro solution, the user may set the target point to be reached: (i) by receiving a push message from the CS for an intervention, (ii) searching and browsing for the exits, collecting areas, position of colleague, POI. The identified target position can be near or far from the current position; it can be in another building, or at a different floor. Thus the navigation system decomposes the problem in segments of indoor and/or outdoor navigation subtasks taking into account the position of the best doors to enter/exit in/out of the buildings, and thus changing the navigation modality.

When the mobile is outdoor, the Outdoor Path Manager exploits and displays the required path to reach the target point. For example: reaching a building at precise accessible door. In this case, the navigation system is provided by the iOS map and

navigation support is activated with the right parameters on the basis of the closest path to enter in the building from the right door. In the case of ground floor, the passing from outdoor to indoor or viceversa has to lead at the activation of the indoor/outdoor according to the knowledge of the GPS position of the building entrance, which are coded and associated with the map information, as stored into the Mobile Emergency database.

In Mobile Emergency Pro, the in/out detector takes a decision on the basis of the: GPS accuracy, previous position (if any), and in the case of uncertainty ask to the user to get the point acquiring a QR code. Once the position is identified, the corresponding map is shown to the user (taking from the cache or downloaded) and the user may navigate in the context. The distance between the last GPS position (that should be just out of the door) and the QR coded position should be coherent. In any case, the QR coded position is taken as a reference, but a confirmation may be requested to the user in some cases. If the building is correct, and floor has to be changed the system provides direction to the closest stairs showing them on the map. Once reached the correct floor, the navigator displays a map showing the current position of the user and the location of the target position. In both cases, the Indoor Path Manager helps the user to easily get closer and reach the destination.

#### IV.C Inertial Integrated Indoor Navigation

Within indoor environments, the system cannot obtain updated GPS information with the needed accuracy. Therefore, the proposed solution computes the current position by taking into account of: (i) the information acquired from the last taken QR information (the map and the position of the map with respect to the coordinates, position of the QR in the map, building, floor, etc.), (ii) the movements considering the device sensors such as gyroscopes, magnetic compass and accelerometers as an Inertial Navigation System.

In [3], the current position was calculated in relation to the last position (QR) by using an improved dead reckoning algorithm with respect to that proposed by [10]. The idea was to exploit the signals extracted from the smartphone sensors to construct a pedometer: a tool for measuring / counting the steps performed by a human, and associating them to the corresponding direction. Among the signals that can be used to analyse the walking, those coming from the accelerometer are the most relevant.

The Mobile Emergency Pro has been developed on iOS platform. The available sensors on iPhone allow estimating the acceleration, the magnetometer and the gyroscope. All these sensors take the measures along three axes, and thus once the measure is performed, it is possible to perform the inverse of the rotation to get a measure independent on the device orientation. On the other hand, the measures obtained are affected by relevant errors. To make the usage of these data possible, a strong reduction of the noise is needed, and thus the Kalman filtering has been adopted [17]. The Kalman filter uses state space models which relate inputs, outputs and state variables by first order differential equations. A Kalman filter performs the prediction of state  $x \in \mathbb{R}^n$ , given the observation

(measure)  $z \in \mathbb{R}^m$ . Both the state and the observation are modelled as:

$$\begin{aligned} x_k &= Ax_{k-1} + Bu_{k-1} + w_{k-1} \\ z_k &= Hx_k + v_k \end{aligned}$$

Where  $w_k$  and  $v_k$  model the noise of process and that of measure, they are assumed to be independent, white and normal probability distribution of noise:

$$p(w_k) \sim N(0, Q_k), \quad p(v_k) \sim N(0, R_k),$$

Where  $R_k$  and  $Q_k$  are the covariance matrices. Low covariance of  $Q_k$  and  $R_k$  will give rise to high trust, with growing covariances the uncertainty in the equations will also grow. The extreme case  $Q_k \rightarrow 0$  would imply complete trust in the previous state  $x_{k-1}$  while the case  $Q_k \rightarrow \infty$  implies no trust at all. The variable  $u_k$  is an optional  $m$  dimensioned control input and is related to  $x_k$  by the  $n \times m$  matrix  $B$ . The filter process alternates between these two steps: predict the future state at the *time update* and adjust the predicted state in the *measurement update*.

Predict:

- state estimate:  $\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1}$
- estimate covariance:  $P_{k|k-1} = AP_{k-1|k-1}A^T + Q_k$

Update (a posteriori):

- gain:  $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R_k)^{-1}$
- state:  $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H\hat{x}_{k|k-1})$
- covariance:  $P_{k|k} = (I - K_kH)P_{k|k-1}$

The a-posteriori estimation of the covariance is a measure of the accuracy about state estimation.

Typically, Kalman filter addresses linear systems. In this case, there are parts of the model which introduce non-linear features: the time between the steps is not constant, the trigonometric functions sine and cosine used to compute the position. The solution was to use an Extended Kalman filter which uses a linearized non-linear model. The linearization approach has led us to make some small changes into the above presented equations

In practice, two Kalman filters have been applied for the accelerations along the  $x$  and  $y$  axes. The state vector is represented by position, velocity and acceleration assuming that in each time interval (between two measures) the motion is uniformly accelerated. Thus the acceleration between two consecutive instant is constant. Thus the state transition matrix  $A$  holds (where  $\Delta T$  is the time interval):

$$A = \begin{pmatrix} 0 & \Delta T & \frac{1}{2}\Delta T^2 \\ 0 & 1 & \Delta T \\ 0 & 0 & 1 \end{pmatrix}$$

$Q$  is a 3x3 diagonal matrix with constant values,  $R$  is a float, and the measurement model  $H = (0,0,1)$ .  $R$  has been measured by keeping the mobile in the same position for a while and taking the average of the covariance noise along  $x$  and  $y$ . The obtained value:  $R=0.00064$ .  $Q$  cannot be directly estimated and

thus after a number of trials a value of  $Q_0 = 15$  was set as a compromise.

The results obtained with the described implementation of Kalman filter were not satisfactory due to the high level of noise on the measures. As will be shown in the next section, the experimental results obtained by using dead reckoning and the above described simple and direct implementation of Kalman Filter were quite similar, especially in the cases in which several changes of directions have to be followed.

For the above reasons, in order to improve the solution an Extended Adaptive Kalman Filter has been adopted to cope with non-linearity [18]. Thus, in the second solution proposed, matrix  $Q$  is adaptively estimated by using  $Q$  at the previous time instant and a corrective scale factor (estimated on the basis of the ratio from the innovation covariance and the predicted value). The process was initiated by starting from the value empirically estimated,  $Q_0$ . In some cases, the adaptively estimated value of  $Q$  may become zero, and thus, in order to avoid singularity in the global estimation, when this happen,  $Q$  is reset to  $Q_0$ . That is the value that we early estimated. This approach improved the quality of the path estimation and following as described in the next Section of validation.

The assumption performed about the uniformly accelerated motion is not always true. That is due to the fact that the acceleration between two consecutive instant is not constant as happen when the device is not moving. In these cases:  $v = v_0 + at$ , therefore, in the presence of relevant noise on measures, the velocity is neither zero nor constant and thus  $v$  tends to increase even when  $a$  is zero or noisy around zero. This problems has been avoided constraining the velocity to zero if the estimated velocity is lower than 0.76. This decreases the chance to see the continuous moving cursors when the user stopped.

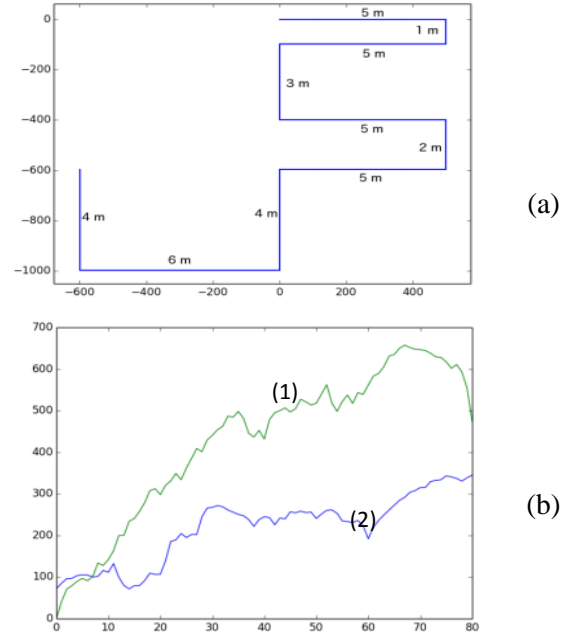


Figure 6 - (a) the indoor path for test (starting from the entrance and exploring some of the rooms, thus multiple changes of directions), (b) trend of the mean error of estimated for navigation with (1) dead reckoning, (2) Kalman filtering. Error is in mm and the path length was of 40 mt.

## V. EXPERIMENTAL RESULTS

The experimental results have been carried out to the global assessment of the advantages of using the Mobile Emergency App in the context of emergency in hospitals. The general indoor/outdoor navigation has been assessed and the results reported in [5]. The general advantage of adopting Mobile Emergency has been the reduction of time needed to reach the target point for the rescue team of the 18%. In addition, in this paper, we focused about the improvement related to the integrated indoor/outdoor navigation system, and in particular on the new algorithms for the inertial indoor navigation. In [5], the indoor navigation was based on an enhanced dead reckoning algorithm as described in Section IV.C.

For this reason, the first step has been to compare the results of the former algorithm of dead reckoning with the first implementation of Kalman filtering with Q constant. The results are reported in Figure 6, where the typical trend path is reported together with the trend of the mean error (estimated on the basis of 8 indoor navigations) on the same path of Figure 6a. In Figure 7, some navigation screens taken during the navigation are depicted. The results depicted in Figure 6b for Kalman (2) show a relevant improvement with respect to the basis solution of dead reckoning [5].



Figure 7 - In the order: indoor navigation to reach an exit, outdoor navigation segment, indoor navigation to reach a colleague.

According to Section IV.C, a second algorithm has been proposed, thus obtaining a further improvement by adopting the Adaptive Extended Kalman filter. In Figure 8, the comparison of the trend of mean error for Extended Kalman and the Adaptive Extended Kalman Filer are reported. Therefore, the final error is lower than 20 cm at the end of the path.

All the experiments and measures have been obtained by sampling the curve and keeping aligned the time code for measuring the data coming from the internal sensors with respect the position of the person passing by the marked points. The measures have been recorded with a TV camera to allow review and verification according to the time code. Due to the high number of samples taken for the measures, and the time code in ms, the error of measure was smaller than 2 cm. Therefore, proposed solutions are better ranked with respect to the state of the art solutions reviewed in Section IV.A.

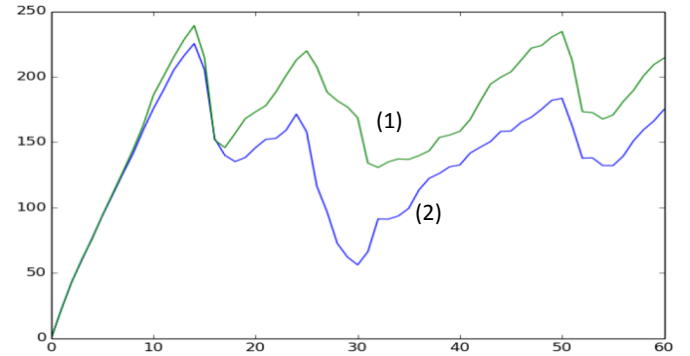


Figure 8 - Trend of the mean error of estimated positions during navigation for (1) Extended Kalman, (2) Adaptive Extended Kalman filtering. Errors is in mm.

## VI. DISCUSSIONS AND CONCLUSION

Wide infrastructures as industry plants and hospitals are vulnerable to causality events. In this paper, we presented a solution addressing the guiding personnel during maintenance and/or emergency conditions. The aim is to reduce the time needed to react and to cope with organization and maintenance support, while facilitating communication, and indoor / outdoor navigation. The solution is based on the formalization of protocol, the modelling of knowledge for navigation, the algorithms and the development of a mobile application and corresponding server device for integrated indoor/outdoor navigation. The introduction of the integrated indoor/outdoor navigation has constrained to a major restructuring of the software architecture of the application. The details are reported in the paper. The navigation algorithms are based on low costs mobile sensors and Adaptive Extended Kalman Filter. The solution has been validated within a large medical infrastructure, thus demonstrating the validity of the identified modalities and procedures, measuring the advantage from both qualitative and quantitative aspects. The indoor navigation solution has been compared with other former solutions based on classical Kalman and dead reckoning (the former solution). The proposed solution for indoor navigation resulted to be better ranked with respect to the solution reviewed at the state of the art.



The present solution presents some intelligence aspects in the navigation systems (on client side) and may could integrate some intelligence into the Central Station to select the personnel, provide suggestions, arrange the groups for the interventions. Both these aspects could be subjected to a further addition of some intelligence and may of some Slow Intelligence solutions [19]. The selection of the best teams on the basis of their profiles, position, and intervention can be a complex task. In more details, the slow intelligence approach could be adopted for managing the status evolution of the several groups and single persons involved into the emergency or maintenance conditions.

#### ACKNOWLEDGMENTS

The authors would like to thank the colleagues of the Medical area and from the Maxi Emergency Group of University of Florence, and to Leonardo Sequi that performed early experiment on former algorithms.

#### REFERENCES

- [1] Levy G, Blumberg N, Kreiss Y, Ash N, Merin O., "Application of information technology within a field hospital deployment following the January 2010 Haiti earthquake disaster", *Journal Am Med Inform Assoc*. 2010 Nov 1;17(6):626-30.
- [2] José H. Canós, Gustavo Alonso, Javier Jaén, "A Multimedia Approach to the Efficient Implementation and Use of Emergency Plans", *IEEE Multimedia*, Vol.11, N.3, pp.106-110, 2004.
- [3] Bellini P., I. Bruno, D. Cenni, A. Fuzier, P. Nesi, M. Paolucci, "Mobile Medicine: Semantic Computing Management for Health Care Applications on Desktop and Mobile Devices", *Multimedia Tools and Applications*, Springer, Vol.58, n.1, pp.41-79, May 2012.
- [4] Protogerakis, M.; Gramatke, A.; Henning, K., "A System Architecture for a Telematic Support System in Emergency Medical Services", *ICBBE 2009. 3rd International Conference on Bioinformatics and Biomedical Engineering*, 2009, 11-13 June 2009 Page(s):1-4.
- [5] P. Bellini, S. Boncinelli, F. Grossi, M. Mangini, P. Nesi, L. Sequi, "Mobile Emergency: supporting emergency in hospital with mobile devices", *Theme Issue Media Tablets & Apps* (Guest editors: Pincirolì & Pagliari), *JMIR RESEARCH PROTOCOLS*, 2013
- [6] Hui Li; Xiangyang Gong, "An approach to integrate outdoor and indoor maps for books navigation on the intelligent mobile device," *Communication Software and Networks (ICCSN)*, 2011 *IEEE 3rd International Conference on*, vol., no., pp.460,465, 27-29 May 2011.
- [7] Pengfei Zhou, Yuanqing Zheng, Zhenjiang Li, Mo Li, and Guobin Shen. 2012. IODetector: a generic service for indoor outdoor detection. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems (SenSys '12)*. ACM, New York, NY, USA, 113-126.
- [8] G.N. DeSouza and A.C. Kak, "Vision for Mobile Robot Navigation: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Feb. 2002, pp. 237-267.
- [9] Schindhelm, C.K.; Gschwandtner, F.; Banholzer, M., "Usability of apple iPhones for inertial navigation systems," *Personal Indoor and Mobile Radio Communications (PIMRC)*, 2011 *IEEE 22nd International Symposium on*, vol., no., pp.1254,1258, 11-14 Sept. 2011
- [10] Kougiori, M. and Kurata, T. 2003. Personal Positioning based on walking locomotion Analysis with self-contained sensor and a wearable camera. In *Proceedings of ISMAR2003*, pp. 103-112.
- [11] Woodman, O. and Harle, R. 2008. Pedestrian Localisation for Indoor Environments. In *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp)*, Seoul, Korea, ACM 2008, 114-12
- [12] K. Okuda, S. yuan Yeh, C. in Wu, K. hao Chang, and H. hua Chu. The GETA Sandals: A Footprint Location Tracking System. In *Workshop on Location- and Context-Awareness (LoCa 2005)*, also published as LNCS 3479: *Location- and Context-Awareness*, pages 120-131. Springer, 2005.
- [13] Chang Y.J., Tsai S. K., Wang T. Y and Chou L. D. A novel way finding system based on geo-coded QR codes for individuals with cognitive impairments. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, Arizona, USA, 2007, 231 – 232.
- [14] Gehring S., Löchtefeld M., Schöning J. and Krüger A. Exploring the Usage of an Electronic Compass for Human Navigation in Extreme Environments. *Haptimap 2010: Multimodal Location Based Techniques for Extreme Navigation*, In conjunction with *Pervasive 2010*, Helsinki, Finland, 2010.
- [15] Mulloni A., Wagner D., Barakonyi I., Schmalstieg D., *Indoor Positioning and Navigation with Camera Phones*, *IEEE Pervasive Computing*, vol. 8, no. 2, pp. 22-31, Apr.-June 2009
- [16] Serra A., Dessi T., Carboni D., Popescu V., Atzori L., *Inertial Navigation Systems for User – Centric Indoor Applications*, in *Proceedings of NEM Summit – Towards Future Media Internet Barcelona*, Spain 2010.
- [17] Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems". *Journal of Basic Engineering* 82 (1): 35–45.
- [18] A. Almagbile, J. Wang, and W. Ding Evaluating the Performances of Adaptive Kalman Filter Methods in GPS/INS Integration, *ournal of Global Positioning Systems* (2010)
- [19] Shi-Kuo Chang, "A General Framework for Slow Intelligence Systems", *International Journal of Software Engineering and Knowledge Engineering*, Volume 20, Number 1, February 2010, 1-16.



# A System Design for Surveillance Systems Protecting Critical Infrastructures

Erland Jungert<sup>1</sup>, Niklas Hallberg<sup>1,2</sup>, Niclas Wadströmer<sup>1</sup>

<sup>1</sup>FOI (Swedish Defence Research Agency)  
Box 1165, SE-581 11 Linköping, Sweden  
{jungert, nikha, niclas.wadstromer}@foi.se

<sup>2</sup>School of Computer Science and Communication  
KTH Royal Institute of Technology  
SE-100 44 Stockholm, Sweden

**Abstract**---Critical infrastructures are attractive targets for attacks by intruders with different hostile aims. Modern information and sensor technology provides abilities to detect such attacks. The objective of this work is to outline a system design for surveillance systems aimed at protection of critical infrastructures, with the focus on early threat detection at the perimeters of critical facilities. The outline of the system design is based on an assessment of stakeholder needs. The needs were identified from interviews with domain experts and system operators. The system design of the surveillance system and the user requirements in terms of capabilities were then determined. The result consists of the systems design for surveillance systems, comprising the systems capabilities, the systems structure, and the systems process. The outcome of the work will have an impact on the implementation of the surveillance systems with respect to the sensors utilized, the sensor data algorithms and the fusion techniques.

**Keywords:** security systems, surveillance systems, critical infrastructure, user requirements,

## I. INTRODUCTION

In recent times, the risk for critical infrastructures to be subject to attacks from various groups of terrorists or criminals has become increasingly high and therefore they must be protected. To accomplish sufficient surveillance, modern information technology could be used. Such surveillance systems need to be based on modern sensors and sensor systems with advanced sensor-data analysis and data fusion. However, to accomplish systems of high quality, they must be based on the stakeholders needs, so that needed capabilities can be provided that support the system operators in their work to handle upcoming events and incidents enforced to the facilities to be protected from attacks [1]. Hence, to accomplish such surveillance systems it is essential to put a sufficient amount of resources on the early stages of the development, which is to identify the stakeholder needs and to define the users requirements in terms of system capabilities. To enhance the realization of such systems they should be based on an adequate system design. Thereby, the probability to get useful systems that provide the means to support

handling of incidents and crisis management will increase and help to avoid or at least minimize the consequences of attacks on critical infrastructure facilities. The approach taken for determination of the system design is based on the assessment of stakeholder needs through a series of interviews with a number of especially appointed domain experts and security personnel.

This work has been carried out as a part of the EU project, *The Privacy Preserving Perimeter Protection Project (P5)*. The objective of the project is to demonstrate an intelligent perimeter surveillance system that will operate in all weather and light conditions and with privacy preserving properties. The system will monitor a part of the area just outside the boundary of critical infrastructure facilities and, thereby, provide early warnings to detected terrestrial and airborne threats.

The system should have a low false alarm rate, e.g. due to animals and other innocuous events, combined with high level of threat detection sensitivity and privacy standards will be central ambitions of the project.

The objective of the work described here is to outline a systems design for surveillance systems aimed at protection of critical infrastructures. In particular, the surveillance system should be able to support the security staff at the facility to respond to attacks from intruders at an early stage and thus the protection and surveillance of the perimeter of the facilities will be in focus to make it possible to give early warnings of attacks. Thus the overall objective of this part of the project is *to identify a systems design including user requirements in terms of capabilities*. The system should thus be able to warn for threats carried out by different types of objects, (persons or vehicles etc.). Eventually, these capabilities of warnings should be realized by state of the art sensor solutions. To determine what sensors to be used and the methods for sensor-data analysis and fusion is, however, outside the scope of this part of the P5 project.

## II. PHYSICAL CONTEXT

The physical context in which surveillance systems of critical infrastructures operate is varying from facility to facility; especially with respect to the perimeter, which also differ with respect to the type of infrastructure that should be protected. Generally, the critical infrastructure facility can be described as containing a central complex, i.e. the surveilled area of the critical infrastructure with one or several buildings or installations as illustrated in Fig.1. The area surrounding the surveilled area makes up the perimeter of the facility that may differ in width with respect to its extension. Thus, the perimeter can be defined as illustrated in Fig. 1 where the perimeter is made up by the Restricted area, the Facility boundary, a strip of the outside area, and the airspace above the facility. In some cases, there is no restricted area and the boundary of the facility coincides with the boundary of the Surveilled area. The outside strip may also vary from facility to facility depending on its context. Further, the terrain type at different facilities, differ as well and the surveillance system must be able to adapt to such differences.

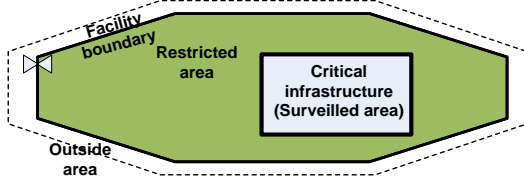


Figure 1. An illustration of an extended perimeter surrounding a critical infrastructure facility.

## III. METHODS

The work was carried out as two main activities: a needs assessment activity and an outline of the design of the surveillance systems.

### A. Needs assessment

The needs assessment was performed in six steps. The initial step was to determine who the stakeholders are, which of them should be given the opportunity to influence the development of the system, and how their statements should be collected. The work was carried out during a workshop, involving the project management, in which different categories of stakeholders were identified, such as systems operators, business managers, and security managers. Thus, it was decided which categories of stakeholders that should be provided the opportunity to influence the design of the system. The second step was to capture the stakeholders' statements. The respondents were selected as good representatives of the selected stakeholder categories. Interview questions that focus on the specific problems subject to the studies were developed. Each interview was carried out by two persons; one that asked the questions and another

responsible for recording the answers through note taking. The third step of the needs assessment was to interpret the collected statements to determine the actual needs. When asking stakeholders about what needs they have, they will use descriptions of, e.g., problematic situations that they have experienced and technical solutions that they believe can be useful to them [2]. The Voice of the customer table (VCT) was used for analyzing statements to reveal the actual needs [1][3]. The outcome from this step was a large set of unstructured and unsorted needs.

The fourth step was to thoroughly analyse the identified needs, to unify the formulation of the needs and, thereby, identify and discard duplicates of needs. Further, the analysis also included to determine if any needs had been left out, and if appropriate add the missing ones. To accomplish this and due to the amount of needs it is likely necessary to categorize the needs. This step was performed by using affinity diagrams and hierarchy diagrams [3]. The fifth step is to validate the needs. This was carried out in a workshop with stakeholder representatives. During the workshop the identified needs were presented to the stakeholders' representatives; based on their comments inaccuracies were corrected. The sixth step was to prioritize the needs in terms of stakeholder value. During a workshop, the stakeholders' representatives were asked to prioritize, on a scale from 0 to 3, the identified needs based on how important they considered the needs to be; taking into account the scope of the project. In this case 3 means *the highest importance* and 0 *not important*. The highest prioritized needs, 3 and 2, were established as the needs that should constitute the foundation for the determination of the capabilities.

### B. Outline of the system design

The outline of the system design was carried out in three steps as definition of (1) the systems capabilities, (2) the systems structure, and (3) the systems process. The initial step, to define the systems capabilities was based on the needs with the highest priorities. The capabilities were compiled and elaborated to a coherent set of capabilities, documented in a hierarchical diagram [3]. The validation of the capabilities was performed during several workshops where the developers present the capabilities to the stakeholder representatives. The provided comments and suggestions are analyzed and appropriately incorporated in the set of capabilities. The second step was to define the structure of the systems, i.e., systems components that together provide the defined capabilities. The third step was to define systems process. That is

outlining how the systems accomplish the capabilities, as activity diagrams [3].

After the first version of the system design was completed, validation of the design was performed with stakeholder representatives and technical experts. The latter were experts on surveillance techniques.

#### IV. SYSTEM CAPABILITIES

The outcome of this study is a system design for surveillance systems for protection of critical infrastructures. The design is based on the systems capabilities, the systems structure, and its process.

##### A. Outline of the system design

The systems capability structure is hierarchical and will thus include sets of subordinate capabilities and eventual a number of leaf capabilities. Leaf capabilities with attachments to the *Sensor data analysis and fusion* modules are either of the type *Track* (an entity of an event), *Detect* and *Watch* (entities). *Detect* and *Watch* are concerned with the detection of objects deviating from normal inside the boundary of the facility respectively outside whereas *Track* is concerned with tracking of the deviating objects. *Watch* is similar to *Detect* but since different legislation is applied to the outside area (in most countries), sensor types that directly can identify individuals are not allowed on the outside due to privacy considerations that must be followed and for this reason *Watch* is introduced.

The general interpretation of the capabilities is that a certain capability is not enabled until all its subordinate capabilities have been enabled or when a leaf node has been supplied with relevant event related information or in some other way terminated.

The ultimate capability for the protection and the surveillance of critical infrastructures in this work is *Handle facility protection*. This capability has four subordinate capabilities *Handle surveillance*, *Handle deviation*, *Handle incident* and *Handle user interactions* (Fig. 2).

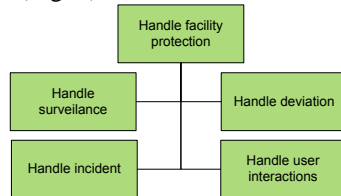


Figure 2. The two top levels of the capability structure,

The capability *Handle surveillance* is concerned with surveillance of the perimeter of the facility, which on the next lower level is utilized by four capabilities: *Surveil Restricted area*, *Surveil Boundary fence*, *Surveil Airspace*, and *Watch Outside area* (Fig. 3). These capabilities are aimed at the surveillance of the facility perimeter. Furthermore, the capability *Handle surveillance* relates to aspects like: the system

should be able to operate e.g. without interruption, during all weather conditions and with respect to privacy considerations.

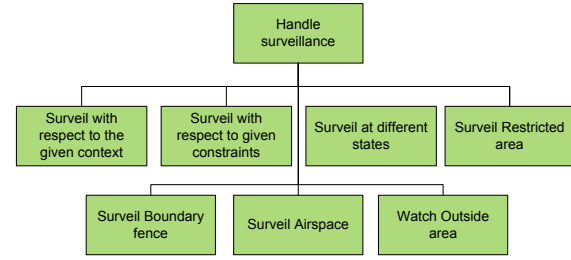


Figure 3. The *Handle surveillance* branch of the capability structure.

An illustration to the deeper levels of the four perimeter surveillance oriented capabilities in Fig. 3 can be seen in Fig. 4, which shows how they are divided further into surveillance of the land and the sea areas and on the lowest level the detection of various types of events requested to release an alarm; this is indicated with either one of the alarm types: verified alarm (VA) or unverified alarm (UA) (section IV D). Thus, a *Detect* capability for protection of the Restricted area is formulated as *Detect person behaving in a deviating way on land of restricted area (UA)*, which indicates that the capability determined to detect a person behaving in a deviating way and that the alarm in this case is unverified and consequently needs to be verified in some way, e.g., guards can be sent out by the operator to determine what kind of event that has occurred.

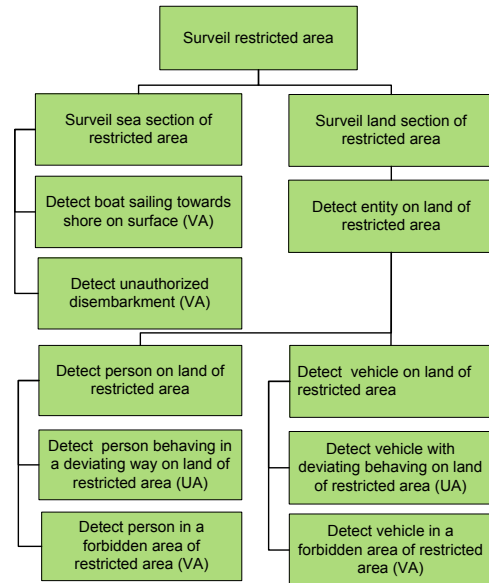


Figure 4. The capabilities for surveillance of the facility perimeter.

The second capability, on the second level, is the *Handle deviation* that has to be enabled once a deviation has been determined (Fig. 5). These capabilities are concerned with the handling of the type of alarm that has been released but also to the analysis of the occurred deviation type that eventually should be reported.

The third capability, *Handle incident* (Fig. 6), is enabled when an incident has been determined, by a verified alarm, and is concerned with the monitoring of the incident, that is, to track incident related objects, to gather, analyse and store various kinds of incident related information but also to request orders from and to forward information to participating persons in the crisis management organization. Eventually, the incident will be terminated.

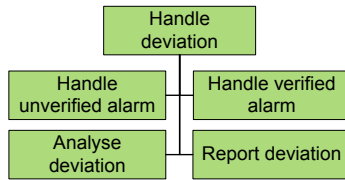


Figure 5. The main part of the *Handle deviation* capability structure.

The fourth and final capability at the top level, i.e., *Handle user activities and support* (Fig. 7). This capability is aimed at directly support the operators to control sensors, basically of visual type, and to gather information from them. Further, information gathered by the surveillance system should also be handled, stored and aggregated in such a way that a situational picture can be built up. The situational picture must also be possible to adapt to the needs of the operator, e.g., by zooming and panning to follow the events going on during an incident. Among other things, the situational picture will include information of deviating objects such as detected deviations, objects locations and tracks and alarm related information.

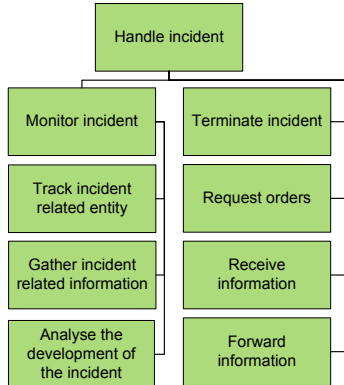


Figure 6. The *Handle incident* capability structure.

## B. System structure

The main components of the system structure are: the *User interface module*, the *Command & control module* and the *Sensor system module* (Fig. 8). Inside

the *Command & control module* there is the *User support module* that handles the surveillance processes and on its upside it supports the user interface; this means that it is serving the operator of the surveillance system. At the down side, the surveillance processes are served by the *Sensor data analysis & fusion module*. The sensor system delivers information about detected objects and tracks from these objects.

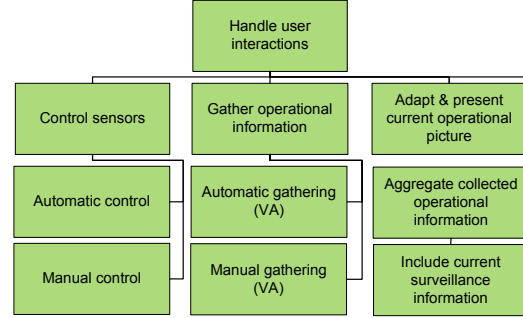


Figure 7. The *Handle user interactions* capabilities.

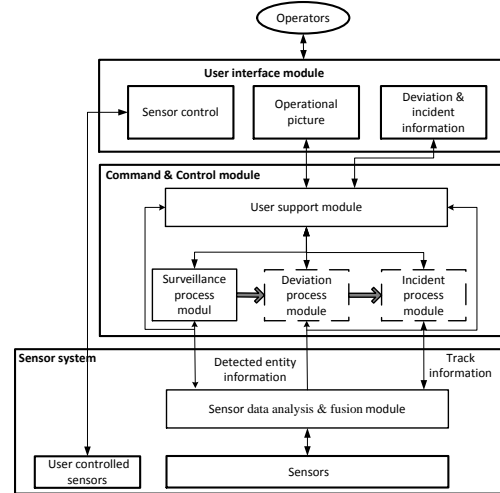


Figure 8. A schematic description of the system structure of the surveillance system.

The users can, besides the user interface, which contains the operational picture with its deviation and incident related information, also control visual sensors outside the sensor system. The purpose of these sensors is to allow the operator to follow and verify what is going on during an event. Thus, some video screens must be available for this purpose.

## C. Surveillance structure

The subordinate capabilities of *Handle facility protection* (Fig. 2) are the most influential capabilities in the surveillance process. The capabilities *Handle surveillance*, *Handle deviation* and *Handle incident* can all be transformed into corresponding processes, i.e., the *Surveillance process*, the *Deviation process* and the *Incident process*. The relationship between these processes

can be seen in Fig. 8. The *Surveillance process* is a process that will be running for as long as the system is running. A *Deviation process* is the cause of an alarm and is also the result of a single registered event (a single instance), and as many events may occur simultaneously, due to some hostile and coordinated activities this leads to the initiation of multiple deviation processes that each in turn may initiate a single *Incident process*. All detected incidents may thus lead to the initiation of an *Incident process* but multiple incident processes may, on a higher level, be part of the same ongoing incident with multiple events or actions taken by a group of intruders. Once any of the events in any of the deviation and incident processes have been solved the corresponding processes must terminate.

#### D. Alarm handling

The alarm handling is essential to surveillance systems, because false alarm rate must be kept low and the type of alarm precise so that the operators know when and how to act in case of serious events. An alarm is either directly verified or unverified. At a verified alarm the cause of the alarm is known with a certainty high enough to initiate a response to the event. The cause of an unverified alarm is uncertain and in such a case the immediate action is to determine its cause.

A verified alarm may turn out to be either false or the result of a failing system component. Both cases must be adjusted promptly. The third case is an incident corresponding to an event that is anything from harmless and up to something serious. In any of these cases the operators must act to keep up the security of the facility. An illustration of the type of alarm that can be determined by the sensor system may include the following information:

*Sub-area type: Land section of Restricted area*  
*Deviation: vehicle observed at forbidden area*  
*Alarm type: VA (verified alarm)*

This will lead to the release of a verified alarm in the Restricted area with respect to the type of sub area and the determined type of deviation (Fig. 4).

#### V. RELATED WORKS

In [4] a system for surveillance of critical infrastructures is described, which concerns early warnings with respect to attacks from e.g. terrorists although in this work no strong efforts for determination of user requirements have been performed. Some other work on surveillance systems design with relationship to the work carried out here is the work by Jungert et al. [5]. In Goodall [6] gathering of user requirements for a visualization system with capabilities for intrusion detection analysis is discussed. Shan, Wang, Li, and Chen [7] present a

comprehensive design for decision support systems within emergency response. Hansson et al. [8] demonstrates the intensions to determine the general context for security systems as a foundation for user and system requirements.

#### VI. CONCLUDING REMARKS

The objective of the work presented is this work was to define a system design of surveillance systems for protection of critical infrastructures, which includes a set of capabilities and a systems structure. The stakeholders' needs have been identified and used to define the users' requirements in terms of capabilities and a systems design in collaboration with the stakeholders. The result of this work will be used as input the other parts of the P5 project, i.e. to complete the systems architecture of the surveillance system with its sensor system and network of selected sensors.

#### ACKNOWLEDGEMENT

This work has been performed as part of the Privacy Preserving Perimeter Protection Project (Grant number 312784), which is a European and FP7 funded project for the protection of critical infrastructures

#### REFERENCES

1. N. Hallberg, T. Timpka, and H. Eriksson, "The Medical Software Quality Deployment Method," *Methods of Information in Medicine*, vol. 38, 1999, pp. 66-73.
2. N. Hallberg, S. Pilemalm, and T. Timpka, "Quality Driven Requirements Engineering for Development of Crisis Management Systems", *Int. J. of Information Systems for Crisis Response and Management (IJISCRAM)*, vol. 4, 2012, pp. 35-52.
3. N. R. Tague, *The quality toolbox*. Milwaukee: ASQ Quality Press, 2005.
4. F. Flammini, A. Gaglione, N. Mazzocca, V. Moscato, and C. Pragliola, "On-line integration and reasoning of multi-sensor data to enhance infrastructure surveillance," *J. of Information Assurance and Security*, vol. 4, 2009, pp. 183-191.
5. E. Jungert, C. Grönwall, N. Hallberg, B. Kylesten, F. Lantz, and L. Eriksson, "A Generic Architecture for Surveillance Systems," *Proc. of the Int. conf. on Distributed Multimedia Systems*, pp. 57-63, Oct 2010.
6. J. R. Goodall, "User requirements and design of a visualization for intrusion detection analysis", *Proc. 2005 Workshop on Information Assurance and Security*, pp. 394-401, June 2005.
7. S. Shan, L. Wang, L. Li, and Y. Chen, "An emergency response decision support system framework for application in e-government," *Information Technology and Management*, vol. 13, 2012, pp. 411-427.
8. J. Hansson, R. Granlund, N. Hallberg, F. Lantz, and E. Jungert, "A reference context module for development of security systems," *Proc. of the Int. conf. on Distributed Multimedia Systems*, pp. 64-69, Aug. 2011.

# Joint Fingerprinting and Encryption in Hybrid Domains for Multimedia Sharing in Social networks

Conghuan Ye, Zenggang Xiong, Yaoming Ding, Guangwei Wang, Jiping Li, Kaibing Zhang  
College of Computer and Information Science  
Hubei Engineering University  
Xiaogan, China  
p2pgrid@gmail.com

**Abstract:** The advent of social networks and cloud computing has made social multimedia sharing in social networks easier and more efficient. The lowered cost of redistribution, however, also invites much motivation for large-scale copyright infringement. It is necessary to safeguard multimedia sharing for security and privacy. In this paper, we propose a novel framework for joint fingerprinting and encryption (JFE) based on Cellular Automata (CA) and social network analysis (SNA) with the purpose of protecting media distribution in social networks. The motivation is to map the hierarchical community structure of social networks into the tree structure of Discrete Wavelet Transform (DWT) for fingerprinting and encryption. Firstly, the fingerprint code is produced using SNA. Secondly, fingerprints are embedded in the DWT domain. Thirdly, CA is used for permutation in the DWT domain. Finally, the image is diffused with XOR operation in the spatial domain. The proposed method, to the best of our knowledge, is the first JFE method using CA and SNA in hybrid domains for security and privacy in social networks. The use of fingerprinting along with encryption can provide a double-layer of protection for media sharing in social networks. Theory analysis and experimental results show the effectiveness of the proposed JFE scheme.

**Keywords:** security and privacy; fingerprinting; multimedia encryption; social multimedia sharing;

## I. INTRODUCTION

The advent of social networks and cloud computing makes multimedia sharing in social networks very easy. Multimedia content may be generated, processed, transmitted, retrieved, consumed or shared in social networks [1]. Content distribution in social networks offer distinctive challenges such as privacy and security issues. In order to decrease the loss of multimedia owners, secure multimedia sharing in social networks is becoming more and more urgent for practical applications. To prevent illegal use in social networks, techniques, such as watermarking (or fingerprinting) and encryption [2] of these media for security and privacy need to be carried out.

Multimedia encryption is one way which may ensure the content security and prevent an unauthorized access. Chaotic maps are employed to generate a permutation table for confusion and a pseudo-random keystream for diffusion [3], therefore, the chaos-based approach is a promising direction for multimedia encryption. A substantial amount of research work

on chaos-based image encryption has been carried out [4-6]. However, these schemes only focus on encrypting. In fact, when the ciphered data is deciphered by the authorized user, it is unprotected, and it is still possible for a legal user to deliver decrypted data to an unauthorized person. The content could be copied and redistributed at their option. There are not ways to continue the work of protecting the multimedia content, therefore the privacy of content owner may be leaked. In this case, extra protection schemes should be adopted to deter content redistribution, therefore, encrypted data need an additional level of protection in order to keep control on them after the decryption phase. Watermarking is another technology to protect copyright further. It enables a distributor to hide additional bits into multimedia content while preserving its quality [7]. The use of watermarking along with encryption can provide a double-layer of protection for multimedia sharing.

There have been some related works on watermarking in the encrypted domain over the past few years. Commutative Encryption and Watermarking (CEW) could be used for providing comprehensive security protection for multimedia content. D. Bouslimi et al. proposed a joint encryption and watermarking algorithm in [8]. The convergence of the two technologies is now facilitating privacy and security studies [9]. An interactive buyer-seller watermarking protocol for invisible watermarking was proposed in [10]. And in [11], the encryption is performed on most significant bit planes while watermarking the rest of lower significant bit planes. Two robust watermarking algorithms were proposed to watermark compressed JPEG Images in encrypted domain [12] and JPEG2000 compressed and encrypted images [13] respectively. However, watermarking can't trace somebody who redistributed the copies. To solve this problem, digital fingerprinting methods have been intensively investigated.

Digital fingerprinting is a technique for identifying users who might try to use multimedia content for unintended purposes [14]. Fingerprint, which is used to identify adversary who leak copies of the content, represents the ID of a user [15]. Although the approach of embedding and extracting fingerprints is similar to that of watermarking, the goals of each method are quite different. Basically, watermarks embedded into multimedia data for enforcing copyrights [16] must uniquely identify the data, but fingerprinting is aimed at traitor

---

Identify applicable sponsor/s here. (sponsors)

tracing. Fingerprinting can further safeguard security and privacy for content sharing in social networks. Kundur and Karthik [17] proposed a novel architecture for joint fingerprinting and decryption (JFD) that holds promise for a better compromise between practicality and security. The scheme provided a good framework for JFD, but the encrypted data is not secure in visual perception since the encryption of signs of DCT coefficients cannot fully scramble the original data. A joint fingerprinting and decryption (JFD) scheme based on vector quantization is proposed with the purpose of protecting multimedia distribution in [18-20]. In [21], the JFE scheme in the compressed domain is proposed. In order to map the community structure of social networks into the tree structure Haar (TSH) transform, the authors proposed a secure content sharing method in the TSH transform domain [22].

Although the above joint encryption and watermarking (fingerprinting) methods meet the requirements of protecting multimedia distribution, they are performed on either the transform domain or pixel domain, and none of them can be applied to hybrid domains for security and privacy in social networks. In addition, the traditional fingerprinting methods do not consider the relationship between users in social networks; then they cannot be applied to secure sharing in social networks. Undoubtedly, safeguarding privacy and security of personal information in social networks is still in its infancy, therefore a fast and simple encryption procedure is required for real time request. In fact, CA is capable of developing chaotic behavior using simple operations or rules offering the benefit of high speed computation, which makes CA an interesting platform for digital image scrambling [23]. With the different wavelet bases and decomposition levels, the DWT can extract different kinds of information from the multimedia, and is therefore very likely to map community structure of social networks into tree structure of DWT for fingerprinting and encryption. To encrypt the important data only, transform domain algorithm can improve the encryption speed, but the encryption effect is weaker obviously. In practice, permutation and diffusion are often combined in order to get high computational security.

In this paper, the first JFE method in hybrid domains using SNA to deal with the issues of multimedia sharing is proposed. The proposed JFE method offers a discussion of how to use SNA for the JFE to realize secure content sharing in social networks. This paper addresses the issue of protecting multimedia distribution using fingerprinting/encryption in the hybrid domains for social networks. Firstly, we describe a method for the fingerprint code produced by the dendrogram of hierarchical and overlapping structure of social network, and conduct to get wavelet decomposition with the structure of fingerprint code. Secondly, we propose a JFE method in hybrid domains, where the fingerprints are embedded in the DWT domain, and the encryption process is carried out in both the spatial domain and the DWT domain. By using our technique, one is well able to design a privacy-preserving and secure sharing system in social networks. By using the proposed scheme, two properties of multimedia content transmission can be ensured, including privacy preserving and traitor tracing, which sometimes deter traitor behaviors. The remainder of this paper is organized as follows. In Section 2, techniques used

in this paper will be introduced. Section 3 details the proposed JFE scheme based on CA and SNA. Then, the experimental results will be given in Section 4. Finally, conclusions are drawn in Section 5.

## II. BASIC THEORY

### A. Social network

A social network is a cluster of people or groups of people with some pattern of contacts or interactions between them. Our intent here is to suggest that SNA [24] can help design secure multimedia sharing systems. Graph theories are available to measure networks. Given a graph  $G = (V, E)$ , the elements of  $V \equiv \{v_1, v_2, \dots, v_n\}$  are the nodes, while the elements of  $E \equiv \{e_1, e_2, \dots, e_n\}$  are edges. Two nodes are connected if they regularly talk to each other.

### B. Chaotic maps

The Logistic Map is a well-known continuous dynamical system. A 1D Logistic map is described as follows:

$$x_{n+1} = ux_n(1 - x_n) \quad (1)$$

where  $u \in [0, 4]$ ,  $x_n \in (0, 1)$ ,  $n=0, 1, 2, \dots$ . The research result shows that the system is in a chaotic state under the condition that  $3.56994 < u \leq 4$ . This Logistic Map generates continuous values between  $[0, 1]$ , which are discretized (binaries) in order to fulfill the initial CA to later encryption. The piecewise linear chaotic map (PWLCM) can be described in Eq. (2):

$$y_{n+1} = F(y_n, \eta) = \begin{cases} y_n / \eta, & 0 \leq y_n < \eta \\ (y_n - \eta) / (0.5 - \eta), & \eta \leq y_n < 0.5 \\ 0, & y_n = 0.5 \\ F(1 - y_n, \eta), & 0.5 \leq y_n < 1 \end{cases} \quad (2)$$

where  $y_n \in (0, 1)$ ,  $n=0, 1, 2, \dots$ . When control parameter  $\eta \in (0, 0.5)$ , Eq. (2) evolves into a chaotic state, and  $\eta$  can serve as a secret key.

### C. Cellular Automata

CA [23] are dynamical complex space and time discrete systems. GL (Game of Life) is governed by its local rules and by its immediate neighbors, which specifies how CA evolves in time. In general, the state of a cell at the next generation depends on its own state and the sum of the neighbor cells. At every time step, all the cells update their states synchronously by applying rules (transition function). Each cell has eight neighbors which are the cells that are horizontally, vertically, or diagonally adjacent. Each cell computes its new state by applying the following transition rules.



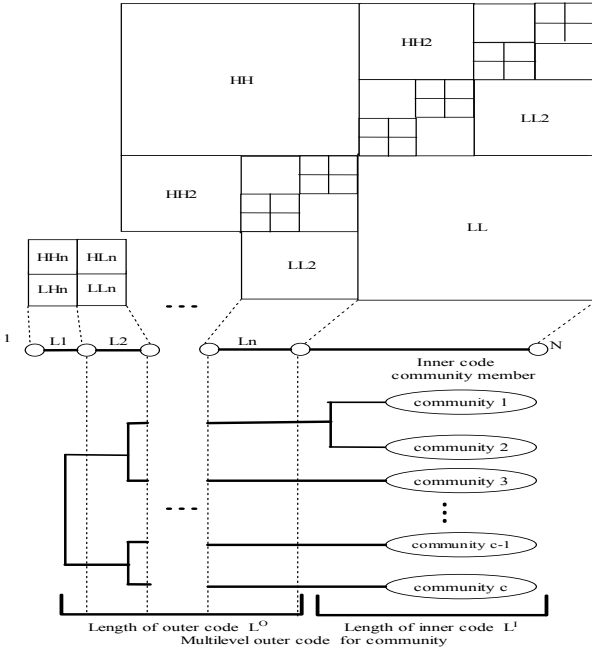


Figure 1. Encoding and DWT using social network analysis

- Any live cell with fewer than two live neighbors dies
- Any live cell with two or three live neighbors lives on to the next generation.
- Any live cell with more than three live neighbors dies, as if by overcrowding.
- Any dead cell with exactly three live neighbors becomes a live cell.

For binary cells  $c_1, c_2, \dots, c_9$ , we say that the transition function, at any time  $t$ , for GL (Game of Life) rule [25] is of the form:

$$\phi \begin{pmatrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \\ c_7 & c_8 & c_9 \end{pmatrix} = \begin{cases} 1, & \text{if } \sum_{i=1}^9 s(c_i, t) = 3 \\ 1, & \text{if } \sum_{i=1}^9 s(c_i, t) = 3, i \neq 5 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

CA capable of exhibiting chaos is attractive in cryptography because of the large keyspace. We propose performing pixel scrambling with the help of the GL.

### III. THE PROPOSED JFE ALGORITHM

The proposed technique uses a multimedia content such as an image and gives a fingerprinted and encrypted image which can be decrypted later for various purposes. The proposed technique consists of four phases. In the first phase, the fingerprint code for users in social networks is produced by the dendrogram of hierarchical and overlapping structure of social networks, followed by the

second phase, image  $I$  is transformed with DWT through mapping community structure of social networks into tree structure of DWT. For the third phase, the fingerprints are embedded in the DWT domain, and the permutation process is carried out in both the spatial domain and the DWT domain, and the low-pass subband coefficients of image DWT decomposition are permuted by GL in DWT domain. Finally, the image after IDWT reconstruction is diffused with PWLCM map and XOR operation in the spatial domain.

#### Notations

For ease of reference, important notations used throughout the paper are listed below.

$N_u$	the number of users
$X^O$	the robust coefficients vector for the outer code
$X^I$	the robust coefficients vector for the inner code
$L^O$	the length of the outer code
$L^I$	the length of the inner code
$Q_\Delta(\cdot)$	the quantization function with step size $\Delta$
$F_k$	the fingerprint information for user $k$
$d_k$	the dither sequence
$Y_k$	the fingerprinted coefficients vector
$w^*$	the codeword
$\hat{m}$	the traitor
$G^0$	the initial two-dimensional grids of cells
$I^{JFS}$	the scrambled and fingerprinted image
$R$	the number of iteration times for scrambling
$P$	the pixel sequence
$FP$	the chaotic sequence for encryption
$CP$	the encrypted coefficients sequence
$I^{JFE}$	the encrypted fingerprinted image

#### A. Fingerprint Encoding Using Social network analysis

Given a multimedia social networks, we try to use the method in [26] to get the overlapping and hierarchical structure of social networks. In the Fig.1, the dendrogram shows the social relations between members in a given social networks. The dendrogram can provide a good concatenated fingerprinting code design by the tree-based fingerprint scheme to reduce the length of code. As shown in Fig.1, users are placed into  $c$  four communities. These communities are encoded by outer code that is constructed by BS code [27], and



the users in each community are encoded by the inner code produced with Tardos scheme [28]. Therefore, for  $N_u$  users can be concatenated by a multilevel outer code for communities and an inner Tardos code for users in the communities [29]. In Fig.1, note that every level outer code can be small if users are grouped appropriately and the collusion probability in the same community is higher than that of collusion between communities.

### B. DWT Using Social network analysis

As a kind of frequency transformation, DWT provides a time-frequency representation of an image. In the DWT transform [30], an image is split into *LL*, *LH*, *HL*, and *HH* subband. In this paper, we transform middle-frequency subbands repeatedly. This process can be repeated until the height or width of the area to be transformed is no longer divisible by two.

For example, in Fig.1, the number of layers of community structure is  $n+1$ , then, the interval *Intv* will be split into  $n+1$  intervals, while the sizes of these intervals are decided by the length of the outer codes in Fig.1. The *LH* and *HL* subbands are then themselves split into a second-level approximation and details, and the process is repeated. For a given code scheme, we define the splitting scheme for multi-level DWT through social network analysis. For example, in Fig.1, the number of the layers of community structure is  $n+1$ , then the number of the layers of outer code is  $n$ , and the *LH* and *HL* subbands for community code embedding will be split into  $n$  levels according to Fig.1. An example of decomposing an image by a 4-level wavelet transformation is shown in Fig. 2.

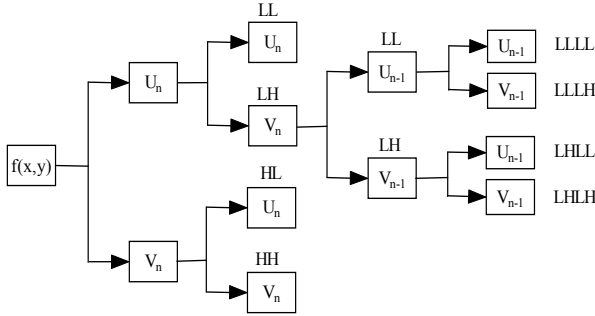


Figure 2. Logarithmic tree decomposition scheme in two dimensional case

### C. The JFE process

The architecture of joint fingerprinting and encryption (JFE) algorithm based on DWT and chaotic CA is designed and shown in Fig.3. The JFE process is composed of two processes: fingerprinting and substitution in the DWT domain, and diffusion in the spatial domain.

#### 1) Fingerprint embedding

Digital fingerprinting is a technique for identifying traitor who uses multimedia content for unintended purposes, such as redistribution. Digital fingerprinting system could realize traitor tracing. Once a traitord copy is detected, the owner extracts the fingerprint of the traitord copy and carries out

traitor tracing algorithms to identify the traitor.

In this paper, we focus on blind watermarking to embed fingerprints because the watermark is detected without reference to the original image once a traitord image was found. To simplify the description of embedding method, we only discuss embedding of a unique fingerprint using an improved QIM scheme.

Suppose  $N_u$  is a set of users. We choose the robust coefficients in all *LH*-level and *HL*-level subbands to create a vector,  $X^O = (x_1, x_2, \dots, x_{L^O})$  of host signals to embed community fingerprint code, and choose another robust coefficients sequence in *LL* subband to create a vector,  $X^I = (x_1, x_2, \dots, x_{L^I})$ , where  $L^O$  and  $L^I$  is the length of the outer codeword and the inner codeword, respectively. So the length of fingerprint code is  $L = L^O + L^I$ . The outer code hiding scheme is described in Eq. (4), and the inner code embedding scheme is similar to that of the outer codeword.

$$Y_k = Q_\Delta(X_k^O + F_k + d_k) - F_k - d_k, k = 1, 2, \dots, N_u \quad (4)$$

where  $Q_\Delta(\cdot)$  is the quantization function with step size  $\Delta$ ,  $F_k$  is the fingerprint information for user  $k$ , and  $d_k$  is a dither sequence which follows a uniformly distribution over  $(-\Delta/2, \Delta/2)$ .

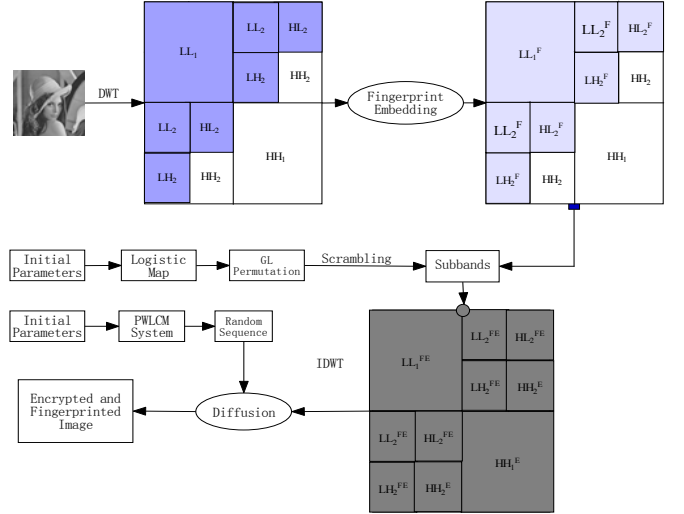


Figure 3. The architecture of image fingerprinting and encryption algorithm

#### 2) The traitor tracing

The traitors tracing algorithm takes a codeword  $w^*$  and outputs at least one traitor. The fingerprinting system is formed by  $\Sigma = \{0, 1\}$ , where  $\Sigma$  is a base alphabet. An  $(N_u, L)$  code is an ordered subset of  $\Sigma^L$ , where  $L$  is the code length. A set  $\Gamma = \{w^{(1)}, w^{(2)}, \dots, w^{(n)}\}$  of codewords will be called an  $(N_u, L)$  code. The codeword  $w^{(i)}$  ( $i \in \{1, 2, \dots, N_u\}$ ), the

$i$ -th codeword of  $(N_u, L)$  code, is assigned to the user  $u_i$ . In our implementation, we apply the minimum-distance detector technique to trace the traitor who leaked information. The robust coefficients extracted from all LH, HL, and LL subband compose a long vector  $Z$  with size  $L$ . By calculating the difference as follow:

$$\hat{m} = \arg \min_{k=1,2,\dots,N_u} \|Z - Y_k\|^2 \quad (5)$$

The detector produces the value of the  $\hat{m}$  th user, which is declared the traitor.

### 3) Encryption and decryption algorithm

The contents are encrypted totally via chaotic CA permutation process and XOR diffusion process. The proposed encryption algorithm can be divided into the following steps:

**Step 1:** Divide the original image of size  $M \times N$  into two parts,  $I = I_1 + I_2$ , where  $I_*$  denotes half of the image. Then calculate the sum of both parts denoted by  $S_{I_1}$  and  $S_{I_2}$ , respectively. **Subtract** these sums and multiply the total number of gray levels in the image to get  $Th$ , which is used to generate the initial value using MD5, which is a widely used cryptographic hash function with a 128-bit hash value [31]. The MD5 hash value of  $Th$  is  $V^{Th}$ . According to the order of bits, we segment  $V^{Th}$  into eight 16-bit parts  $V_1^{Th}, V_2^{Th}, \dots, V_8^{Th}$ , and compute the values of these parts in decimal numbers. We can compute initial values,  $x_0, y_0$ , and parameters  $u, \eta$ , which are viewed as the secret keys in this algorithm. Our encryption algorithm actually does have some of the following secret keys: (1) The initial values  $x_0$  (Logistic map) and  $y_0$  (PWLCM system); (2) The control parameters  $u$  (Logistic map) and  $\eta$  (PWLCM system).

$$x_0 = \frac{V_1^{Th}}{2^{16}}, \quad y_0 = \frac{V_2^{Th}}{2^{16}}, \quad u = 3.57 + \frac{V_5^{Th}}{2^{16}} \times 0.43, \quad \eta = \frac{V_6^{Th}}{2^{17}}$$

**Step 2:** We calculate the one-level DWT coefficient matrix of the image  $I$ . Then we can get four sub-bands: the approximation coefficients LL, and the detailed coefficients HL, LH, and HH;

**Step 3:** Use a logistic map to generate the sequence  $(x_1, x_2, \dots, x_{M/2 \times N/2})$ . Then we create a two-dimensional grids of cells  $G^0$ , as the seeds of GL, where  $G^0$  is used to permute the LL coefficient matrix. The rule is that if the value of  $x_i$  is bigger than the mean value of the sequence, the corresponding cell is alive, else is dead;

**Step 4:** When producing the  $k$  th generation  $G^k$  by the rules of GL, the corresponding plain coefficients are inserted in the scrambling matrix one by one;

**Step 5:** After  $R$  iterations, we stop and put the rest of the value into the scrambling coefficient matrix;

**Step 6:** Perform two-level IDWT reconstruction with the encrypted wavelet transform coefficients. We have now the scrambled and fingerprinted image  $I^{JFS}$ ;

**Step 7:** Convert 2D image  $I^{JFS}$  into a 1D pixel sequence  $P = \{p_i\}$  by using the pixel positions;

**Step 8:** Using the PWLCM map to generate chaotic sequence  $FP = \{fp_1, fp_2, \dots, fp_{M \times N}\}$ , then compute

$$cp_i = (dt_i \oplus p_i) \oplus (c_{i-1} \oplus r_i), i = 1, 2, \dots, M \times N \quad (6)$$

where  $dt_i = (fp_i \times 10^{14}) \bmod 256$ ,  $r_i = r_{i-1} \oplus p_i, r_0 = 128$ . This produces the new sequence  $CP = \{cp_0, cp_1, \dots, cp_{M \times N}\}$ ;

**Step 9:** Convert  $CP = \{cp_i\}$  into a 2D image  $I^{JFE}$  according to element positions, i.e., encrypted image, by using the element positions.

The decryption algorithm uses the inverse process of the encryption algorithm.

## IV. EXPERIMENT RESULTS AND SECURITY ANALYSIS

The performance of the proposed JFE technique demonstrated using MATLAB platform on a computer having a Pentium(R) Dual-Core E5700 CPU and 2-GB RAM. A number of experiments have been performed on a set of grayscale test images, which include images *Lena*, *Peppers*, *Airplane*, *Couple*, *Fishingboat*, *Bridge*, *Baboon*, and *Watch*. Six parameters are used as the keys: the initial values  $x_0$  (Logistic map);  $y_0$  (PWLCM system); the parameters  $u$  (Logistic map),  $\eta$  (PWLCM system),  $k$ , and the iteration times  $R$ . In our experiments  $x_0 = 0.986372185231$ ,  $u = 3.95374324256$ ,  $y_0 = 0.475291583612$ ,  $\eta = 0.419673893132$ .

### A. Perceptual Security

The visual impact of the proposed encryption scheme is demonstrated in Fig.4. It is clear that all the encrypted images become noise-like images and are all actually unintelligible. Therefore, the proposed scheme indeed possessed high perceptual security. The fingerprint is embedded in the DWT domain using social network analysis. In order to preserve visual quality, the fingerprint in the fingerprinted copy should be imperceptible and perceptually undetectable. Fig.4 (b), (f), (j), and (n) show some experimental results of decrypted fingerprinted images. It can be observed that the quality of the fingerprinted image doesn't have any observable change.

### B. Ability of resisting brute-force attack

Key space size is the total number of different keys that can be used in an encryption algorithm. The total key space includes two processes: confusion and diffusion. Our encryption algorithm actually does have some of the following secret keys: (1) Initial values  $x_0$  (Logistic map),  $y_0$  (PWLCM system); (2) Parameters  $u$  (Logistic map),  $\eta$  (PWLCM system),  $k$ ; (3) The iteration times  $R$ . The sensitivity to  $x_0$ ,  $y_0$ ,  $u$  and  $\eta$  is considered as  $10^{-16}$  [32], The total key space is about  $10^{16 \times 4} = 10^{64}$ . This key space is large enough to resist the brute-force attack.

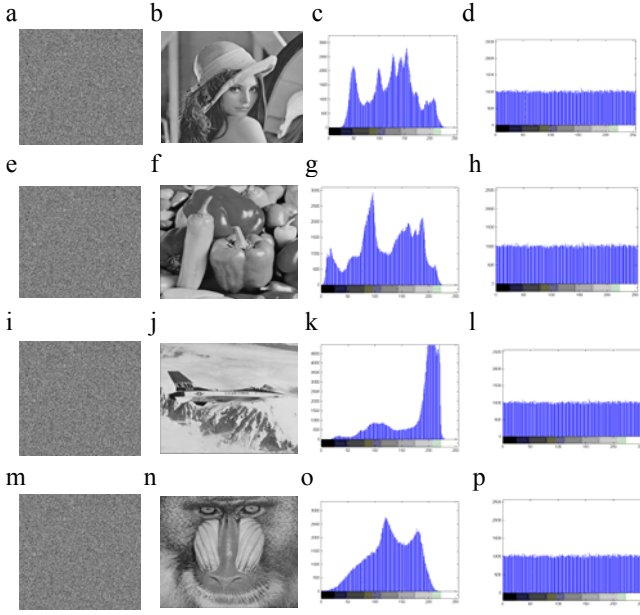


Figure 4. The experimental results: (a), (e), (i), and (m) are the encrypted images, (b), (f), (j), and (n) are the decrypted images with fingerprints, (c), (g), (k), and (o) are the grey histograms of the original images, (d), (h), (l), and (p) are the grey histograms of the encrypted images

### C. Resistance to statistical attack

The basic idea is to compare the histograms of the original and encrypted images. Fig. 4 show the grey-scale histograms. Comparing the two histograms we find that the pixel grey values of the original images are concentrated on some values, the histogram of encrypted images is very uniform. The features of the original images are destroyed during the encryption process, which makes statistical attacks difficult.

An effective encryption algorithm can reduce the correlation between adjacent pixels. In order to test the correlation of two adjacent pixels, we randomly select 3000 pairs (horizontal, vertical and diagonal) of adjacent pixels from the original image and the encrypted image. **Using the following formulas for the correlation coefficient, we obtain:**

$$D(x) = \frac{1}{N} \sum_{i=1}^N (x_i - E(x))^2 \quad (7)$$

$$\text{cov}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - E(x))(y_i - E(y)) \quad (8)$$

$$r_{xy} = \frac{\text{cov}(x, y)}{\sqrt{D(x)} \times \sqrt{D(y)}} \quad (9)$$

where  $x$  and  $y$  are the grey values of two adjacent pixels in the image,  $\text{cov}(x, y)$  is the covariance,  $D(x)$  is the variance, and  $E(x)$  is the mean. Fig. 5(a), (b) show the correlation of two adjacent pixels in Lena image and its encrypted image, where the correlation coefficients are 0.9468 and 0.0036, respectively. It can clearly be seen that our algorithm can destroy the relativity effectively; the proposed image encryption algorithm has a strong ability to resist statistical attack.

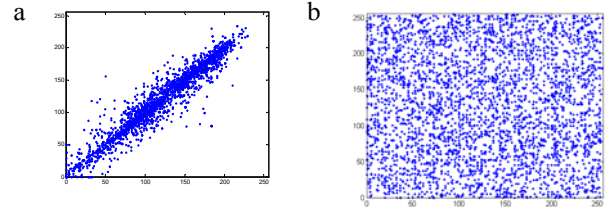


Figure 5. Correlation of two adjacent pixels in the original image and in the encrypted image.

### D. Resistance to differential attack

Attackers often make a slight change to the original image, and use the proposed algorithm to encrypt it before and after changing. Then they compare two encrypted images to find out the difference, which is called differential attack. Such difference can be measured by means of two criteria namely, the number of pixel change rate (NPCR) and the unified average changing intensity (UACI). The proposed cryptosystem can ensure two ciphered images completely different, even if there is only one bit difference between them. The following formulas are used to calculate the NPCR and the UACI.

$$C(i, j) = \begin{cases} 0, & \text{if } T_1(i, j) = T_2(i, j) \\ 1, & \text{if } T_1(i, j) \neq T_2(i, j) \end{cases} \quad (10)$$

$$\text{NPCR} = \frac{\sum_{i=1}^M \sum_{j=1}^N C(i, j)}{M \times N} \times 100\% \quad (11)$$

$$UACI = \frac{\sum_{i=1}^M \sum_{j=1}^N \|T_1(i, j) - T_2(i, j)\|}{M \times N} \times 100\% \quad (12)$$

where  $M$  and  $N$  are the height and width of the image, and  $T_1(i, j)$  and  $T_2(i, j)$  denote the grey value of the encrypted images before and after one pixel of the plain image is changed. We obtained  $NPCR_{mean}=0.9965$  and  $UACI_{mean}=0.3256$ , from the simulation of the images. This result demonstrates that our algorithm has a strong ability to resist differential attack.

#### E. Information entropy

If the distribution of grey values is more uniform, the information entropy is greater. The information entropy is defined as follows:

$$H(m) = -\sum_{i=0}^L P(m_i) \log_2 P(m_i) \quad (13)$$

where  $m_i$  is the  $i$ th pixel grey value for an  $L$  level grey image,  $P(m_i)$  is the emergence probability of  $m_i$ , so  $\sum_{i=0}^L P(m_i) = 1$ . For an ideal random image, the value of the information entropy is 8. An effective encryption algorithm should make the information entropy tend to 8. We obtained an information entropy  $H=7.9946$ , that is very close to 8. It can be seen that the proposed algorithm is very effective.

#### F. Discussion of the encryption process

According to Section 3, we know that the diffusion process in Fig. 6 only enhances the unintelligibility of the encrypted image and is optional for the proposed method. Therefore, even if the chaotic map used in GL is cracked, the hacker still cannot decrypt the image since the random sequence of diffusion remains secret. Fig. 6 shows the comparison of when a diffusion process is and is not applied. It is clear that the diffusion process in the proposed scheme can enhance perceptual security. Therefore, if confidentiality is in high demand, the proposed first method with diffusion can be applied. Otherwise, the encryption method with only permutation can be performed since only a rough sketch without details would be revealed, making the perceptual quality unacceptable.

#### G. The adaptability of the algorithm

According to Section 3, the proposed algorithm encrypts the original image of size  $M \times N$  based on the height and width of the image, therefore, the algorithm can encrypt any images adaptively. In addition, the images are encrypted totally via chaotic CA permutation process and XOR diffusion process, after the two processes, the relativity of the original images is destroyed, in the end, the encrypted images become noise-like images. According to the grey histograms of the encrypted images shown in Fig.4 and the resistance to statistical attack in

Fig.5, the proposed algorithm can also apply to a larger set of images.

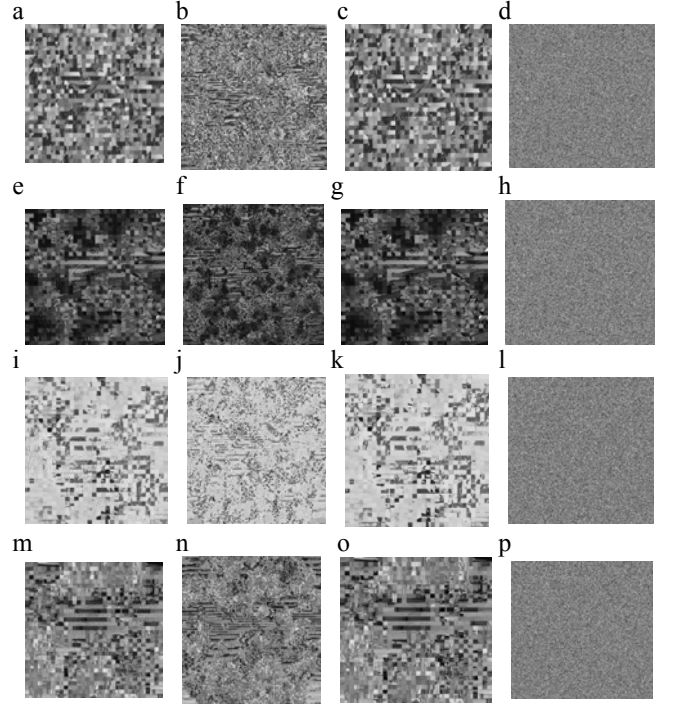


Figure 6. Evaluation of the encryption process: (a), (e), (i), (m) are  $4 \times 4$  block permutation in the 2-level LL subband via GL, (b), (f), (j), (n) are single coefficient permutation in 2-level LL subband via GL, (c), (g), (k), (o) are images which are permuted by  $4 \times 4$  block in all subbands of 1-level DWT via GL, (d), (h), (l), (p) are encrypted images with permutation on 2-level LL subband via GL and diffusion in spatial domain

#### H. Encryption Efficiency

This subsection presents a comparative analysis of the proposed technique with the joint encryption/watermarking algorithm for verifying medical image presented by D. Bouslimi et al. [8]. The authors have suggested the merging of a stream cipher algorithm (RC45) and watermarking approaches. However, the stream cipher algorithm for encryption still has a high time complexity according to the abundant data in images. The approach is inefficient because a large amount of time is spent on the encryption process. The proposed algorithm is able to overcome the aforementioned weaknesses by confusing LL subband with CA and diffusion in the spatial domain. This proves an improvement by the proposed technique over the existing watermarking and encryption technique.

In the case of multimedia distribution in social networks, if a technique requires a huge amount of time to encrypt/decrypt an image, then it is not considered a feasible technique. Therefore, the time efficiency of the proposed technique is evaluated in this subsection. In the proposed technique, the time efficiency is depicted in Table 1. These experiments are run on a computer having a Pentium(R) Dual-Core E5700, and with MATLAB 7.1 version. From the table, it is clear that time taken for the encryption process is completed in 0.5s or so. Therefore, we can say that the proposed JFE scheme is time



efficient, and it can provide security services within strict time deadlines to users.

TABLE I. TIME EFFICIENCY

Images	Lena	peppers	airplane	baboon	watch
Time(s)	0.560	0.5228	0.4997	0.5188	0.5397

## V. CONCLUSION

The traditional JFE methods don't consider the relationship between users, therefore they cannot be applied to secure content sharing for social networks because of the tremendous scale of social networks. In this paper, the first JFE method based on CA and SNA in hybrid domains to deal with the issues of multimedia sharing and traitor tracing is proposed. The experiment results and algorithm analyses show that the new algorithm possesses a large key space and can resist brute-force, differential and statistical attacks. Our method does not require a great deal of computation time because the proposed algorithm confuses the important data using chaotic CA in the DWT domain. This algorithm is simple, secure, fast, and easy to be realized. The fundamental goal of our research has been to provide a useful synthesis of SNA for the field of secure multimedia distribution in social networks.

## ACKNOWLEDGMENT

This work is supported by NSF of China Grants 61272409, 61370092, 61370223, Natural Science Foundation of Hubei Province of China (No.2013CFC005), Key Project in Hubei Provincial Department of Education (No.D20142703), and Youth innovation team project in Hubei Provincial Department of Education (No.T201410).

## REFERENCES

- [1] P. Belimpasakis and A. Saarinen, "Sharing with people: a system for user-centric content sharing," *Multimedia Syst.*, vol. 16, pp. 399-421, 2010.
- [2] J. Dittmann, P. Wohlmacher, and K. Nahrstedt, "Using cryptographic and watermarking algorithms," *IEEE Multimedia* vol. 8, pp. 54-65, 2001.
- [3] C. H. Yuen and K. W. Wong, "A chaos-based joint image compression and encryption scheme using DCT and SHA-1," *Appl. Soft. Comput.*, vol. 11, pp. 5092-5098, 2011.
- [4] G. Chen, Y. Mao, and C. K. Chui, "A symmetric image encryption scheme based on 3D chaotic cat maps," *Chaos Solitons Fractals*, vol. 21, pp. 749-761, 2004.
- [5] S. Behnia, A. Akhshani, H. Mahmodi, and A. Akhavan, "A novel algorithm for image encryption based on mixture of chaotic maps," *Chaos Solitons Fractals*, vol. 35, pp. 408-419, 2008.
- [6] G. Alvarez and S. Li, "Some basic cryptographic requirements for chaos-based cryptosystems," *Int. J. Bifurcation Chaos*, vol. 16, pp. 2129-2151, 2006.
- [7] B. Jankowski, W. Mazurczyk, and K. Szczypiorski, "Information hiding using improper frame padding," *arXiv preprint arXiv:1005.1925*, 2010.
- [8] D. Bouslimi, G. Coatrieux, and C. Roux, "A joint encryption/watermarking algorithm for verifying the reliability of medical images: Application to echographic images," *Comput. Meth. Programs Biomed.*, vol. 106, pp. 47-54, 2012.
- [9] T. Bianchi and A. Piva, "Secure watermarking for multimedia content protection: A review of its benefits and open issues," *IEEE Signal Processing Magazine*, vol. 30, pp. 87-96, 2013.
- [10] N. Memon and P. W. Wong, "A buyer-seller watermarking protocol," *IEEE Trans. Image Process.*, vol. 10, pp. 643-649, 2001.
- [11] M. Cancellaro, F. Battisti, M. Carli, G. Boato, F. De Natale, and A. Neri, "A commutative digital image watermarking and encryption method in the tree structured Haar transform domain," *Signal Process.-Image Commun.*, vol. 26, pp. 1-12, 2011.
- [12] A. Subramanyam and S. Emmanuel, "Robust watermarking of compressed JPEG images in encrypted domain," 2011, pp. 37-57.
- [13] A. Subramanyam, S. Emmanuel, and M. S. Kankanhalli, "Robust Watermarking of Compressed and Encrypted JPEG2000 Images," *IEEE Trans. Multimedia* vol. 14, pp. 703-716, 2012.
- [14] H. Feng, H. Ling, F. Zou, W. Yan, M. Sarem, and Z. Lu, "A collusion attack optimization framework toward spread-spectrum fingerprinting," *Applied Soft Computing*, vol. 13, pp. 3482-3493, 2013.
- [15] J. Dittmann, P. Schmitt, E. Saar, and J. Ueberberg, "Combining digital watermarks and collusion secure fingerprints for digital images," *J. Electron. Imaging*, vol. 9, pp. 456-467, 2000.
- [16] T. Thomas, S. Emmanuel, A. Subramanyam, and M. S. Kankanhalli, "Joint watermarking scheme for multiparty multilevel DRM architecture," *IEEE Trans. Inf. Forensic Secur.*, vol. 4, pp. 758-767, 2009.
- [17] D. Kundur and K. Karthik, "Video fingerprinting and encryption principles for digital rights management," *Proc. IEEE*, vol. 92, pp. 918-932, 2004.
- [18] C. Y. Lin, P. Prangjarote, L. W. Kang, W. L. Huang, and T. H. Chen, "Joint fingerprinting and decryption with noise-resistant for vector quantization images," *Signal Process.*, vol. 92, pp. 2159-2171, 2012.
- [19] C.-Y. Lin, P. Prangjarote, C.-H. Yeh, and H.-F. Ng, "Reversible joint fingerprinting and decryption based on side match vector quantization," *Signal Processing*, vol. 98, pp. 52-61, 2014.
- [20] M. Li, D. Xiao, Y. Zhang, and H. Liu, "Attack and improvement of the joint fingerprinting and decryption method for vector quantization images," *Signal Processing*, vol. 99, pp. 17-28, 2014.
- [21] C. Ye, H. Ling, F. Zou, Z. Lu, Z. Xiong, and K. Zhang, "A novel JFE scheme for social multimedia distribution in compressed domain using SVD and CA," in *Digital Forensics and Watermarking*, ed: Springer, 2013, pp. 507-519.
- [22] C. Ye, H. Ling, F. Zou, and C. Liu, "Secure content sharing for social networks using fingerprinting and encryption in the TSH transform domain," in *Proceedings of the 20th ACM international conference on Multimedia*, 2012, pp. 1117-1120.
- [23] S. Wolfram, "A new kind of science," *Wolfram Media*, 2002.
- [24] S. Wasserman and K. Faust, *Social network analysis: Methods and applications* vol. 8: Cambridge university press, 1994.
- [25] A. Adamatzky, *Game of life cellular automata*: Springer, 2010.
- [26] H. Shen, X. Cheng, K. Cai, and M. B. Hu, "Detect overlapping and hierarchical community structure in networks," *Physica A*, vol. 388, pp. 1706-1712, 2009.
- [27] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inf. Theory* vol. 44, pp. 1897-1905, 1998.
- [28] G. Tardos, "Optimal probabilistic fingerprint codes," *J. ACM* vol. 55, p. 10, 2008.
- [29] C. Ye, H. Ling, F. Zou, and Z. Lu, "A new fingerprinting scheme using social network analysis for majority attack," *Telecommunication Systems*, vol. 54, pp. 315-331, 2013.
- [30] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674-693, 1989.
- [31] H. Liu and X. Wang, "Color image encryption based on one-time keys and robust chaotic maps," *Comput. Math. Appl.*, vol. 59, pp. 3320-3327, 2010.
- [32] M. K. Khan, J. Zhang, and K. Alghathbar, "Challenge-response-based biometric image scrambling for secure personal identification," *Future Generation Computer Systems*, vol. 27, pp. 411-418, 2011.

# An Improved Collaborative Movie Recommendation System using Computational Intelligence

Zan Wang

Department of Software Engineering,  
School of Computer Software,  
Tianjin University  
Tianjin, 300072, P.R. China  
wangzan@tju.edu.cn

Xue Yu\*, Nan Feng

Department of Information  
Management & Management Science,  
College of Management and  
Economics, Tianjin University  
Tianjin, 300072, P.R. China  
{yuki, fengnan}@tju.edu.cn

Zhenhua Wang

American Electric Power  
Gahanna, OH 43230, United States  
zhw.powersystem@gmail.com

**Abstract**—Recommendation systems have become prevalent in recent years as they dealing with the information overload problem by suggesting users the most relevant products from a massive amount of data. For media product, online collaborative movie recommendations make attempts to assist users to access their preferred movies by capturing precisely similar neighbors among users or movies from their historical common ratings. However, due to the data sparsely, neighbor selecting is getting more difficult with the fast increasing of movies and users. In this paper, a hybrid model-based movie recommendation system which utilizes the improved K-means clustering coupled with genetic algorithms (GA) to partition transformed user space is proposed. It employs principal component analysis (PCA) data reduction technique to dense the movie population space which could reduce the computation complexity in intelligent movie recommendation as well. The experiment results on Movielens dataset indicate that the proposed approach can provide high performance in terms of accuracy, and generate more reliable and personalized movie recommendations when compared with the existing methods.

**Keywords**—Movie recommendation, Collaborative filtering, Sparsity data, Genetic algorithms, K-means

## I. INTRODUCTION

Fast development of internet technology has resulted in explosive growth of available information over the last decade. Recommendation systems (RS), as one of the most successful information filtering applications, have become an efficient way to solve the information overload problem. The aim of Recommendation systems is to automatically generate suggested items (movies, books, news, music, CDs, DVDs, webpages) for users according to their historical preferences and save their searching time online by exacting useful data.

Movie recommendation is the most widely used application coupled with online multimedia platforms which aims to help customers to access preferred movies

intelligently from a huge movie library. A lot of work has been done both in the academic and industry area in developing new movie recommendation algorithms and extensions. The majority of existing recommendation systems is based on collaborative filtering (CF) mechanism [1-3] which has been successfully developed in the past few years. It first collects ratings of movies given by individuals and then recommends promising movies to target customer based on the “like-minded” individuals with similar tastes and preferences in the past. There have been many famous online multimedia platforms (e.g., youtube.com, Netflix.com, and douban.com) incorporated with CF technique to suggest media products to their customers. However, traditional recommendation systems always suffer from some inherent limitations: poor scalability, data sparsity and cold start problems [3, 4]. A number of works have developed model-based approaches to deal with these problems and proved the benefits on prediction accuracy in RS [5-8].

Model-based CF uses the user-item ratings to learn a model which is then used to generate online prediction. Clustering and dimensionality reduction techniques are often employed in model-based approaches to address the data sparse problem [5, 8-9]. The sparsity issues arise due to the insufficiency of user’s history rating data and it is made even more severe in terms of the dramatically growth of users and items. Moreover, high-dimensional rating data may cause it difficult to extract common interesting users by similarity computation, which results in poor recommendations. In the literature, there have been many model-based recommendation systems developed by partitioning algorithms coupled, such as K-means and self-organizing maps (SOM) [15-18, 20]. The aim of clustering is to divide users into different groups to form “like-minded” (nearest) neighbors instead of searching the whole user space, which could dramatically improve the system scalability. It has been proved that clustering-based recommendation systems outperform the pure CF-based ones in terms of efficiency and prediction quality [7, 9-11]. In many works, the clustering methods are conducted with the entire dimensions of data which might lead to somewhat

---

\*Corresponding author. Tel.: +86 22 27406125; Fax: +86 22 87401540

inaccuracy and consume more computation time. In general, making high quality movie recommendations is still a challenge, and exploring an appropriate and efficiency clustering method is a crucial problem in this situation.

To address challenges aforementioned, a hybrid model-based movie recommendation approach is proposed to alleviate the issues of both high dimensionality and data sparsity. In this article, we construct an optimized clustering algorithm to partition user profiles which have been represented by denser profile vectors after Principal Component Analysis transforming. The whole system consists of two phases, an online phase, and an offline phase. In offline phase, a clustering model is trained in a relatively low dimensional space, and prepares to target active users into different clusters. In online phase, a TOP- $N$  movie recommendation list is presented for an active user due to predicted ratings of movies. Furthermore, a genetic algorithm (GA) is employed in our new approach to improve the performance of  $K$ -means clustering, and the improved clustering algorithm is named as GA-KM. We further investigate the performance of the proposed approach in Movielens dataset. In terms of accuracy and precision, the experiment results prove that the proposed approach is capable of providing more reliable movie recommendations comparing with the existing cluster-based CF methods.

The remainder of this paper is organized as follows: section 2 gives a brief overview on collaborative recommendation systems and clustering-based collaborative recommendation. Then we discuss the development of our proposed approach called PCA-GAKM movie recommendation system in detail in Section 3. In section 4, experiment results on movielens dataset and discussion are described. Finally, we summarize this paper and the future work is given.

## II. RELATED WORK

### A. Movie Recommendation Systems based on Collaborative Filtering

Recommendation systems (RS), introduced by Tapestry project in 1992, is one of the most successful information management systems [12]. The practical recommender applications help users to filter mass useless information for dealing with the information overloading and providing personalized suggestions. There has been a great success in e-commerce to make the customer access the preferred products, and improve the business profit. In addition, to enhance the ability of personalization, recommendation system is also widely deployed in many multimedia websites for targeting media products to particular customers. Nowadays, Collaborative filtering (CF) is the most effective technique employed by movie recommendation systems, which is on the basis of the nearest-neighbor mechanism. It is on the assumption that people who have similar

history rating pattern may be on the maximum likelihood that have the same preference in the future. All “like-minded” users, called neighbors, are derived from their rating database that is recording evaluation values to movies. The prediction of a missing rating given by a target user can be inferred by the weighted similarity of his/her neighborhood.

Reference [6] divides CF techniques into two important classes of recommender systems: memory-based CF and model-based CF. Memory-based CF operate on the entire user space to search nearest neighbors for an active user, and automatically produce a list of suggested movies to recommend. This method suffers from the computation complexity and data sparsity problem. In order to address computational and memory bottleneck issues, Sarwar et al. proposed an item-based CF in which the correlations between items are computed to form the neighborhood for a target item [4]. In their empirical studies, it is proved that item-based approach can shorten computation time apparently while providing comparable prediction accuracy.

Model-based CF, on the other hand, develops a pre-build model to store rating patterns based on user-rating database which can deal with the scalability and sparsity issues. In terms of recommendation quality, model-based CF applications can perform as well as memory-based ones. However, model-based approaches are time-consuming in building and training the offline model which is hard to be updated as well. Algorithms that often used in model-based CF applications include Bayesian networks [6], clustering algorithms [9-11], neural networks [13], and SVD (Singular Value Decomposition) [5, 14]. While traditional collaborative recommendation systems have their instinct limitations, such as computational scalability, data sparsity and cold-start, and these issues are still challenges that affect the prediction quality. Over the last decade, there have been high interests toward RS area due to the possible improvement in performance and problems solving capability.

### B. Clustering-based Collaborative Recommendation

In movie recommendation, clustering is a widely used approach to alleviate the scalability problem and provides a comparable accuracy. Many works have proved with experiments that the benefits of clustering-based CF frameworks [15-18]. The aim of clustering algorithms is to partition objects into clusters that minimize the distance between objects within a same cluster to identify similar objects. As one of model-based CF methods, clustering-based CF is used to improve  $k$ -nearest neighbor ( $k$ -NN) performance by prebuilding an offline clustering model. Typically, numbers of users can be grouped into different clusters based on their rating similarity to find “like-minded” neighbors by using the clustering technique. Then the clustering process is performed offline to build the model. When a target user arrived, the online module assigns a cluster with a largest

similarity weight to him/her, and the prediction rating of a specified item is computed based on the same cluster numbers instead of searching whole user space.

According to early studies in [3, 6], CF coupled with clustering algorithms is a promising schema to provide accuracy personal recommendations and address the large scale problems. But they also concluded that good performance of clustering-based CF depends on appropriate clustering techniques and the nature of dataset as well. Li and Kim applied fuzzy  $K$ -means clustering method to group items which combined the content information for similarity measurement to enhance the recommendation accuracy [9]. In the conclusion of their work, it shows that the proposed cluster-based approach is capable of dealing with the cold start problem. Furthermore, Wang et al. developed [19] a new approach to cluster both the rows and columns fuzzily in order to condense the original user rating matrix. In Kim and Ahn's research, a new optimal  $K$ -means clustering approach with genetic algorithms is introduced to make online shopping market segmentation [10]. The proposed approach is tested to exhibit better quality than other widely used clustering methods such as pure  $K$ -means and SOM algorithms in the domain of market segmentation, and could be a promising tool for e-commerce recommendation systems.

Liu and Shih proposed two hybrid methods that exploited the merits of the Weighted RFM-based and the preference-based CF methods to improve the quality of recommendations [20]. Moreover,  $K$ -means clustering is employed to group customers based on their enhanced profile. The experiments prove that the combined models perform better than the classical  $K$ -NN mechanism. Xue et al. proposed a novel CF framework that uses clustering technique to address data sparsity and scalability in common CF [7]. In their work,  $K$ -means algorithm is employed to classify users for smoothing the rating matrix that is to generate estimated values for missing ratings corresponding to cluster members. In latter recommendation phrase, the clustering result is utilized to neighborhood selection for an active user. The experiment results show that the novel approach can demonstrate significant improvement in prediction accuracy. Georgiou and Tsapatsoulis developed a genetic algorithm based clustering method which allows overlapping clusters to personalized recommendation, and their experiment findings show that the new approach outperforms  $K$ -means clustering in terms of efficiency and accuracy [21].

The above works have proved that clustering-based CF systems show more accuracy prediction and help deal with scalability and data sparse issues.

### III. PCA-GAKM BASED COLLABORATIVE FILTERING FRAMEWORK

In this section, we aim at developing a hybrid cluster-based model to improve movie prediction accuracy, in which offline and online modules are coupled to make

intelligent movie recommendations. Traditional CF search the whole space to locate the  $k$ -nearest neighbors for a target user, however, considering the super high dimensionality of user profile vectors, it is hard to calculate a similarity to find like-minded neighbors based on ratings which leads to poor recommendation because of sparse. To address such an issue, our offline clustering module involves two phases: 1) to concentrate feature information into a relatively low and dense space using PCA technique; 2) To build an effective GA-KM clustering algorithm based on the transformed user space.

Fig 1 shows an overview of the new approach: offline module represented by light flow arrows, is used to optimize and train the user profiles into different clusters on the basis of history rating data; online module is real-time movie recommendation noted with dark flow arrows, to which a target user's rating vector input, and come out with a TOP-N movie recommendation list. We explain the details in the following.

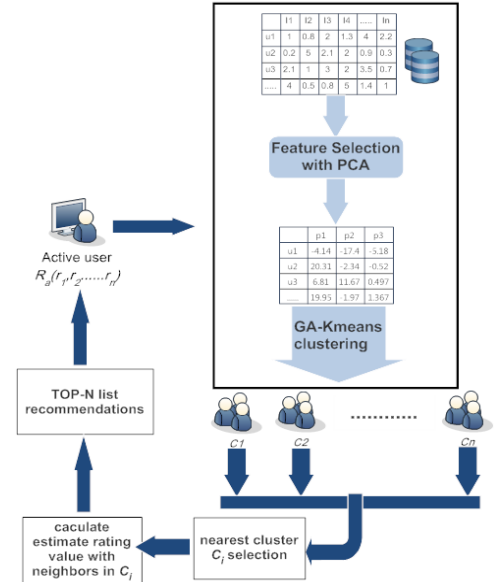


Fig 1. Overview of proposed movie recommendation system framework

#### A. Pre-processing Data using PCA

In this section, we employ a linear feature extraction technique to transfer the original high space into a relatively low space in which carries denser feature information. Since the high dimensionality of user-rating matrix which is mostly empty at the beginning makes the similarity computation very difficult, our approach is started with PCA-based dimension reduction process. As one of the most successful feature extraction techniques, PCA is widely used in data prefilling and dimensional reduction of collaborative filtering systems [14][22-23].

The main idea of PCA is to convert the original data to a new coordinate space which is represented by principal component of data with highest eigenvalue. The first principal component vector carries the most



significant information after ordering them by eigenvalues from high to low. In general, the components of lesser significance are ignored to form a space with fewer dimensions than the original one. Suppose we have user-rating  $m \times n$  matrix in which  $n$ -dimension vector represents user's profile. It turns out the  $n$  principal components after performing eigenvalue decomposition, and we select the only first  $d$  components ( $d \ll n$ ) to keep in the new data space which is based on the value of accumulated proportion of 90% of the original one. As a result, the reduced feature vectors from PCA are prepared to feed to GA-KM algorithm for classification.

#### B. An Enhanced K-means Clustering Optimized by Genetic Algorithms

Memory-based CF systems suffer from two main common flaws: cold-start and data sparse. Many research works have proved benefits of cluster-based CF in terms of the increased quality of recommendation and robustness. The objective of this section is to propose an effective classification method to ensure the users who have the same preference could fall into one cluster to generate accurate like-minded neighbors. The GA-KM algorithm we employed in this work can be roughly performed in two phases:

- *K*-means clustering

*K*-means algorithm is one of the most commonly used clustering approaches due to its simplicity, flexibility and computation efficiency especially considering large amounts of data. *K*-means iteratively computes  $k$  cluster centers to assign objects into the most nearest cluster based on distance measure. When center points have no more change, the clustering algorithm comes to a convergence. However, *K*-means lacks the ability of selecting appropriate initial seed and may lead to inaccuracy of classification. Randomly selecting initial seed could result in a local optimal solution that is quite inferior to the global optimum. In other words, different initial seeds running on the same dataset may produce different partition results.

Given a set of objects  $(x_1, x_2, \dots, x_n)$ , where each object is an  $m$ -dimensional vector, *K*-means algorithm aims to automatically partition these objects into  $k$  groups. Typically, the procedure consists of the following steps [24-25]:

- 1) choose  $k$  initial cluster centers  $C_j, j=1, 2, 3 \dots k$ ;
- 2) each  $x_i$  is assigned to its closest cluster center according to the distance metric;
- 3) compute the sum of squared distances from all members in one cluster:

$$J = \sum_{j=1}^k \sum_{i \in C_{temp}} \|x_i - M_j\|^2 \quad (1)$$

where  $M_j$  denotes the mean of data points in  $C_{temp}$ ;

- 4) if there is no further change, then the algorithm has converged and clustering task is end; otherwise, recalculate the  $M_j$  of  $k$  clusters as the new cluster centers and go to step2.

To overcome the above limitations, we introduce genetic algorithm to merge with *K*-means clustering process for the enhancement of classification quality around a specified  $k$ .

- Genetic algorithms

Genetic algorithms (GA) are inspired by nature evolutionary theory which is known for its global adaptive and robust search capability to capture good solutions [26]. It can solve diverse optimization problems with efficiency due to its stochastic search on large and complicated spaces. The whole process of GA is guided by Darwin's nature survival principle and provides a mechanism to model biological evolution. A GA utilizes a population of "individuals" as chromosomes, representing possible solutions for a given problem. Each chromosome contains number of genes which is used to compute fitness to determine the likelihood of reproduction for the next generation. Typically, a chromosome with the fittest value will be more likely to reproduce than unfit ones. GAs iteratively creates new populations replace of the old ones by selecting solutions (chromosomes) on the basis of pre-specified fitness function. During each successive iteration, three genetic operators are executed to construct the new generation known as selection, crossover and mutation. Selection process selects a proportion of the current population to breed a new generation according to their fitness value. Crossover operator allows swapping a portion of two parent chromosomes for each other to be recombined into new offspring. Mutation operator randomly alters the value of a gene to produce offspring. All above operators provide the means to extend the diversity of population over time and bring new information to it. Finally, the iterations tend to terminate when the fitness threshold is met or a pre-defined number of generations is reached.

A common drawback of *K*-means algorithm has described above that sensitivity selection of initial seeds could influence final output and easy to fall into local optimum. In order to avoid the premature convergence of *K*-means clustering, we considered a genetic algorithm as the optimization tool for evolving initial seeds in the first step of *K*-means process in order to identify optimal partitions. In our study, a chromosome with  $k$  genes is designed for  $k$  cluster centers as  $(x_1, x_2, \dots, x_k)$ , where  $x_i$  is a vector with  $n$  dimensions. During the evolution process, we applied fitness function to evaluate the quality of solutions that is:

$$f(\text{chromosome}) = \sum_{x_j \in X} \min_{1 \leq i \leq k} (\text{dist}(C_i, x_j)) \quad (2)$$

The fitness value is the sum of distances for all inner points to their cluster centers and tries to minimize the values which correspond to optimized partitions. In every successive iteration, three genetic operators precede to construct new populations as offspring according to the fittest survival principles. The populations tend to converge to an optimum chromosome (solution) when

the fitness criterion is satisfied. Once the optimal cluster centers have come out, we use them as initial seeds to perform  $K$ -means algorithm in the last step of clustering. Pseudo-code of the hybrid GA-KM approach is presented as follows, and other configuration parameters will be pointed in Fig 2.

**Algorithm: GA-KM Pseudo-code**

Initialization:

Parameters Initialization:

Maximum Iterations  $Tmax = 200$  and iteration  $t=0$ ;

Population Size  $popnum = 50$ ;

Clusters Number:  $k$ ;

Probability of Crossover:  $Pc$ ;

Probability of Mutation:  $Pm$ ;

Fitness function: minimize the total distance of every sample to its nearest center

$$f(chromosome) = \sum_{x_j \in X} \min_{1 \leq i \leq k} (dist(C_i, x_j))$$

Population Initialization:

Generate the initial population  $P(0)$  randomly, each individual consists of  $k$  centers;

While (  $t < Tmax$  )

{

For  $i = 1$  to  $popnum$

$fi = f(chromosomei)$ ;

End

Optimization Reserved for Each Population;

Selection Operator:

Roulette Selection;

Crossover Operator:

Select  $\alpha \in [1, k]$  randomly;

Crossover parent chromosomes with probability  $Pc$ :

Parent chromosomes  $C_{i1}, C_{i2}$ , new pair of Children chromosomes  $C_{i1}', C_{i2}'$ :

$C_{i1}' = [C_{i1} (1: \alpha, :); C_{i2} (\alpha+1: end, :)]$

$C_{i2}' = [C_{i2} (1: \alpha, :); C_{i1} (\alpha+1: end, :)]$

Compare  $C_{i1}', C_{i2}'$  with  $C_{i1}, C_{i2}$ , if the children chromosomes are better than the parents, the parents will be replaced with the children;

Replace the worst two chromosomes in population with  $C_{i1}', C_{i2}'$ ;

Mutation Operator:

Mutate each center of the best chromosomes with probability  $Pm$  respectively:

For each center  $C_i$ :

Replace  $C_i$  with random center in all samples;

Compare  $C_{new}$  with  $C_i$ , if it is better,  $C_i$  will be replaced with  $C_{new}$ ;

$t = t + 1$ ;

}

Get the initial  $k$  centers with the optimal fitness value.

$K$ -means Optimization:

Generate new clusters and new  $k$  centers;

Fig 2. GA-KM clustering algorithm procedure

The offline model was constructed by our PCA-GAKM approach, and once the target user is reached, we calculate the most interesting movies as TOP-N recommendation list online from a cluster neighborhood instead of searching the whole user space. The estimated rating value for an un-rated movie given by  $U_a$  is predicted as follows [27]:

$$P_{Ua, item} = \overline{R_u} + \frac{\sum_{y \in C_x} sim(U_a, y) \times (R_{y, i} - \overline{R_y})}{\sum_{y \in C_x} (|sim(U_a, y)|)} \quad (3)$$

where  $\overline{R_u}$  is average rating score given by  $U_a$ ,  $C_x$  is a set of neighbors belonging to one common cluster with

$U_a$ ,  $\overline{R_y}$  denotes the average rating given by  $U_a$ 's neighbor  $y$ ,  $\text{sim}(U_a, y)$  is a similarity function based on Pearson correlation measure to decide the similarity degree between two users.

#### IV. EXPERIMENTS AND RESULTS

In this section, we describe the experimental design, and empirically investigate the proposed movie recommendation algorithm via PCA-GAKM technique and compare its performance with benchmark clustering-based CF. Finally the results will be analyzed and discussed. We carried out all our experiments on Dual Xeon 3.0GHz, 8.0GB RAM computer and run Matlab R2011b to simulate the model.

##### A. Data Set and Evaluation Criteria

We consider the well-known Movielens dataset to conduct the experiments, which is available online, including 100,000 ratings by 943 users on 1,682 movies, and assigned to a discrete scale of 1-5. Each user has rated at least 20 movies. We use  $\phi$  to describe the sparsity level of dataset:  $\phi_{ml}=1-100,000/943 \times 1682=0.9369$ . Then the dataset was randomly split into training and test data respectively with a ratio of 80%/20%. We utilized training data to build the offline model, and the remaining data were used to make prediction. To verify the quality of recommendation, we employed the mean absolute error (MAE), precision, recall as evaluation measures which have been widely used to compare and measure the performance of recommendation systems. The MAE is a statistical accuracy metric which measures the average absolute difference between the predicted ratings and actual ratings on test users as shown in Eq.(4). A lower MAE value corresponds to more accurate predictions.

$$MAE = \frac{\sum |\tilde{P}_{i,j} - r_{i,j}|}{M} \quad (4)$$

where  $M$  is the total number of predicted movies,  $\tilde{P}_{i,j}$ , represents the predicted value for user  $i$  on item  $j$ , and  $r_{i,j}$  is the true rating.

To understand whether users are interested with the recommendation movies, we employ the precision and recall metrics which are widely used in movie recommender systems to evaluate intelligence level of recommendations. Precision is the ratio of interesting movies retrieved by a recommendation method to the number of recommendations. Recall gives the ratio of interesting movies retrieved that is considered interesting. These two measures are clearly conflict in nature because increasing the size of recommended movies  $N$  leads to an increase in recall but decrease the precision. The precision and recall for  $Top-N$  ( $N$  is the

number of predicted movies) recommendation are defined in (5) & (6) respectively.

$$\text{precision} = \frac{|\text{interesting} \cap \text{TopN}|}{N} \quad (5)$$

$$\text{recall} = \frac{|\text{interesting} \cap \text{TopN}|}{|\text{interesting}|} \quad (6)$$

##### B. Experimental Designs

We try to conduct different clustering algorithms –  $K$ -means, SOM, and the proposed GA-KM on a relatively low dimension space after PCA transformation.  $K$ -means is easy with efficiency, but sensitive for initial cluster and often convergence to a local optimum. SOM, as an artificial neural network, has been applied to many intelligent systems for its good performance. In our GA optimal processing, we use Euclidean distance measure to decide the similarity of  $n$ -dimensional vectors in search space. The initial population is generated with the size of 50 and number of generation:  $Tmax=200$ . Parameter  $N$  represents the number of movies on the recommendation list. To decide the cluster number  $K$  suitably, we first make a robust estimation on unrated scores by performing global  $K$ -means clustering operations on dataset where  $K$  varies from 4 to 28. It has been seen in Fig 3 that smaller MAE values occurs when  $K$  is between 12 and 18. Thus we set  $K$  to 16 as the total number of clusters to guide our numerical experiments.

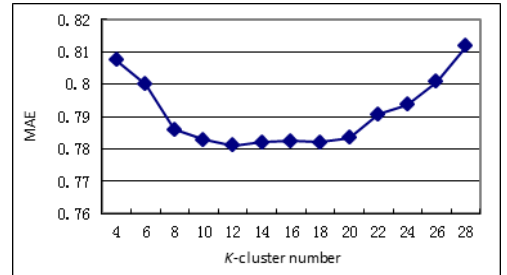


Fig 3. Precision error on different cluster numbers

In this part, we use an all-but-10 method, in which we randomly held out ten ratings for each user in test data that should be predicted based on model. Prediction is calculated for the hidden votes by three clustering algorithms to compare the recommendation accuracy. In addition, to examine the prediction precision, we employed UPCC (Pearson Correlation Coefficient based CF) and up-to-date clustering-based CF methods to compare with the proposed hybrid clustering approach.

We also conduct experiments to verify quality of the new hybrid model. Here cold-start users are defined as users who have rated less than 5 movies. We get five ratings visible for each test user; the rest of ratings

replaced with null, and tried to prediction their values. Given5 data is designed to test cold-start problem caused by a new user with little history information. To normal target users, we also build Given10 to Given20 and Allbut5 to Allbut10 test data to generate recommendation. In each above experiment, we repeat five times for randomly training and test datasets, and average the results.

### C. Results and Discussion

The sparse of user-item rating matrix makes it hard to find real neighbors to form the final recommendation list. In our experiments, we compare the performances and some trends of the existing baseline CF movie recommendation systems with our approach, while the neighborhood size varies from 5-60 in an increment of 5. Detail explanation is showed as the follows from experiment results:

#### 1) Performance of PCA-GAKM CF approach

We first try to evaluate the movie recommendation quality with the traditional cluster-based CFs. Fig 4 shows that all methods tries to reach the optimum prediction values where the neighborhood size varies from 15-20, and it becomes relatively stable around 60 nearest neighbors. All clustering with PCA algorithms performed better accuracy than pure cluster-based CFs. We consider that PCA process could be necessary to dense the original user-rating space, and then improve the partition results. Without the first step of dimensional reduction, GAKM and SOM gave very close MAE values and it seems that GAKM produce slightly better prediction than SOM. When coupling with PCA technique, GAKM shows a distinct improvement on recommendation accuracy compared with SOM. Moreover, the proposed PCA-GAKM performs apparently high accuracy among all the algorithms, and produces the smallest MAE values continually where the neighbor size varies. All *K*-means clustering CF generate increasing MAE values which indicate the decreasing quality for recommendation due to sensitiveness of the algorithm. Traditional user-based CF produces relatively worse prediction compared with the basic clustering-based methods.

To exam the difference of predictive accuracy between our proposed method and other comparative cluster-based methods, we applied *t*-test in the recommendation results. As shown in Table 1, the differences between MAE values are statistically significant at the 1% level. Therefore, we can affirm that the proposed PCA-GAKM outperforms with respect to the comparable cluster-based methods.

TABLE I. THE T-TEST RESULTS FOR VARIOUS CLUSTER-BASED METHODS IN TERMS OF MAES.

Method	Mean	Std. dev	PCA-GAKM	
			<i>t</i> -Value	Sig*
PCA-SOM	0.8018	5.900e-03	9.0633	.000
SOM-CLUSTER	0.8078	2.589e-03	16.5194	.000
UPCC	0.8232	1.292e-03	28.9869	.000
KMEANS-CLUSTER	0.8175	2.244e-03	23.3691	.000
PCA-KMEANS	0.8412	2.845e-03	37.0515	.000
GAKM-CLUSTER	0.8040	2.786e-03	13.8541	.000
<b>PCA-GAKM</b>	<b>0.7821</b>	<b>4.747e-03</b>		

\*Statistically significant at the 1% level.

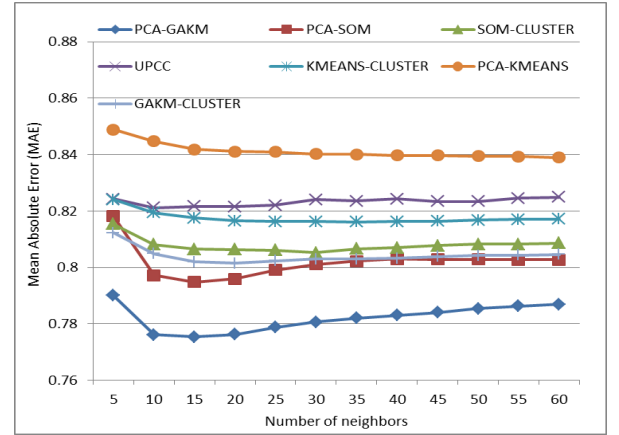


Fig 4. Comparing accuracy with existing clustering-based CF

#### 2) Precision of PCA-GAKM CF approach

To analyze precision of recommendation, we fix the neighbor size  $n=20$ . As seen from Fig 5, the overall precisions improve with the increasing number of recommendation, and the PCA-GAKM generates higher precision rates which indicate that it can recommend more interesting and reliable movies to users than other clustering-based algorithms when a relatively small number of movies on recommendation list are considered. In addition, Fig 6 compares the recall rates of user interesting movies, and it's apparently that PCA-GAKM still provide greater recall rates with each value in  $N$  (the number of recommendation). The existing cluster-based CFs show lower precision and recall rates comparing to our optimal clustering approach.

#### 3) Recommendation of cold-start users

We finally experiment the cold-start problem with "less information" users who has rated few movies in history. It is understandable that searching neighbors in high dimensional space become difficult for cold-start users with few ratings. Fig 7 enables us to discover that clustering coupled with PCA methods may produce a generalized improvement in prediction accuracy for cold-start users. Among examined clustering methods, the proposed PCA-GAKM seems to have the best

performance in alleviating cold-start with the satisfactory MAE values. With increasing number of ratings used to make prediction, all approaches show the similar trend that prediction accuracy is getting higher as presented in Fig 7.

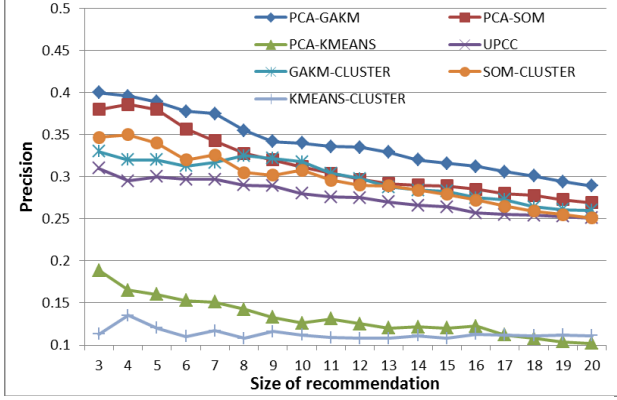


Fig 5. Precision comparison with existing cluster-based CF

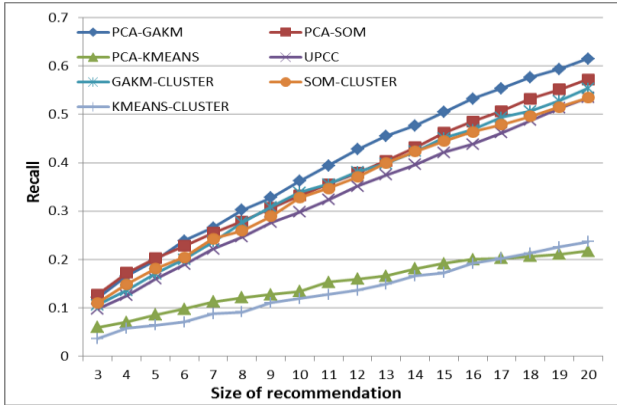


Fig 6. Recall comparison as the recommendation size grows

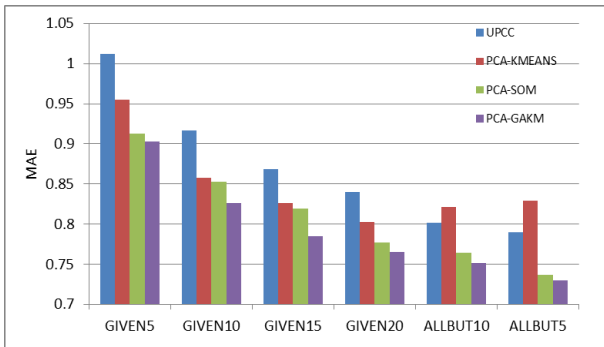


Fig 7. MAE comparisons in different rating reveal level

## V. CONCLUSIONS AND FUTURE WORK

In this paper we develop a hybrid model-based CF approach to generate movie recommendations which combines dimensional reduction technique with

clustering algorithm. In the sparse data environment, selection of “like-minded” neighborhood on the basis of common ratings is a vital function to generate high quality movie recommendations. In our proposed approach, feature selection based on PCA was first performed on whole data space, and then the clusters were generated from relatively low dimension vector space transformed by the first step. In this way, the original user space becomes much denser and reliable, and used for neighborhood selection instead of searching in the whole user space. In addition, to result in best neighborhood, we apply genetic algorithms to optimize *K*-means process to cluster similar users. Based on the Movielens dataset, the experimental evaluation of the proposed approach proved that it is capable of providing high prediction accuracy and more reliable movie recommendations for users’ preference comparing to the existing clustering-based CFs. As for cold-start issue, the experiment also demonstrated that our proposed approach is capable of generating effective estimation of movie ratings for new users via traditional movie recommendation systems.

As for future work, we will continue to improve our approach to deal with higher dimensionality and sparsity issues in practical environment, and will explore more effective data reduction algorithms to couple with clustering-based CF. Furthermore, we will study how the variation number of clusters may influence the movie recommendation scalability and reliability. To generate high personalized movie recommendations, other features of users, such as tags, context, and web of trust should be considered in our future studies.

## ACKNOWLEDGEMENTS

The research was supported by the National Natural Science Foundation of China (Nos. 71271149, 70901054 and 61202030). It was also supported by the Program for New Century Excellent Talents in University (NCET). The authors are very grateful to all anonymous reviewers whose invaluable comments and suggestions substantially helped improve the quality of the paper.

## REFERENCES

- [1] G. Adomavicius, A. Tuzhilin, “Toward the next generation of recommender system: A survey of the state-of-the-art and possible extensions,” *IEEE Trans on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.
- [2] G. Linden, B. Smith and J. York, “Amazon.com recommendations: Item to item collaborative filtering,” *IEEE Internet Computing*, vol.7, no.1, pp. 76–80, 2003.
- [3] B. M. Sarwar, G. Karypis, J. Konstan and J. Riedl, “Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering,” in *Proceedings of the international conference on computer and information technology*, 2002.

- [4] B. M. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithm," in Proceedings of the 10<sup>th</sup> International World Wide Web Conference. Hong Kong, 2001, pp. 285–295.
- [5] B. M. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Application of Dimensionality Reduction in Recommender System-A Case Study," in ACM 2000 KDD Workshop on Web Mining for e-commerce-Challenges and Opportunities, 2000.
- [6] J. S. Breese, D. Heckerman and C. Kadie, "Empirical analysis of Predictive Algorithms for collaborative filtering," in Proceedings of the 14<sup>th</sup> Conference on Uncertainty in Artificial Intelligence, 1998, pp. 43–52.
- [7] G. Xue, C. Lin, Q. Yang, et al, "Scalable Collaborative Filtering Using Cluster-based Smoothing," in Proceedings of the 28<sup>th</sup> annual international ACM SIGIR conference on Research and development in information retrieval, Brazil: ACM Press, 2005, pp. 114–121.
- [8] F. Gao, C. Xing and Y. Zhao, "An effective algorithm for dimensional reduction in collaborative filtering," in Lecture Notes on Computer Science, Springer Berlin, 2007, pp. 75–84.
- [9] Q. Li and B. M. Kim, "Clustering approach for hybrid recommendation system," in Proceedings of the International Conference on Web Intelligence, 2003, pp. 33–38.
- [10] K. Kim and H. Ahn, "A recommender system using GA K-means clustering in an online shopping market," Expert Systems with Application, vol. 34, no. 2, pp. 1200–1209, 2008.
- [11] A. Kohrs and B. Merialdo, "Clustering for collaborative Filtering Applications," in Proceedings of CIMCA'99, Vienna: IOS Press, 1999, pp. 199–204.
- [12] D. Goldberg, D. Nichols, B. M. Oki and D. Terry, "Using collaborative filtering to weave an information tapestry," Communications of the ACM, vol. 35, no. 12, pp. 61–70, 1992.
- [13] D. Billsus and M. J. Pazzani, "Learning collaborative information filters," in Proceedings of the 15th International Conference on Machine Learning, Madison, 1998, pp. 46–53.
- [14] K. Goldberg, T. Roeder, D. Gupta and C. Perkins, "Eigentaste: A constant time collaborative filtering algorithm," Information Retrieval, vol. 4, no. 2, pp. 133–151, 2001.
- [15] K. Q. Truong, F. Ishikawa and S. Honiden, S, "Improving accuracy of recommender system by item clustering," Transactions on Information and Systems, E90-D(9), pp. 1363–1373, 2007.
- [16] G. Pitsilis, X. L. Zhang and W. Wang, "Clustering recommenders in collaborative filtering using explicit trust information," in Proceedings of the 5<sup>th</sup> IFIP WG 11.11 International Conference on Trust Management, vol.358, 2011, pp. 82–97.
- [17] M. Zhang and N. Hurley, "Novel item recommendation by user profile partitioning," in Proceedings of the IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, 2009, pp. 508–515.
- [18] C. Huang and J. Yin, J, "Effective association clusters filtering to cold-start recommendations," in Proceedings of the 7<sup>th</sup> International Conference on Fuzzy Systems and Knowledge Discovery, 2010, pp. 2461–2464.
- [19] J. Wang, N.-Y. Zhang, J. Yin, J, "Collaborative filtering recommendation based on fuzzy clustering of user preferences," in Proceedings of the 7<sup>th</sup> International Conference on Fuzzy Systems and Knowledge Discovery, 2010, pp. 1946–1950.
- [20] D. -R. Liu, Y. -Y. Shih, "Hybrid approaches to product recommendation base on customer lifetime value and purchase preferences," Journal of Systems and Software, vol. 77, no. 22, pp. 181–191, 2005.
- [21] O. Georgiou and N. Tsapatsoulis, "Improving the scalability of recommender systems by clustering using genetic algorithms," in Proceedings of the 20<sup>th</sup> Int. conf. Artificial Neural Networks, 2010, pp. 442–449.
- [22] K. Honda, N. Sugiura, H. Ichihashi and S. Araki, "Collaborative filtering using principal component analysis and fuzzy clustering," in Proceedings of the 20th International Conference on Artificial Neural Networks: Part I, 2001, pp. 442–449.
- [23] A. Selamat and S. Omatu, "Web page feature selection and classification using neural networks," Information Science, vol.158, pp. 69–88, 2004.
- [24] J. Han and M. Kamber, Data Mining: Concepts and Techniques. San Francisco CA: Morgan Kaufmann Publishers, 2001.
- [25] J.T. Tou, R.C. Gonzalez, Pattern Recognition Principle. Massachusetts: Addison Wesley, 1974.
- [26] D. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning. New York: Addison-Wesley, 1989.
- [27] P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom and J. Riedl, "GroupLens: An open architecture for collaborative filtering of netnews," in Proceedings of the 1994 Computer Supported Cooperative Work Conference, 1994, pp. 175–186.



# An Original-Stream Based Solution for Smoothly Replaying High-definition Videos in Desktop Virtualization Systems

Kui Su, Zonghui Wang\*, Xuequan Lu, Wenzhi Chen  
College of Computer Science  
Zhejiang University, Hangzhou, 310027, China  
Email: sukuias12@zju.edu.cn

## Abstract

*How to smoothly replay high-definition (HD) videos in desktop virtualization systems has been a much needed yet challenging problem. Existing desktop virtualization systems apply well to classic office-applications but offer a very limited performance for multimedia applications, especially for replaying HD videos. In existing solutions, video is decoded on the server, and then decoded video data is highly compressed before being delivered to the client. However, high compression ratio requires a large amount of processing power and causes response delay, poor video quality and dropped frames that greatly deteriorate user experience. Although some solutions have been optimized for video replay, they are forced to modify media players and only support specific video formats. Therefore, we propose an original-stream based solution to provide good user experience for replaying HD videos in desktop virtualization systems without any modification on applications as well as support most of prevalent HD video formats. In our solution, video content is directly delivered to the client in its originally encoded state, and the actual video decoding and rendering are executed by the client's GPU. The experimental results validate our method and show that this proposed approach measurably outperforms state-of-the-art solutions.*

## 1 Introduction

As an emerging trend, virtualization [1, 2, 3] has been widely used in cloud computing [4, 5, 6] over the past decade. Among those virtualization applications, desktop virtualization has become an important branch [7, 8]. In desktop virtualization environment, all applications and operating system code are executed in a server which lies in a

remote data center. End user only needs a thin client which handles display, keyboard and mouse combined with adequate processing power for graphical rendering and network communication. The client no longer has to keep user state and communicate with server by using a remote protocol. The protocol allows graphical displays to be virtualized, and transmit user input from the client to the server [9]. Many productive desktop virtualization systems have been developed and applied to various commercial applications since they provide a lot of advantages for IT enterprises such as reducing maintenance and operating costs and improving resource utilization efficiency.

However, existing desktop virtualization systems still suffer from a number of problems before being widely applied: they can not provide high fidelity display and good interactive experiences for end users, especially on multimedia applications which are commonly used in desktop computing. Current remote display protocols such as Remote Framebuffer protocol (RFB) [10] and Remote Desktop Protocol (RDP) [11] are widely used in desktop virtualization systems [12]. They are mainly designed for low-motion graphical applications, such as text editors whose graphic changes are minor with low frequency. However, those protocols cannot effectively support high-motion scenarios such as video playback and real-time interactions. First, because the transport of multimedia data over those protocols is inefficient, requiring high bandwidth to ensure the delivery of all frames to the client in real time. Second, intensive computation for video decoding imposes a heavy burden on server's CPU, which greatly decreases the overall performance of a desktop virtualization system with increasing clients. The problems become even worse when it comes to replaying high definition (HD) video [13] which has a much larger amount of data than standard definition (SD) video. Real-time re-encoding of the video data can definitely save bandwidth but it is computationally expensive, even with modern CPUs, and it causes high response delay, poor video quality and dropped frames that greatly deteriorate user experience.

---

\*This research is funded by National Natural Science Foundation of China under grant NO. 60970125.

To address the existing problems, we propose an original-stream based solution for replaying HD videos in desktop virtualization systems, which is named HDR for the sake of simplicity and readability. Combined with virtualization technology, HDR provides great user experiences of replaying HD videos without any changes on applications or the window system. The HDR prototype is implemented in Virtualbox [14], an open source virtualization software. The experimental results show that our system could reach almost 100% video quality and full frame rate in full screen on HD video playback in both 100 Mbps and 10 Mbps network environments while classic systems only achieve no more than 20% quality and very low frame rate which is not enough to replay HD videos smoothly. We have tested most of prevalent video formats such as H264, MPEG-2 and VC-1 with generally used resolutions for HD videos such as 720P, 1080i and 1080P. Besides, a number of popular media players have been tested in our experiments. The results show that all the tested video formats and media player applications can be well supported in HDR while some other systems only support specific video formats and media player applications. Additionally, our solution greatly reduces both server and client's CPU usage by using GPU-accelerated video decoding technology [15].

## 2 Related Work

Many productive desktop virtualization systems have been developed and applied to various commercial applications since they provides a lot of advantages for IT enterprises such as reducing maintenance and operating costs and improving resource utilization efficiency. VNC [10] and THINC [16] are famous thin-client systems proposed in academic research while in industry there are Microsoft Remote Desktop [11], Citrix XenDesktop [17], VMware View [18], Sun Ray and HP Remote Graphics and so on. However, most of them cannot provide a satisfactory performance for replaying HD videos.

VNC (Virtual Network Computing) is a popular remote display system with RFB protocol. It uses a virtual driver to maintain local copy of the framebuffer state used to refresh its display and forward user input directly to the server. VNC provides a good performance for office applications but not for video, because the "Client-Pull" mode of screen update in VNC is very sensitive to network latency. It takes encoding time, data transmission time and round trip latency time for every frame to be fully processed that it is not suitable for frequently updated video replay.

THINC and its portable version pTHINC intercepts low-level video driver commands and adopts a push mode to interact with client. Its codec is efficient for UI compression but suffers from compression performance degradation over multimedia content encoding. As a result, it can achieve

a great multimedia playback performance with sufficient bandwidth but not for network environments with low bandwidth.

RDP (Remote Desktop Protocol) is widely used in desktop virtualization products such as Microsoft RDS and VMware view. For office applications, such as a text editor or a spread-sheet, RDP is highly optimized and the display changes are quite small and have a sufficiently low frequency to cope with. However, with the emergence of multimedia applications, existing remote display protocol cannot reach the high levels of crisp. As a result, RDP provides a poor performance for video replay. In recent years, multimedia applications has been playing a significant role in remote display, for example, Microsoft is taking much more efforts to optimize RDP and VMware is trying to find a more efficient solution for its desktop virtualization systems.

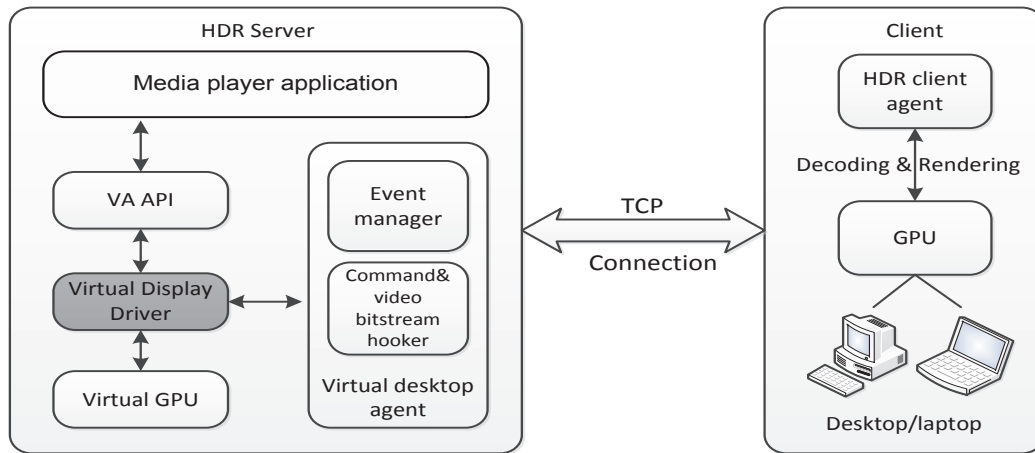
For the past few years, popular commercial products Xen Desktop and RemoteFx [19] have devoted much effort to the optimization for video replay. They try to deliver video files to the client before they are decoded on the server, this method will achieve good video quality with low bandwidth, but the main drawback is that they capture the video stream from application layer by taking advantage of Windows media foundation so that it only supports certain video formats and media player applications that use the necessary Windows media framework. Considering multitudinous video formats and media player applications and system scalability, these solutions are inappropriate to be used in desktop virtualization systems.

## 3 Design of Architecture

We propose HDR, an original-stream based method to solve existing problems of replaying HD videos in desktop virtualization systems. In the HDR, encoded video content and decoding API calls are intercepted from the display driver of the server and delivered to the client through the network. The client re-executes the API calls to display the video on the screen. The intercepted video content is in its originally encoded state, which means that the video has never been decoded in the server. The proposed method has two main goals: i) improving user experience of watching HD videos remotely in terms of video quality, fluency, bandwidth and generality and ii) an obvious reduction on server and client's CPU utilization and an exclusive use of server's GPU.

### 3.1 Overview of HDR

Figure 1 shows the overview of the HDR. The clients are connected to the desktop virtualization server through ethernet or wireless network. User's applications are executed in Guest OS virtualized by the server. A HD video file played



**Figure 1. Overall architecture of HDR**

by a media player of the server will be eventually displayed on the client's screen. Generally, media players call video acceleration (VA) APIs to leverage GPU-accelerated video decoder for better decoding a HD video. Then the encoded video content is passed to the display driver and finally decoded and rendered by GPU. But in the HDR, the video content and API calls will be intercepted in the display driver of the server and delivered to the client by the virtual desktop agent. The virtual desktop agent composes of commands & video bitstream hooker and event manager. Actually the hooker is implemented by modifying the display driver, which is used to intercept video content and API calls. Then the event manager delivers the intercepted data to the client through the network. The hooker and the event manager communicate with each other by shared memory. On the client side, the client agent re-executes the API calls from the server to display the video on the screen.

In the workflow of HDR, there is no need to compress the video data for transmitting to the client, because the HD video is not decoded on the server and it is still at the originally encoded state (e.g. H264, MPEG2, VC-1) which is very suitable for transmitting. No compression and decompression mean little quality loss, low response latency and low CPU usage. Also the GPU of the server is not used. Moreover, HDR intercepts the video data and commands from the display driver layer which is transparent to applications so that it is able to work seamlessly with existing applications without any modifications.

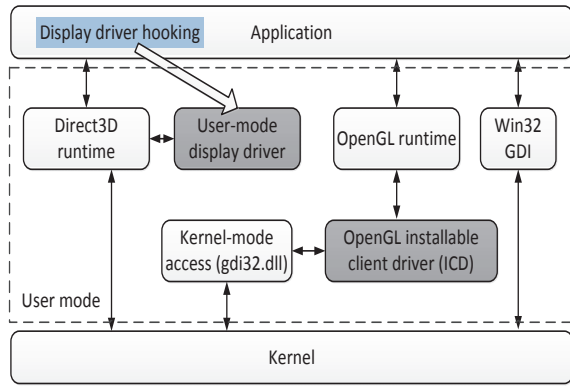
### 3.2 Original-stream based streaming

In the HDR, we deliver HD video content in its originally encoded state instead of highly compressed decoded video data to the client through the network for two reasons: Firstly, HD videos produced by cameras are always

encoded to H264 [20], MPEG-2 or VC-1 formats due to the large amount of data so that they can be easily stored and transported in computer systems. For a desktop virtualization system, it is very convenient to deliver and process HD videos in those formats. Secondly, in a desktop virtualization system, the objective is to enable client users to achieve the same desktop experiences as in the local PCs. In traditional solutions, HD videos are decoded on the server and the decoded video data is highly compressed to reduce bandwidth before transmitted to the client. On the client, the compressed video data has to be decompressed before displayed on the screen. The complicated process on video data greatly damages the quality of HD videos, causes high response delay and consumes more CPU power. Therefore, HDR transmits HD video content in its originally encoded state. In the HDR, high compression is not needed for video data on the server and decompression is also eliminated on the client, HD video can be replayed on the client at its full quality with ideal frame rate. User can achieve the same experience as that in the local PCs with low CPU usage of both the server and client.

### 3.3 Driver-based video hooking

In the previous subsection, we discuss that it is much better to deliver video data in its originally encoded state instead of highly compressed decoded video data. In this subsection, we discuss where to intercept the encoded video data. In practice, video data can be intercepted at different layers: media player application, display driver and framebuffer. In framebuffer, the video has been decoded to pixel data which has to be highly compressed for transmitting to client. In application layer, a video file is in the encoded state but often encapsulated into different formats such as MKV, WMV and MP4, etc. Besides, a variety of media



**Figure 2. Video hooking from WDDM**

player applications have been developed such as Windows media player, MPlayer, KMPlayer, etc. Meanwhile, various development frameworks for media player applications such as Windows media foundation (WMF) and DirectShow have been proposed. In a virtual desktop, it is necessary to support most of popular media player applications and video formats. If we intercept the video data from the application layer, we have to modify those different applications and the client agent will become much more complicated. Therefore, the best choice is to intercept the video data from the display driver layer which is transparent to applications, and in the display driver, the video is also in its encoded state. It enables HDR to work seamlessly with existing applications without any modifications.

To intercept the encoded video data from the display driver, we assume that all the media players of the server adopt GPU-accelerated technology to decode and render the video. The GPU-accelerated technology guarantees the encoded video data to be passed to the display driver for GPU decoding. GPU-accelerated video decoding technology is very popular and widely used by most of media player applications because that decoding tasks for HD videos are often computation-intensive which imposes a heavy burden on classic low-end CPUs, and high-end CPU is much more expensive than GPU and consume a lot of power.

## 4 Implementation

We have implemented a prototype server based on Windows 7 system in a virtual environment created by Virtualbox and a client based on Windows 7 system in real physical machines. In the HDR, we have assumed that all the media players of the server adopt GPU-accelerated technology to decode and render HD videos so that the encoded video data can be intercepted from the display driver. In Windows 7 system, DirectX Video Acceleration (DXVA) [21] technology is the most widely used GPU-accelerated

technology. It is a Microsoft API specification for the Microsoft Windows and Xbox 360 platforms that allows video decoding to be GPU accelerated. The pipeline of DXVA allows certain CPU-intensive operations such as IDCT, motion compensation and deinterlacing to be offloaded to the GPU. The DXVA is used by software video decoders to define a codec-specific pipeline for GPU-accelerated decoding and rendering of the codec. The pipeline starts at the CPU which is used for parsing the media stream and conversion to DXVA-compatible structures. DXVA specifies a set of operations that can be hardware accelerated and device driver interfaces (DDIs) that the graphic driver can implement to accelerate the operations.

In the HDR, we intercept encoded video content and decoding API calls By modifying the implementation of DXVA DDIs in server's display driver. However, most of the display drivers especially for Windows OS are commercial proprietary closed, to achieve our goal, we have developed a virtual display driver based on Windows Device Driver Model (WDDM) [22] on the server. The main modifications are made in the user-mode display driver in case of severe effects on system kernel as shown in Figure 2.

In order to deliver the intercepted data from the display driver to the client, the virtual desktop agent takes advantage of shared memory technology for the communication between the event manager and the modified display driver (i.e., the hooker). First, the event manager creates a shared memory region and maps the region to its process space. Then, the driver also maps the same shared region to its process space so that both the event manager process and the driver process can notify each other to read and write the shared memory by holding an event handle. The driver writes the intercepted data to the shared memory and notify the event manager. The event manager reads the data and delivers them to the client through the network.

Generally, media players need to query the ability information of the local GPU before using DXVA to decode HD videos. This is because that DXVA is only available for suitable GPUs. In the HDR, the actual decoding and rendering operations for HD videos are executed by the client's graphic device. To work compatibly with the server, the client should deliver the ability information of its graphic device to the server in advance. Considering various video cards of clients, we adopt a simple and adaptive method to finish this job. Firstly, when a connection between a client and server is built, the client delivers the ability information of its video card to the server and the server will save the information as a local file which shall be valid until the client disconnects with the server. Once a media player begins to query the ability, the information of the file will be extracted to it. This method effectively solves the ability query problem and dynamically adapts diverse client video cards.

## 5 Performance Evaluation

In this section, system performance is evaluated in real applications under different network conditions to demonstrate the effectiveness of HDR. We mainly evaluate the performance of HDR in terms of bandwidth consumption, CPU usage, video quality and frame rate. Several prevalent remote display systems are involved for comparison. They are TightVNC [23], Microsoft Remote Desktop and THINC.

### 5.1 Experimental Setup

In our experiments, we use a 100 Mbps, 1ms latency LAN network to emulate different network conditions. The bandwidth emulated by the widely used network emulator WANem [24] is 100 Mbps and 10 Mbps. The Server machine has a 2.66 GHz Intel Core i7 – 920 processor and 8 GB of RAM. Client\_1 is a 2.0 GHz Intel Core II laptop with 1 GB of RAM, Client\_2 is a thin client with a 1.6 GHZ Intel ATOM N270 processor and 512 MB memory. the client\_1 runs Windows 7 SP1 system and the client\_2 runs Windows Embedded Standard 7 system. For THINC, we use VirtualBox 4.1.16 to run Ubuntu 12.04 system on both server and client hardware. For HDR we use VirtualBox 4.1.16 to run Windows 7 system on server hardware. WANem emulator is also installed on a virtual machine created by VirtualBox 4.1.16 on server hardware. The tested videos include various formats with different resolutions, but in the following, we just give the results of the two H264 HD video clips for the limited space of this paper: (1)Video\_1.avi (1280 \* 720p, 30 fps, H264 codec, time: 119 s); (2)Video\_2.mkv (1920 \* 1080p, 30 fps, H264 codec, time: 96 s). The media player used is Windows media player for Windows system, and for linux we use MPlayer.

### 5.2 Experimental Results

In the following, we introduce our experiments in detail and show the experimental results of all the tested solutions in terms of bandwidth consumption, CPU utilization, video quality and frame rate. Besides, we have also tested that how the FPS in HDR is affected by network delay.

#### 5.2.1 Bandwidth consumption

This experiment is designed to show the detailed bandwidth consumption of each participating system under different network conditions while replaying HD videos of different resolutions. For the 100 Mbps high-bandwidth and 10 Mbps low-bandwidth environments, we compute the average bandwidth consumption during the video replay. We can see the results from Figure 4 (a) and (b), all the solutions consume different bandwidth for HD videos of differ-

ent resolutions. 1080P video\_2 consume much more bandwidth than 720P video\_1. HDR and TightVNC consume much less than other solutions under both 10 Mbps and 100 Mbps, but TightVNC does not consume much bandwidth simply because the quality of video is extremely low.

#### 5.2.2 CPU consumption

**Table 1. CPU utilization for video\_1**

Protocol \ Role	Server	Client_1	Client_2
TightVNC	3.7%	4.9%	21.2%
RDP	12.5%	11.1%	35.4%
THINC	10%	10.4%	38.3%
HDR	3.7%	2.4%	10%

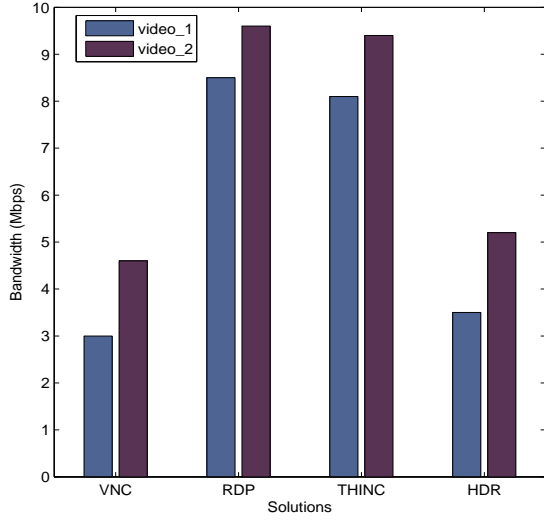
**Table 2. CPU utilization for video 2**

Protocol \ Role	Server	Client_1	Client_2
TightVNC	4.8%	5.6%	22.5%
RDP	15.1%	12.9%	46.8%
THINC	12%	14.5%	44%
HDR	4.1%	3%	13.2%

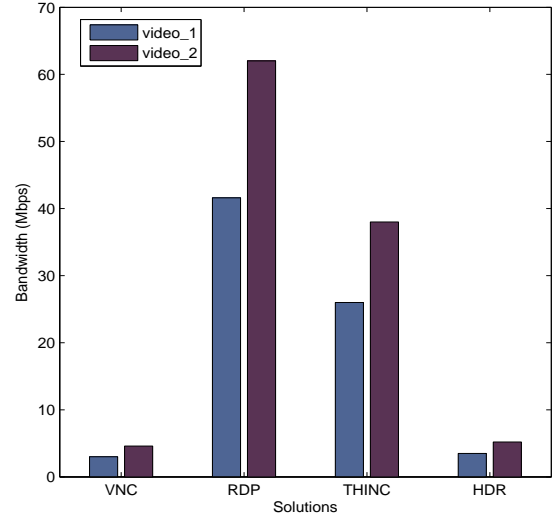
For desktop virtualization systems, CPU consumption is an essential metric for system performance. In this experiment, we have tested the average CPU consumption of video replay on the two clients and the server. Table 1 and Table 2 show the CPU utilization of the participating systems for the two videos, respectively. As shown in the tables, video\_2 consume much more CPU than video\_1. HDR performs better than RDP and THINC, because HDR delivers video data to the client without high compression which costs much CPU resource and the client's CPU does not need to decompress the video data. Additionally, video decoding on the client is executed by GPU and CPU is just used to process network I/O requests in HDR.

#### 5.2.3 Video quality

Video quality is measured by using slow motion technique [25], which takes both playback delays and frame drops into consideration. Define Video Quality (V.Q.), the video quality is calculated according to formula 1. 100% video quality is the optimal quality, which means all video frames are played at real-time speed. Figure 5 (a) and (b) show the results of video quality for the two videos under 100 Mbps and 10 Mbps network environments, respectively. Among the tested solutions, TightVNC provides the worst

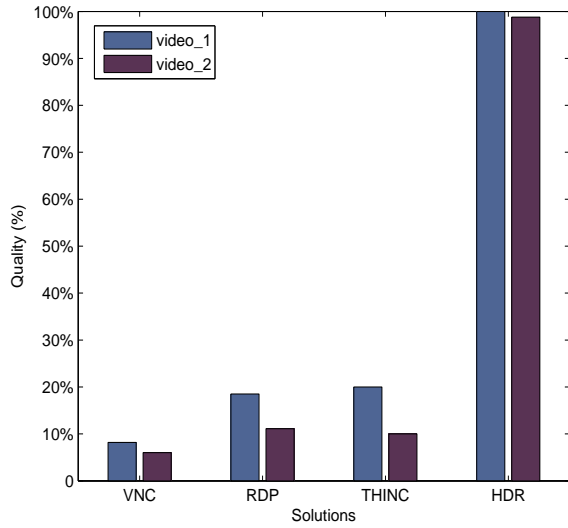


(a) Bandwidth consumption under 10 Mbps network environment

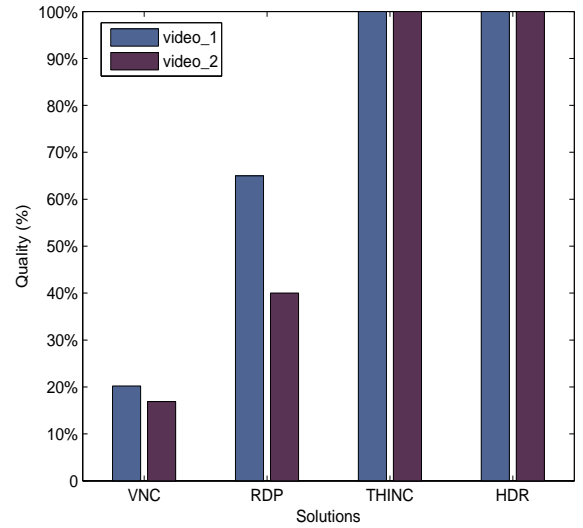


(b) Bandwidth consumption under 100 Mbps network environment

**Figure 3. Bandwidth consumption for different HD videos under different network environments**



(a) Video quality under 10 Mbps network environment



(b) Video quality under 100 Mbps network environment

**Figure 4. Video quality for different HD videos under different network environments**

quality of video. RDP performs better than TightVNC, it can achieve almost 65% for 720P video\_1 but only 40% for 1080P video\_2 under 100 Mbps. It becomes much worse while the network bandwidth is reduced to 10 Mbps. THINC does very well under 100 Mbps but it only achieve the video quality no more than 20% under 10 Mbps even for 720P video\_1. By contrast, HDR provide 100% quality of the two videos under 100 Mbps, and the only quality loss is

for 1080P video\_2 under 10 Mbps. Thus, our proposed solution outperforms all the tested solutions in terms of video quality.

$$V.Q = \frac{\frac{(DataTransferred(30fps))}{(PlaybackTime(30fps))}}{\frac{(DataTransferred(1fps))}{(PlaybackTime(30fps))}} \quad (1)$$



#### 5.2.4 Frames per second (FPS)

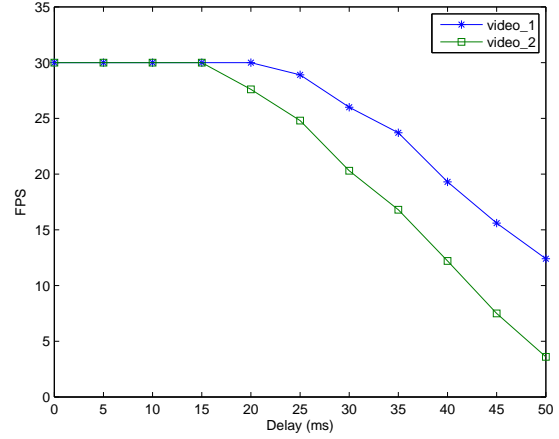
**Table 3. Frame Rate under different network environments**

Protocol Video	TightVNC	RDP	THINC	HDR
10Mb/s ( <i>video_1</i> )	4.1	6.6	11.2	30
10Mb/s ( <i>video_2</i> )	3.7	5.3	6.2	28.8
100Mb/s ( <i>video_1</i> )	6.2	12.4	20.8	30
100Mb/s ( <i>video_2</i> )	5.8	10.1	16.1	30

In this experiment, the evaluated factor is FPS for video replay on the clients, we compute the average FPS during the video replay. FPS is a count of how many pictures a movie displays per second. A frame is a still, visible image. Showing these frames in succession creates the illusion of a motion picture. For *video\_1* and *video\_2*, the full FPS is 30. We have tested these videos in a local PC, all of them can achieve 30 FPS, but in the participating systems, not all of them can make it. As shown in Table 3, HDR can achieve an ideal FPS for both the two videos under 100 Mbps and *video\_1* under 10 Mbps, and it reaches 28.8 even for 1080P *video\_2* under 10 Mbps. However, among the other solutions, only THINC achieves a FPS greater than 20 for 720P *video\_1* under 100 Mbps, all the others have a very low FPS especially for 1080P *video\_2* under 10 Mbps.

#### 5.2.5 Network delay

The above experiments are done under LAN network with 1 ms latency. Although we have limited network bandwidth, latency is not taken into account. In practice, propagation delay between server and client can also affect system performance. The above experimental results show that HDR performs very well under both 100 Mbps and 10 Mbps environments, now we test the performance under different network delay. We use WANem to emulate a 100 Mbps network with different delay to test the average FPS of video replay. As shown in Figure 6, when the delay is lower than 20 ms, the FPS is still ideal, but as the delay increases, the FPS decreases rapidly. When the delay is more than 50 ms, the FPS is too low for users to watch HD videos smoothly. Though the above experiment results indicate that HDR is an excellent solution for replaying HD videos in desktop virtualization systems and performs much better than existing solutions, high network latency limits the performance due to the large amount data of HD videos.



**Figure 5. How FPS is affected by network delay in HDR**

## 6 Conclusion

Desktop virtualization has been widely applied and multimedia applications play a significant role in it. Existing desktop virtualization systems provide good performance for general-purpose applications but still have some challenges for multimedia applications, especially for replaying HD videos which have large amount of data and frequent display updates. We introduce HDR which transmits video content to the clients in its originally encoded state so that video is replayed on the client with ideal FPS, none of quality loss, low CPU cost and network bandwidth. The encoded video content is intercepted from display driver layer on the server which enables HDR to work seamlessly with unmodified media player applications, do not depend on any multimedia framework such as Windows media framework and support most of prevalent video formats and HD resolutions. Besides, HDR uses an adaptive method to dynamically adapts various clients with different video cards.

We have measured HDR's performance on HD video replay in terms of bandwidth consumption, CPU usage, video quality and frame rate under 100 Mbps and 10 Mbps network environments and make a comparison with classic remote display systems. From our experimental results, we can see that HDR costs much less resource (CPU, network bandwidth) and provides better user experience (FPS, video quality) than other systems. It shows that HDR is a very favorable method, which far outperforms other state-of-the-art methods. In HDR, client users will achieve the same good experience as that in a local PC. However, there are still some limitations in our prototype system. In the future, we plan to expand HDR to more client devices such as smartphones and PDAs. We will also make more optimiza-

tions to reduce the complexity of client agents and lower the processing time on the server to adapt for network environments with high latency.

## References

- [1] Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, Tim Harris, Alex Ho, Rolf Neugebauer, Ian Pratt, and Andrew Warfield. Xen and the art of virtualization. *ACM SIGOPS Operating Systems Review*, 37(5):164–177, 2003.
- [2] Rich Uhlig, Gil Neiger, Dion Rodgers, Amy L Santoni, Fernando CM Martins, Andrew V Anderson, Steven M Bennett, Alain Kagi, Felix H Leung, and Larry Smith. Intel virtualization technology. *Computer*, 38(5):48–56, 2005.
- [3] Irfan Habib. Virtualization with kvm. *Linux Journal*, 2008(166):8, 2008.
- [4] Peter Mell and Tim Grance. The nist definition of cloud computing. *National Institute of Standards and Technology*, 53(6):50, 2009.
- [5] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, et al. A view of cloud computing. *Communications of the ACM*, 53(4):50–58, 2010.
- [6] Ajay Gulati, Ganesh Shanmuganathan, Anne Holler, and Irfan Ahmad. Cloud-scale resource management: challenges and techniques. In *Proceedings of the 3rd USENIX conference on Hot topics in cloud computing*, pages 3–3. USENIX Association, 2011.
- [7] Karissa Miller and Mahmoud Pegah. Virtualization: virtually at the desktop. In *Proceedings of the 35th annual ACM SIGUCCS fall conference*, pages 255–260. ACM, 2007.
- [8] Xiaofei Liao, Hai Jin, Liting Hu, and Haikun Liu. Towards virtualized desktop environment. *Concurrency and Computation: Practice and Experience*, 22(4):419–440, 2010.
- [9] Jiewei Wu, Jiajun Wang, Zhengwei Qi, and Haibing Guan. Sridesk: A streaming based remote interactivity architecture for desktop virtualization system. In *Computers and Communications (ISCC), 2013 IEEE Symposium on*, pages 000281–000286. IEEE, 2013.
- [10] Tristan Richardson, Quentin Stafford-Fraser, Kenneth R Wood, and Andy Hopper. Virtual network computing. *Internet Computing, IEEE*, 2(1):33–38, 1998.
- [11] Windows remote desktop protocol (RDP), <http://msdn.microsoft.com/en-us/library>.
- [12] Charles Border. The development and deployment of a multi-user, remote access virtualization system for networking, security, and system administration classes. In *ACM SIGCSE Bulletin*, volume 39, pages 576–580. ACM, 2007.
- [13] Hd-video, [http://en.wikipedia.org/wiki/high-definition\\_video](http://en.wikipedia.org/wiki/high-definition_video).
- [14] Virtualbox, [http://www.virtualbox.org/wiki/vbox\\_vs\\_others](http://www.virtualbox.org/wiki/vbox_vs_others).
- [15] Guobin Shen, Guang-Ping Gao, Shipeng Li, Heung-Yeung Shum, and Ya-Qin Zhang. Accelerate video decoding with generic gpu. *Circuits and Systems for Video Technology, IEEE Transactions on*, 15(5):685–693, 2005.
- [16] Ricardo A Baratto, Leonard N Kim, and Jason Nieh. Thinc: a virtual display architecture for thin-client computing. In *ACM SIGOPS Operating Systems Review*, volume 39, pages 277–290. ACM, 2005.
- [17] Hdx of citrix xendesktop, <http://hdx.citrix.com/hdx>.
- [18] Vmwareview, <http://www.vmware.com/products/horizon-view>.
- [19] Remotefx of microsoft, <http://technet.microsoft.com/>.
- [20] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay Luthra. Overview of the h. 264/avc video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7):560–576, 2003.
- [21] Directx video acceleration (DXVA), <http://msdn.microsoft.com/en-us/library/windows/desktop>.
- [22] Windows display driver model (WDDM), <http://msdn.microsoft.com/en-us/library>.
- [23] Tightvnc, <http://en.wikipedia.org/wiki/tightvnc>.
- [24] Hemanta Kumar Kalitay and Manoj K Nambiar. Designing wanem: A wide area network emulator tool. In *Communication Systems and Networks (COMSNETS), 2011 Third International Conference on*, pages 1–4. IEEE, 2011.
- [25] Jason Nieh, S Jae Yang, and Naomi Novik. Measuring thin-client performance using slow-motion benchmarking. *ACM Transactions on Computer Systems (TOCS)*, 21(1):87–115, 2003.

# Fostering Participation and Co-Evolution in Sentient Multimedia Systems

Federico Cabitza  
Dip. di Informatica  
Università di Milano-Bicocca  
Milano, Italy  
cabitza@disco.unimib.it

Daniela Fogli  
Dip. di Ingegneria dell'Informazione  
Università di Brescia  
Brescia, Italy  
fogli@ing.unibs.it

Antonio Piccinno  
Dip. di Informatica  
Università di Bari "Aldo Moro"  
Bari, Italy  
antonio.piccinno@uniba.it

**Abstract**—User diversity and co-evolution of users and systems are two important phenomena usually observed in the design and use of IT artifacts. In recent years, End-User Development (EUD) has been proposed to take into account these phenomena, by providing mechanisms that support people, who are not software professionals, to modify, adapt, and even create IT artifacts according to their specific evolving needs. This is particularly true in the case of sentient multimedia systems, in which the system is called on to interact with multiple sensors and multiple human actors. However, to motivate and sustain these people, a culture of participation is necessary, as well as proper meta-design activities that may promote and maintain it. To this aim, this paper first proposes a model for describing interaction and co-evolution in sentient multimedia systems enhanced by EUD features. Then it presents four main roles involved in interaction and co-evolution, including that of maieuta-designer, as the “social counterpart” of the meta-designer. Finally, it describes how the maieuta-designer is in charge of carrying out all those activities that are necessary to cultivate a culture of participation, by means of proper ways that are briefly introduced in the paper.

**Keywords**—End-User Development; cultures of participation; co-evolution; meta-designer; maieuta-designer

## I. INTRODUCTION

Sentient multimedia systems are distributed systems that actively interact with the environment through the exchange of multimedia information with many kinds of information sources, such as sensors, robots, actuators, websites and others. End users also belong to such sources of information, since they are called on to communicate and express their feelings, evolving needs and requests to this web of computational nodes. Accordingly, the overall system has to take into account this information flow coming from humans. If we take this stance, a sentient multimedia system can also be seen as a socio-technical system, which encompasses people (rather than just users) that are bound together by social ties and personal relations of acquaintance and that are also linked with each other and with personal devices and other machines. The latter are able to perceive the environment in which people interact, also by considering the capability of people to feel a situation besides perceiving it (e.g., through the issue of a preference, “likes”, and emoticons in tweets), and interpret situations to give people multimedia and multichannel means to act accordingly.

In many IT domains, IT artifacts are usually developed as commodities rather than as ad-hoc projects, that is more for uniform populations of consumers rather than for members of different communities that exhibit local needs and perform situated practices to achieve their goals [1]. Moreover, each end-user community is often characterized by user diversity, due to users’ different physical and/or cognitive abilities, past experiences, roles, responsibilities and work contexts. To this end, in today’s competitive global market, the adoption of product configuration software has recently helped to increasingly speed up the understanding of the customers’ needs for a successful design and implementation of customized products [2-5]. In fact, product configuration is the activity of customizing a product, in order to better meet the needs of a particular end user more quickly. However, fulfilling the needs of end users is a “moving target” [6], since they evolve (e.g., regarding their proficiency of use, skills, expectations, needs, wishes and domain knowledge) by using software systems, and they can also change their practices (to accommodate the new artifact [7]). Acknowledging this twofold evolution (i.e., of users and their tasks) entails the requirement that IT artifacts should be designed to be very flexible, in order to be easily adapted to the specific needs of the user communities and, hopefully, to be personalized by the individual users to better fit their own evolving needs. This overall phenomenon has been called co-evolution of users and systems, to denote the variety of situations where users and their systems must co-evolve in a continuously self-adapting mutual fit [8].

End-User Development (EUD) has been proposed as one possible solution to cope with the challenges posed by user diversity and co-evolution, since it encompasses techniques that allow end users to modify and extend their own IT artifacts without necessarily delegating these modifications to software professional developers. Taking co-evolution seriously sheds light on the fact that continuously relying on professional actors for these interventions would not be feasible in the long run. Indeed, one kind of unintended consequence related to IT artifact deployment, which is reported most frequently in the specialist literature, regards the never-ending request for modifications, corrections and evolution of artifacts by the users [9]. In other words, in a EUD perspective, software systems are viewed as “continuously evolving sociotechnical

systems driven by design activities of both professional software engineers and users” [10].

For the particular human-oriented extension of the traditional definition of sentient multimedia systems (see above) proposed by us, we uphold that an EUD approach is necessary also for the design and continuous evolution of this novel class of applications. In particular, through the use of EUD methods and techniques, a sentient multimedia system can resemble a living system, with some degree of intelligence, that reacts to the end users’ evolution through the consequent self-adaptation and in turn favors the adaptation of end users to the evolution of the system. In this way, the word “sentient” would also imply “alive” and “intelligent”, as this kind of emergent behavior is what characterizes many complex systems, like socio-technical systems, that do have feedback loops between perception and action.

However, this far-reaching objective also requires the “cultivation” of a culture of participation, in order to motivate and sustain end users in their contribution to system evolution. This avoids the risk of participation inequality [11] and of replicating the current gap between IT professionals and end users at the shop floor level. The meta-design framework, which aims to help “users to become co-designers at use time”, has been proposed to this latter aim [12]. However, this framework seems to have neglected some important aspects that might make it more operative in real settings, like its relationship with activities promoting a culture of participation. Therefore, in this paper, we would like to investigate how to extend the original proposal of meta-design with mechanisms that are more specifically aimed at cultivating a culture of participation and thus enabling a suitable environment for the sustainable co-evolution of users and their systems.

To this end, we draw on our research experience in a variety of application domains (e.g., medicine, mechanical engineering, e-government and others) and on the analysis of the existing literature about a variety of EUD projects. We first propose a model for interaction and co-evolution that aims at clarifying the dynamics occurring in EUD settings among the different professionals involved, their tasks and the systems they use. In particular, we expand the technical activities a meta-designer should perform to support system adaptation and growth over time, and then provide indications about the activities of his/her “social counterpart”, namely the *maieuta-designer*, who is in charge of activating all the necessary social mechanisms that may sustain the co-evolution phenomenon.

The paper is organized as follows: Section II presents the related works; Section III describes the proposed model for interaction and co-evolution between users and systems, by clarifying in particular the role played by each different professional; Section IV expands the role of the *maieuta-designer* and proposes a framework for supporting his/her activities and Section V concludes the paper.

## II. RELATED WORKS

Since the eighties, the human-computer interaction (HCI) literature has proposed different techniques for the design of interactive systems. They start from user-centered methods [13] – including field studies, interviews, task analysis, usability

testing – and move on to participatory design techniques [14], where users are directly involved in the creation of interaction scenarios [15] and/or static and semi-static prototypes [16].

However, while HCI scholars have been considering user-centered and participatory design approaches as consolidated and successful practices for interactive system development, only in recent years the need for continuous system development with the participation of end users also at use time has received adequate attention. Consequently, end-user programming (EUP) techniques have been embedded in commercial software, such as macro recording in word processors, formula composition in spread sheets or filter definition in e-mail clients. EUP is defined in [17] as “programming to achieve the result of a program primarily for personal, rather than public use.” However, as highlighted in [18], the problem with end-user programming is that the programs created by end users are often of too low quality in terms of efficiency and maintainability. To cope with this problem, the End-User Software Engineering (EUSE) research area has emerged, which studies EUP practices and proposes new kinds of technologies that help end users improve software quality [17]. Software engineering activities, such as specification, reuse, testing, and debugging, are the primary focus of this research area, and therefore attention is put on the software code created by end users.

Recently, the term ‘end-user programming’ has been gradually replaced with the term ‘end-user development’ [19], in order to give user involvement in system design a broader perspective, with respect to mere code development for personal goals. Indeed, EUD denotes any kind of active participation of end users in the software design and development process, ranging from requirement specification through domain-specific modelling (cfr. [20]) to more advanced activities, such as system personalization and modification, or even creation of new software artifacts [21]. Therefore, contrary to participatory design, EUD research advocates end user participation not only during the design phase, but also during system usage. Both research lines are currently very active: on the one hand, new methods and techniques are being studied to better capture and satisfy user requirements; on the other, a variety of mechanisms are being proposed to allow run-time system modifications with increasing complexity and expression power [22] and possible creation of new software artifacts to be used by (or also by) other people [23].

This has led to the conception of a new design paradigm for systems that support EUD activities, namely *meta-design*. In [10], meta-design is regarded as a framework for creating “sociotechnical environments that empower domain experts to engage actively in the continuous development of systems”. Conversely, in [8] Costabile et al. view meta-design as “a design paradigm that includes end users as active members of the design team and provides all the stakeholders in the team with suitable languages and tools to foster their personal and common reasoning about the development of interactive software systems that support end users’ work”. The two proposals focus on different but complementary aspects of meta-design: the former stresses the social as well as the technical nature of the software environment that should foster

users' involvement; the latter highlights the importance of defining proper languages and tools to adequately support different stakeholders' participation in system development.

In both cases, meta-design is regarded in the wider perspective of cultures of participation [12, 24]. Indeed, cultures of participation open up new opportunities and challenges for the design of innovative interactive systems, whose users "are provided with the means to participate and to contribute actively in personally meaningful problems" [12]. With respect to EUSE, cultures of participation thus pay more attention to interaction design and conceptual modelling of EUD environments, cooperation among users and impact on work organizations. Particularly, beyond meta-design, Fischer proposes two other components for establishing a culture of participation [12] that we think are fundamental also in the conception and design of sentient multimedia systems: 1) *social creativity*, in order to allow all possible contributions to be considered to solve a complex problem and to support people interacting with each other and through shared hardware and software artefacts; 2) *richer ecologies of participation*, in order to obtain different levels of participation on the basis of the different roles that people and their devices can play or would like to play.

The importance of cultures of participation has been demonstrated in the literature with reference to a variety of application domains [12]. However, one aspect has been neglected till now, that is how to enable users to appropriate such a culture, in order to obtain software artifacts that successfully evolve in user's hands and co-evolve with users' tasks, abilities, skills, and preferences. This is particularly true for sentient multimedia systems, as they are bound to the environment and the people living in it, and thus this requires users to be even more engaged in their shaping over time. In our view, this requires proper mechanisms and tools that allow the nurturing and cultivation of cultures of participation within organizational settings. To this end, in the following we propose an analysis of the activities a meta-designer should perform throughout the software life cycle. In particular, we suggest refining the meta-designer role, by splitting and specializing it into two main sub-roles – a more technical role and a more social one – which, according to the situation, domain, and budget constraints, can be played by the same person or by different professionals.

### III. ECOLOGY OF PARTICIPANTS IN CO-EVOLUTION OF USERS AND SYSTEMS

This section is aimed at proposing a model for interaction and co-evolution in complex settings, including sentient multimedia systems. To this end, it first presents the traditional view on the interaction and co-evolution between users and systems, and then extends this view to the case of EUD settings. In particular, it focuses on the roles played by different stakeholders in EUD practice, and then clarifies how co-evolution may be "technically" and "socially" sustained.

#### A. Interaction and Co-Evolution of End Users and Systems

The interaction and co-evolution model proposed in [25] describes three types of mutual influence between end users and systems, which give rise to three different cycles (see

Figure 1). The most internal one is the *interaction cycle*, namely a short-term cycle emerging from the exchange of visual, aural, haptic or multi-modal messages between a user and a system during interaction. This cycle has been modelled in the literature in a variety of ways [26-28]. The intermediary cycle is what Carroll and Rosson call the *task-artifact cycle* [7]. It is a mid-term cycle that highlights how the software artifacts created to support some user's tasks usually suggest new possible tasks and, to support these new tasks, new artifacts must be created. Finally, the external cycle is a long-term cycle that concerns the mutual influence between the technology used for artifact implementation and the user organization (*technology-end-user organization cycle*). Since technology advances give software developers new possibilities for improving interactive systems once they are already in use, new interaction possibilities occur that might change users' working habits, thus making their social and work organization evolve itself with technology [8, 29]. The co-evolution phenomenon thus encompasses all the three cycles, and they affect each other as described in [25].

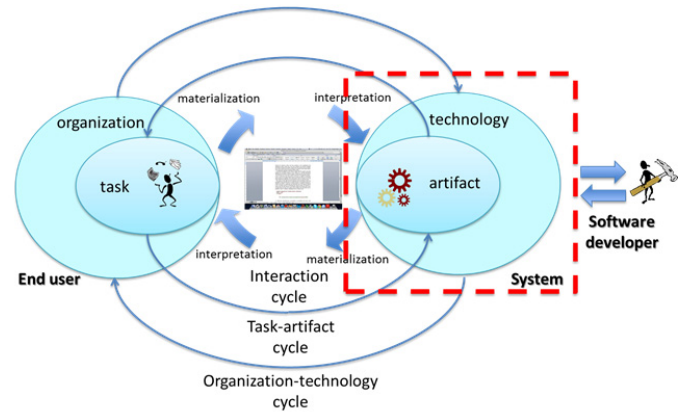


Figure 1. The traditional Interaction and Co-Evolution model.

Figure 1 also highlights the two roles involved in this traditional view on interaction and co-evolution: the *end user* and the *software developer*. The end user is considered a passive user of an interactive system and consumer of its products and services. The software developer is the creator of the system and, during its life cycle, he/she may be called on to modify and extend it for adaptation to the emerging needs and requests of end users.

#### B. Four Main Roles in EUD practice

In EUD literature, a new role is considered: the *meta-designer*. Thus, the two main roles usually highlighted, other than the software developer, are: *end user* and *meta-designer*. End users are increasingly required to act as active contributors at use time, thus becoming "producers" of contents and functionalities, like in Wikipedia, Scratch, SketchUp, and many others [12]. In literature such an "active" end user is called in different ways: "power user" [30], "local developer" [31], "gardener" [32], "end-user developer" [33], "bricolant bricoleur" [34].

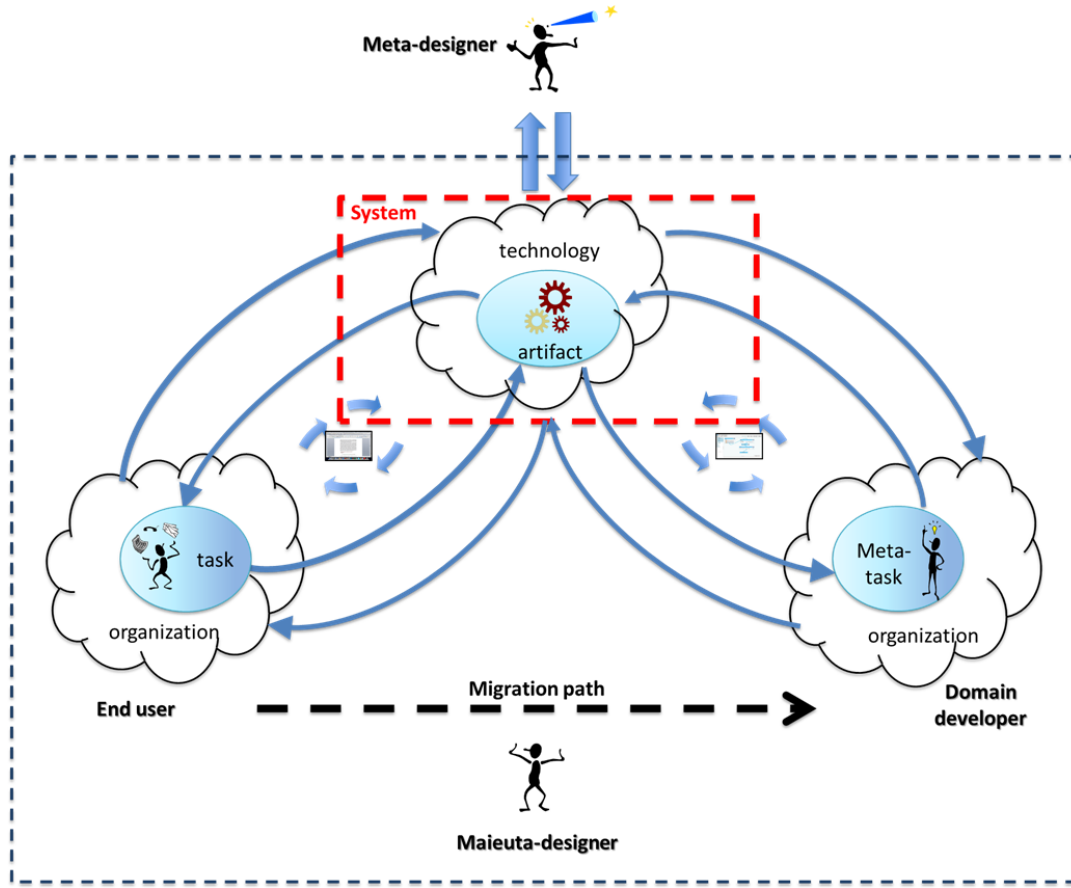


Figure 2. Interaction and co-evolution in EUD settings.

To disentangle this variability of names, we have proposed to refer to such a figure with the term *domain developer* [23]. This term has been chosen because this person is always an expert of the domain in which he/she works and her/his main goal is more the development of the capabilities available in her/his setting than just software code (software is never an end in itself, but always a means). Thus, the domain developer subsumes all those roles denoting people in charge of carrying out software development activities (namely, ‘actual’ EUD activities) without being professional software developers. In some cases, end user and domain developer are roles played by the same person, as in the case described in [10], where a geoscientist decided to spend three months in acquiring programming knowledge, in order to be able to develop software for himself to analyze the data he collected. However, in the majority of situations they are played by different people, who may also belong to different communities, like in multi-tiered proxy design problems [33], i.e. in all those situations where end users cannot or are not willing to act as developers. A typical example is the case of e-government, where citizens using e-government services constitute the community of end users, whilst civil servants called on to develop services for citizens belong to the community of experts in government issues [35], and thus may become domain developers. Another example is the system described in [36], which provides an editing tool that allows caregivers (domain developers) to

customize a simple, wireless prompting system for individuals with cognitive disabilities (end users).

The role of meta-designer, on the other hand, is intended for all professionals who are in charge of creating “socio-technical environments that empower users to engage actively in the continuous development of systems rather than being restricted to the use of existing systems” [37]. In other words, a meta-designer “creates open systems at design time that can be modified by their users, acting as co-designers, requiring and supporting more complex interactions at use time” [12].

Given these definitions, which do not completely clarify the activities a meta-designer should actually perform, we also consider the role of *maieuta-designer* (the term is pronounced just like that of meta-designer, but with a *ju* in the middle: ‘*metā* designer vs. *mei’ju:tā* designer). Like the meta-designer, also the maieuta-designer can be considered as someone in charge of designing the EUD-enabling environment, by which domain developers can build and adapt the artifacts to be used by end users. The role of the maieuta-designer encompasses activities that are involved in the task of supporting the meta-task of the domain developers, namely creating the socio-technical preconditions for: a) having the domain experts appropriate the design culture and technical notions necessary for the meta-task of artifact development and b) involving as many end users as possible in the process of continuous



refinement of the artifact, by improving participation and “produsage” [38]. For this reason we call such a designer a “maieuta”. This is partly in analogy with the Socratic method of making people acquire notions, motivations and self-confidence to undertake challenging tasks and partly in clear assonance with the term meta-designer, of which it is a specialization more oriented to the social aspects of EUD practice than to the technical ones [34].

### C. Co-Evolution of End Users and Systems through EUD

The four roles described above – end user, domain developer, meta-designer and maieuta-designer – interact with each other and with the IT artifact and EUD tools, and each contributes to the co-evolution phenomenon. Fig. 2 presents an extended version of the Interaction and Co-evolution (ICE) model previously described, which encompasses all the four roles.

In EUD practice, the traditional co-evolution process (left-hand side of the figure) is *sustained* by the right-hand side co-evolution process, which involves the domain developer (see Fig. 2). Indeed, requests for system evolution coming from end users reach domain developers, who may directly operate on the system through EUD tools or, if necessary, may in turn ask meta-designers for the evolution of their own tools (by means of the most internal cycle, i.e., the interaction cycle). In particular, the task-artifact cycle that involves end users affects the *meta-task-artifact cycle* of domain developers. For example, in the Electronic Patient Record (EPR) project described in [39], ward physicians can be aware that new data are needed (for example, in the case of specific pathologies), but they cannot find the related module in the EPR. Thus, they have to ask for a new specific module from the head physician, who, using EUD tools, will evolve the current EPR accordingly or request meta-designers to create the new type of module. The most external cycle is also in this case a long-term cycle that regards the mutual influence between the technology used for artifact implementation and the organization of the community of the domain developers (*technology-domain developer organization cycle*). Technology advances give meta-designers new possibilities for improving interactive systems used by domain developers to evolve the system, resulting in new interaction possibilities that might also change users’ work habits.

Thanks to the powerfulness of the current technology, even accessible through the Internet, users have the possibility of increasingly taking an active role in the development of software tools suited to their needs. This has led from a strong dichotomy between the end user and the domain developer to a continuum of roles that constitutes a rich ecology of participants [12, 40] with different skills towards development, responsibility, appropriation and contribution in the whole ecosystem. This is true also in the case in which the same person is and wants to be just a “consumer” (i.e., only a user) in some situations and a “producer” (i.e., domain developer) in others. Therefore, “end user/domain developer” is not an attribute of a person, but a role assumed in a specific context. As we will discuss in the following, the maieuta-designer is in charge of fostering and favoring the migration path from the role of end user to that of domain developer.

The artifact in Fig. 2 is an *intermediary object* between the two co-evolution processes, which can be regarded as being composed of two parts: 1) the software system devoted to the end user and 2) the EUD tools (including the EUD environment and/or software components as building blocks of the software system being developed) that are used by the domain developer to generate and/or adapt the software system for end users.

An intermediary object is usually an object that is shared, exchanged and circulated among members of networks and communities to mediate their interactions [41]. In [23], we have further distinguished intermediary objects as *boundary objects* and *knowledge artifacts* to better characterize their role in EUD contexts. The notion of boundary objects has been introduced by Bowker and Star [42] to account for those artifacts that enable a sort of standardized and effectively simplified communication and coordination between members of different communities of practice. For example, in the e-government project described in [43], a civil servant may generate the XML description of an e-government service through a suitable EUD environment, and this description gives rise to the automatic creation of the web pages to be used by citizens who will apply for the service. Therefore, the XML description and the corresponding web pages can be regarded as a boundary object between the civil servant and the citizen communities. Knowledge artifacts, on the other hand, are artifacts that enable and support learning and innovation within a specific community of practice (what in [44] has been called a “knowing community”), namely processes of knowledge acquisition, accumulation and sharing among its members [45]. For example, the Electronic Patient Record [39, 46, 47] represents a knowledge artifact used within a hospital ward and among different wards for accumulating and sharing knowledge about patients.

### D. Supporting Co-Evolution in EUD settings: the Meta- and Maieuta-Designer roles

Fisher et al. have proposed a set of meta-design guidelines [10], namely indications at a high level of abstraction, on how to carry out a meta-design project. Assuming the validity of such guidelines, we propose here to make a further step, by identifying some more operational indications, in order to make the meta-design activities concrete. In particular, splitting the social activities and technical activities that the meta-designer and maieuta-designer are called on to carry out respectively is done to take into account the socio-technical gap, that is the divide between what is known that should be supported socially and what can actually be technically supported [48].

To this aim, we suggest that the meta-designer be in charge of designing and providing the most effective EUD tools that may sustain the co-evolution between end users, domain developers and IT artifacts. In this way, the meta-designer is not just a software developer playing the role described in Figure 1 (system development and its possible evolution over time, on the basis of the users’ requests); but he/she must be endowed with a set of skills that allows him/her to understand the users’ potential for participation in software co-creation, in order to support them with suitable technical tools. Thus he/she must have competencies in human-computer interaction,

computer-supported cooperative work, interaction design, knowledge management, multimedia and even semiotics [49].

A variety of EUD solutions have been proposed in the literature over the years, from script creation [36], to component-based approaches [50, 51] and from meta-model instantiation [52, 53] to visual programming [8, 54, 55]. The meta-designer is thus called on to choose the most suitable paradigm for the case in hand, and tailor it to the application domain, namely to its habits and users' characteristics and preferences. Furthermore, a meta-designer is requested to develop the infrastructures for communication among end users and between end users and domain developers [56].

The maieuta-designer, on the other hand, is not only what elsewhere has been defined as a facilitator [57], that is, the role responsible for facilitating the *adoption* of an IT artifact within a certain community. Rather, the maieuta-designer is the person who is supposed to facilitate *appropriation* [58], i.e., the process by which end users migrate from their initial role to that of domain developer along the continuum of roles discussed in [40] or, at least, to enable and empower end users to appropriate and contribute to their IT artifacts. Therefore, the end users can commit themselves to improving the artifacts as a way to make them more effective and their work more efficient. Whenever an end user is not capable of, or not interested in, "evolving" into the role of domain developer, the maieuta-designer might favor her/his participation in system evolution, e.g., by simply guiding her/him to report perceived shortcomings and system faults, and suggesting due modifications and appreciated improvements. Moreover, the maieuta designer is also in charge of reducing the socio-technical gap, by creating the conditions and making feasible the use of the IT artifacts and lowering the tendency of the users to just give up using them, especially if they do not have any technical support from IT professional people.

The clarification provided here of the roles of meta-designer and maieuta-designer is based on a critical reflection on the authors' experience in a variety of EUD projects. Actually, in these projects they have often been engaged in playing both roles, at different stages of the system life cycle. As meta-designers, they have studied and proposed different interaction metaphors and environments, to support domain developers in carrying out EUD activities; whereas, as maieuta-designers they have realized that, besides asking users the right questions as any good business analyst does, it is even more important that the users themselves are induced to think about their answers and build their own awareness on how to deal with issues through the system. Thus, the aim of this paper is to shed light on the techniques that could be adopted to sustain system appropriation, by underlying the need for transferring this capability to someone inside the organization who could reiterate the awareness process and make it sustainable.

#### IV. HELPING END USERS HELP THEMSELVES

This section provides some further hints on how the tasks of the maieuta-designer could be performed.

##### A. Identifying the maieuta-designers and their tasks

As has been outlined above, the concept of maieuta-designer means identifying someone who could make the

community gathering around a EUD platform progressively more independent of the IT professionals. To some extent, he/she is the person who guarantees the long-term sustainability of the EUD project. Therefore, this can be an IT professional with an educational curriculum that is quite different from the one traditionally proposed for the common software analyst or engineer. This agenda would encompass, for instance, teaching the basics of social informatics and some qualitative research methods adapted to the IT domain [59], like focus groups [60], insights on current theories on IT impact and risk management [61], as well as notions of socially-informed history of technological evolution [62].

The person playing this role must also train on the job one or more "insiders" of the community of end users that will continue his/her work of facilitation. The latter should be endorsed by the sponsors of the IT project and the organization's top managers, and also be chosen on a voluntarily basis according to their ability and will to encourage colleagues to take part in the development process. The term "designer" is not out-of-place here for at least two reasons: first, the maieuta-designer would be a clear example of a *critical designer*, i.e., someone "who asks carefully crafted questions and makes [people] think" [63], instead of focusing on solving problems and finding answers. Moreover, one of the main tasks of the maieuta designer would be to "design" (or better "co-design") initiatives in which to promote the EUD project, disseminate the underlying values and concepts (i.e., empowerment, co-production, appropriation, co-evolution, produsage, equipotentiality [38], etc.) and enrol the most expert and enthusiastic end users. Then the maieuta-designer should give due visibility of the end users' contributions, and devise simple mechanisms to foster participation and build a real culture of participation. This can be done in many ways: for instance by applying blended gamification, within a competition among colleagues, possibly associated with some reward or compensation policy, e.g., a mechanism by which "the more contributions produced, the higher the rank achieved". Moreover, this can be done by setting up a social media associated with the IT project, e.g., a forum, a blog, a wiki, or something that integrates all of these simpler components, in which to ask for content and contributions and moderate communication within these ad-hoc means. In so doing, such a Web resource would flank the EUD platform as an additional "resource for action" [64] and a virtual meeting place where tasks are coordinated upon the EUD artifacts and the related procedures, FAQs and use instructions are documented/discussed.

##### B. The Google example

Over the years, Google software developers have produced a set of apps and gadgets, technologies for voice recognition, home appliances and entertaining devices. Many of them are smartphone-based and thus accessible and widely used by end users, as well as most of the technologies developed so far by Google. However, even though the specific functionality of each application is usually known, the possibility of connecting or using together different applications, as parts of a unique ubiquitous system, is rarely considered by end users. Indeed, Google apps and technologies are thought to also interact with each other, in order to provide a more pervasive and ubiquitous

experience. For example, Chromecast dongle can be connected to the TV and exchange data with your tablet or phone, while Google Now can exchange data with Google platform, in order to suggest personalized hints. Actually, all these applications can be considered as parts of a pervasive multimedia system surrounding us.

Since people might not be immediately aware of this possibility, Google has set up a team whose aim is to advertise the “hidden” sentient multimedia system already available in our pocket. The team has thus created the Google House Project, an itinerant house toured through New York, London, Paris, Hamburg, Dubai and many other important cities all around the world<sup>1</sup>. The house is split into five rooms: kitchen, living-room, study, travel-room and fashion-room, with a sixth environment outside to try out Google Glass, the company’s latest wearable device (Fig. 3). The majority of technologies available in the house are smartphone-based, linked up to huge displays scattered around the home via Wi-Fi.



Figure 3. The Google House idea.

In such a setting, tablets or phones can be used in the kitchen to look for recipes, by using voice commands, or run a Google-owned YouTube channel with millions of subscribers while cooking, or have the system convert doses from grams to ounces. In the living-room you can use Google Nav to plan your day’s journey or use Google Chromecast dongle. The Chromecast dongle is popped into the back of the television and then connects wirelessly with any tablet or smartphone to stream your favorite movie or, even, use the new photo app Auto Awesome. This application is able to automatically

exclude duplicates and duds from a gallery of holiday pictures and take the smiles from a series of photos of the same people, in order to pool them into one, perfect shot. In the fashion room, your daughter can use Hangout to talk with her friends about choices of dress or to access Google Trends to identify the most popular fashion styles while getting dressed.

In short, Google has created over the years a variety of applications and devices, which have been put in the hands of the users, in order to allow them to provide their feedback, try different usages, contribute their contents (e.g., in Google Maps or Google Sketchup), and eventually become part of the world-wide Google community. Then, Google developers (we can call them meta-designers in this case), have offered the possibility of combining these “pieces” of technology in different ways and at the user’s own pace to obtain a low cost, sentient multimedia system. However, most end users, even though they are perfectly able to exploit existing applications and devices in this new way, without the intervention of computer experts, were not aware of this possibility. Hence, Google had the idea of “advertising” it through a widely understandable example: a real and tangible house. Indeed, the team of researchers welcoming and hosting visitors to try out the Google experience (the maieuta-designers in this case), aimed to make users conscious of the possibility to create their own sentient multimedia system, and thus actually carry out a form of EUD. Furthermore, the Google House project allows end users to appreciate the different usages of each application: for example, the use of Hangout is shown in the fashion-room, but nothing prevents a scientist from deciding to use it in his study, to keep in touch with his foreign colleagues.

In other words, the task of maieuta-designers has been accomplished in this case by creating a real house and exploiting the resonance that the Google name may have all over the world. Like a bait-and-switch, this has enabled a limited number of users to actually visit the house and try its gadgets, but it has allowed millions of users to become aware of the nearly infinite scope of sentient multimedia systems and of the ever-expanding catalog of available applications which can be combined, according to one’s preferences.

Obviously, this approach can rarely be followed in medium and small organizations, and thus other ways should be identified to make users more confident in adopting tools and their capabilities, in order to put them in a position to exploit customization and extension possibilities, for a better fit with their needs. Furthermore, a true maieutic approach would also entail challenging real users, visiting the house, with Cedric Price’s provocative question: “if Google [Technology] is the answer, what is your question?” In looking for possible answers, the end users could also appropriate the underlying idea of such an augmented shelter and could make the best out of it in their situated lives and homes. For these reasons, the following sub-section delineates a more scalable approach that, capitalizing on the experience gained so far in different IT projects, is built around the idea of asking users questions.

<sup>1</sup> <http://www.thenational.ae/uae/technology/google-house-in-dubai-the-home-of-the-future-now>

TABLE I. MACRO-AREAS OF CONCERNS AND LIST OF QUESTIONS

Macro-areas of concern	Sample Questions
Psychological ownership and change management	<p>What's the system for you, and why has it been produced and its adoption encouraged within the organization?</p> <p>Do you think that communication within your team, or with the other teams, has changed lately, and if this is the case, has it been for the better or for the worse?</p> <p>How long have you been using the new system, do you think your work load has changed, that is increased, reduced, or it is just the same?</p>
Work process redesign and adjustment (i.e., fit to task)	<p>To what extent do you think you can exploit the system's full potential? How much do you think the system fits your specific needs currently?</p> <p>Do you still use paper and office applications that you believe the new system will (or should) substitute sooner or later?</p>
Usability shortcomings and room for improvement	<p>Have you found using the system easy so far?</p> <p>Have you realized you have made errors in the process of either entering or retrieving information from the system?</p>
Lack, redundancy or overload of data structures and functionalities	<p>Do you think the system is requiring you to fill in too many data that are not really necessary to proceed in your tasks?</p> <p>Do you think the system provides too many functionalities among which you need to find the right one for your tasks?</p>
Anomaly detection, bug reporting and evolution traceability	<p>Have you found any errors or something you've considered a fault of the system while using it lately?</p> <p>Have you applied some effective solution or workaround to overcome a shortcoming related to the system lately?</p> <p>Do you think that the system has become more difficult to be used after some of its recent updates and new releases?</p>
End-user deskilling and expertise preservation and enhancement	<p>Have you lately experienced problems in the handing over of tasks or in the workflow (like unusual delays, common resources blocked by other teams and the like)?</p> <p>Do you think that using the new system may contribute to preserving or even enhancing your know-how about the work tasks?</p>
Authorship- and privacy-related concerns	<p>Do you think that sharing content and system modifications within your team prevents you from protecting ideas and information?</p> <p>Do you think that either the data you put into the system, or the actions you perform during its use could threaten your security and privacy?</p>
End-user accountability and power issues (also empowerment)	<p>Since the introduction of the new system, do you think that new people or roles have gained more visibility and power within your organization, at the expense of others?</p> <p>What's the main obstacle that prevents you from participating more actively in the IT project (like time, skills, the colleagues already involved, a sense of pointlessness, ...)?</p> <p>What could really convince you to join the IT project, if anything (e.g., explicit acknowledgment by the top management, economic rewards, non-monetary compensations, benefits, social status, ...)?</p>

### C. The maieutic approach

A maieutic approach is mainly characterized by the fact that it “brings others to conceive ‘thoughts or ideas’” with questioning [65], that is by helping others actively understand by themselves how they could make a worthy contribution to the project. To contextualize this approach in EUD settings, we propose a tentative list of items (see Table 1). Each item in the second column of Table 1 is a question that the maieuta-designer could ask (or speak about with) his/her colleagues to address a broader theme or topic regarding the process of digitization of the work setting or the related changes. This can be done in either small polls or surveys, administered through social media, or in informal but scheduled meetings with the members of a specific team at a time. It could also be done even in totally informal and impromptu talks at the coffee break or in similar situations [66]. In particular, the first question in the table (“What’s the system for you, and why has it been produced and its adoption encouraged within the organization?” [67]) is the most important one, as it refers both to the original Socratic main question (What’s this? *Ti esti;*) and to one of the most important matters of concern in requirement engineering.

Far from being comprehensive, the list of questions reported in Table 1 represents just a first contribution within a research strand. This could address more seriously how to contribute to fostering a culture of participation within organizational communities, especially in the context of a digitization project that is supported by EUD and meta-design techniques. Indeed, a more general framework of concerns, that a maieuta-designer should address, can be built from this preliminary list and the study of its impact on real settings (see the first column of Table 1). This framework to date includes eight macro-areas of concern that encompass: psychological ownership and change management [68, 69]; work process redesign and adjustment (i.e., fit to task); usability shortcomings and room for improvement; lack, redundancy or overload of data structures and functionalities; anomaly detection, bug reporting and evolution traceability; end-user deskilling and expertise preservation and enhancement; authorship- and privacy-related concerns; end-user accountability and power issues (also empowerment).

Our point is that there is a need to detect motivated people within organizations, and not only give to the domain developers a set of tools (i.e., the EUD environments), but also (and above all) assign to some people (i.e., the maieuta-designers) precise responsibilities and roles. It is particularly important to provide maieuta-designers with an indication of a set of possible actions to undertake and initiatives to foster, so that they can contribute to building a real culture of participation within their organization. All the actors involved may thus enjoy such a culture within the wider process of co-evolution.

### V. CONCLUSIONS

In this paper we have proposed explicitly extending the notion of a sentient multimedia system to encompass socio-technical networks of humans and non-humans [70], where both kinds of agents are able to perceive and act in the environment, but only the former can feel it and make the

whole ensemble “sentient”. The most important feature of these networks is their ability to autonomously reconfigure their inner relationships (human-human, machine-machine, human-machine) and also to evolve, in the face of the continuous changes in the environment and the network itself. In this paper, we have focused on a particular design approach that does not only allow for continuous evolution, but even fosters it: meta-design. We have extended this framework, originally proposed within the EUD field, by enriching it with an additional role, the maieuta-designer, that could also be seen as a critical design device, and therefore as a way to make sentient multimedia systems able to wonder about and reflect upon themselves. As a critical design device, we suggest supporting the maieuta-designer with a list of questions, organized in turn according to a set of macro-areas of concerns, through which he/she can trigger reflection and cultivate a culture of participation. This can be a powerful way to make sentient multimedia systems more capable of interacting with an ever-changing environment and to be ready for the unexpected.

#### ACKNOWLEDGMENT

This work is partially supported by the Italian Ministry of University and Research (MIUR) under grant PON 02\_00563\_3470993 project “VINCENTE - A Virtual collective INtelligentCe ENvironment to develop sustainable Technology Entrepreneurship ecosystems” and by the Italian Ministry of Economic Development (MISE) under grant PON Industria 2015 MI01\_00294 “LOGIN – LOGistica INtegrata”.

The authors wish to thank also Ms Lynn Margaret Rudd for her careful proofreading.

#### REFERENCES

- [1] W.J. Orlikowski and C.S. Iacono, “Desperately Seeking the “IT” in IT Research? A Call to Theorizing the IT Artifact,” *Information Systems Research*, vol. 12, no. 2, 2001, pp. 121-134.
- [2] G. Kruse and J. Bramham, “You choose [product configuration software],” *Manufacturing Engineer*, vol. 82, no. 4, 2003, pp. 34-37.
- [3] F. Colace, M. De Santo and L. Greco, “An adaptive product configurator based on slow intelligence approach,” *Int. J. Metadata Semant. Ontologies*, vol. 9, no. 2, 2014, pp. 128-137.
- [4] N. Franke and F.T. Piller, “Key research issues in user interaction with user toolkits in a mass customisation system,” *International Journal of Technology Management*, vol. 26 no. 5/6, 2003, pp. 578-599.
- [5] C. Ardito, B.R. Barricelli, P. Buono, M.F. Costabile, R. Lanzilotti, A. Piccinno and S. Valtolina, “An Ontology-Based Approach to Product Customization,” *End-User Development*, LNCS, M. F. Costabile, Y. Dittrich, G. Fischer and A. Piccinno, eds., Springer, 2011, pp. 92-106.
- [6] D.A. Norman, “Human-centered design considered harmful,” *Interactions*, vol. 12, no. 4, 2005, pp. 14-19.
- [7] J.M. Carroll, W.A. Kellogg and M.B. Rosson, “The task-artifact cycle,” *Designing interaction: psychology at the human-computer interface*, J. M. Carroll, ed., Cambridge University Press, 1991, pp. 74-102.
- [8] M.F. Costabile, D. Fogli, P. Mussio and A. Piccinno, “Visual Interactive Systems for End-User Development: a Model-based Design Methodology,” *IEEE Transactions on System Man and Cybernetics Part A-Systems and Humans*, vol. 37, no. 6, 2007, pp. 1029-1046.
- [9] M.I. Harrison, R. Koppel and S. Bar-Lev, “Unintended Consequences of Information Technologies in Health Care—An Interactive Sociotechnical Analysis,” *Journal of the American Medical Informatics Association*, vol. 14, no. 5, 2007, pp. 542-549.
- [10] G. Fischer, K. Nakakoji and Y. Ye, “Metadesign: Guidelines for Supporting Domain Experts in Software Development,” *IEEE Software*, vol. 26, no. 5, 2009, pp. 37-44.
- [11] P.B. Brandtzaeg and J. Heim, “A typology of social networking sites users,” *Int. J. Web Based Communities*, vol. 7, no. 1, 2011, pp. 28-51.
- [12] G. Fischer, “Understanding, fostering, and supporting cultures of participation,” *Interactions*, vol. 18, no. 3, 2011, pp. 42-53.
- [13] D.A. Norman and S.W. Draper, *User Centered System Design; New Perspectives on Human-Computer Interaction*, L. Erlbaum Associates Inc., 1986, p. 526.
- [14] D. Schuler and A. Namioka, *Participatory Design: Principles and Practices*, Lawrence Erlbaum Associates, Inc., 1993, p. 312.
- [15] M.B. Rosson and J.M. Carroll, *Usability engineering: scenario-based development of human-computer interaction*, Morgan Kaufmann Publishers Inc., 2002, p. 422.
- [16] S. Bødker and K. Grønbaek, “Design in action: from prototyping by demonstration to cooperative prototyping,” *Design at work: Cooperative design of computer systems*, J. Greenbaum and M. Kyng, eds., L. Erlbaum Associates Inc., 1992, pp. 197-218.
- [17] A.J. Ko, R. Abraham, L. Beckwith, A. Blackwell, M. Burnett, M. Erwig, C. Scaffidi, J. Lawrance, H. Lieberman, B. Myers, M.B. Rosson, G. Rothermel, M. Shaw and S. Wiedenbeck, “The state of the art in end-user software engineering,” *ACM Computing Surveys*, vol. 43, no. 3, 2011, pp. 1-44.
- [18] M. Burnett, “What Is End-User Software Engineering and Why Does It Matter?,” *End-User Development*, Lecture Notes in Computer Science, V. Pipek, M. B. Rosson, B. Ruyter and V. Wulf, eds., Springer Berlin Heidelberg, 2009, pp. 15-28.
- [19] H. Lieberman, F. Paternò and V. Wulf, eds., *End User Development*, 9, Springer, 2006.
- [20] N. Mehandjiev and L. Bottaci, “User-enhanceability for organisational information systems through visual programming,” *Advanced Information Systems Engineering*, Lecture Notes in Computer Science, P. Constantopoulos, J. Mylopoulos and Y. Vassiliou, eds., Springer, 1996, pp. 432-456.
- [21] M.F. Costabile, D. Fogli, C. Letondal, P. Mussio and A. Piccinno, “Domain-Expert Users and their Needs of Software Development,” *Proc. 2nd International Conference on Universal Access in Human-Computer Interaction*, Lawrence Erlbaum Associates, Inc, 2003, pp. 532-536.
- [22] H. Lieberman, F. Paternò, M. Klann and V. Wulf, “End-User Development: An Emerging Paradigm,” *End User Development*, Human-Computer Interaction Series, H. Lieberman, F. Paternò and V. Wulf, eds., Springer, 2006, pp. 1-8.
- [23] F. Cabitza, D. Fogli and A. Piccinno, ““Each to His Own”: Distinguishing Activities, Roles and Artifacts in EUD Practices,” *Smart Organizations and Smart Artifacts - Fostering Interaction Between People, Technologies and Processes*, Lecture Notes in Information Systems and Organisation, L. Caporarello, B. Di Martino and M. Martinez, eds., Springer, in print.
- [24] G. Fischer, “End User Development and Meta-Design: Foundations for Cultures of Participation,” *Journal of Organizational and End User Computing*, vol. 22, no. 1, 2010, pp. 52-82.
- [25] M.F. Costabile, D. Fogli, A. Marcante and A. Piccinno, “Supporting Interaction and Co-evolution of Users and Systems,” *Proc. International Conference on Advanced Visual Interface*, ACM Press, 2006, pp. 143-150.
- [26] E.L. Hutchins, J.D. Hollan and D.A. Norman, “Direct manipulation interfaces,” *User Centered System Design: New Perspectives on Human-computer Interaction*, D. A. Norman and S. W. Draper, eds., Lawrence Erlbaum, 1986, pp. 87-124.
- [27] G. Abowd and R. Beale, “Users, Systems and Interfaces: A Unifying Framework for Interaction,” *Proc. VI Conference of the British Computer Society Human Computer Interaction Specialist Group - People and Computers (HCI '91)*, Cambridge University Press, 1991, pp. 73-87.
- [28] P. Bottoni, M.F. Costabile and P. Mussio, “Specification and dialogue control of visual interaction through visual rewriting systems,” *ACM Transactions on Programming Languages and Systems*, vol. 21, no. 6, 1999, pp. 1077-1136.

- [29] G. Bourguin, A. Derycke and J.C. Tarby, "Beyond the Interface: Co-evolution inside Interactive Systems - A Proposal Founded on Activity Theory," *Proc. IHM-HCI*, Springer Verlag, 2001, pp. 297-310.
- [30] S. Bandini and C. Simone, "EUD as Integration of Components Off-The-Shelf: The Role of Software Professionals Knowledge Artifacts," *End User Development*, Human-Computer Interaction Series, H. Lieberman, F. Paternò and V. Wulf, eds., Springer Netherlands, 2006, pp. 347-369.
- [31] M. Gantt and B.A. Nardi, "Gardeners and Gurus: Patterns of Cooperation Among CAD Users," *Proc. ACM Conference on Human Factors in Computing Systems (CHI)*, ACM, 1992, pp. 107-117.
- [32] B. Nardi, *A Small Matter of Programming: Perspectives on End User Computing*, The MIT Press, 1993.
- [33] D. Fogli and A. Piccinno, "Co-evolution of End-User Developers and Systems in Multi-tiered Proxy Design Problems," *End-User Development*, Lecture Notes in Computer Science, Y. Dittrich, M. Burnett, A. Mørch and D. Redmiles, eds., Springer, 2013, pp. 153-168.
- [34] F. Cabitza and C. Simone, "Building Socially Embedded Technologies: Implications on Design," *Designing Socially Embedded Technologies: A European Challenge*, D. Randall, K. Schmidt and V. Wulf, eds., Springer, in print.
- [35] D. Fogli, "Towards a new work practice in the development of e-government applications," *Electronic Government, an International Journal*, vol. 10, no. 3, 2013, pp. 238-258.
- [36] S. Carmien, M. Dawe, G. Fischer, A. Gorman, A. Kintsch and J.F. Sullivan Jr., "Socio-technical environments supporting people with cognitive disabilities using public transportation," *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 2, 2005, pp. 233-262.
- [37] G. Fischer, E. Giaccardi, Y. Ye, A. Sutcliffe and N. Mehandjiev, "Meta-design: a manifesto for end-user development," *Communications of the ACM*, vol. 47, no. 9, 2004, pp. 33-37.
- [38] A. Bruns, *Blogs, Wikipedia, Second Life, and beyond : from production to produsage*, Peter Lang Publishing, 2008, p. 418.
- [39] C. Ardito, P. Buono, M.F. Costabile, R. Lanzilotti and A. Piccinno, "End users as co-designers of their own tools and products," *Journal of Visual Languages & Computing*, vol. 23, no. 2, 2012, pp. 78-90.
- [40] G. Fischer, A. Piccinno and Y. Ye, "The Ecology of Participants in Co-evolving Socio-technical Environments," *Engineering Interactive Systems*, Lecture Notes in Computer Science, P. Forbrig and F. Paternò, eds., Springer, 2008, pp. 279-286.
- [41] D. Vinck and E. Blanco, *Everyday Engineering: An Ethnography of Design and Innovation*, MIT Press, 2003.
- [42] G.C. Bowker and S.L. Star, *Sorting Things Out: Classification and Its Consequences*, MIT Press, 1999.
- [43] D. Fogli and L. Parasiliti Provenza, "A meta-design approach to the development of e-government services," *Journal of Visual Languages and Computing*, vol. 23, no. 2, 2012, pp. 47-62.
- [44] F. Cabitza, A. Cerroni and C. Simone, "Knowledge Artifacts within Knowing Communities to Foster Collective Knowledge," *Proc. International Working Conference on Advanced Visual Interfaces (AVI)*, ACM, 2014, pp. 391-394.
- [45] F. Cabitza, G. Colombo and C. Simone, "Leveraging underspecification in knowledge artifacts to foster collaborative activities in professional communities," *International Journal of Human-Computer Studies*, vol. 71, no. 1, 2013, pp. 24-45.
- [46] F. Cabitza and C. Simone, "Affording Mechanisms: An Integrated View of Coordination and Knowledge Management," *Computer Supported Cooperative Work (CSCW)*, vol. 21, no. 2-3, 2012, pp. 227-260.
- [47] C. Ardito, P. Buono, M.F. Costabile, R. Lanzilotti, A. Piccinno and L. Zhu, "On the transferability of a meta-design model supporting End-User Development," *Universal Access in the Information Society Journal (UAIS)*, in print.
- [48] M.S. Ackerman, "The intellectual challenge of CSCW: the gap between social requirements and technical feasibility," *Hum.-Comput. Interact.*, vol. 15, no. 2, 2000, pp. 179-203.
- [49] C.S. De Souza and C.F. Leitão, *Semiotic engineering methods for scientific research in HCI*, Morgan & Claypool Publishers, 2009.
- [50] V. Wulf, V. Pipek and M. Won, "Component-based tailorability: Enabling highly flexible software applications," *International Journal of Human-Computer Studies*, vol. 66, no. 1, 2008, pp. 1-22.
- [51] D. Avola, P. Bottoni and R. Genzone, "Light-Weight Composition of Personal Documents from Distributed Information," *End-User Development*, Lecture Notes in Computer Science, M. F. Costabile, Y. Dittrich, G. Fischer and A. Piccinno, eds., Springer Berlin Heidelberg, 2011, pp. 221-226.
- [52] B. De Silva and A. Ginige, "Meta-model to Support End-User Development of Web Based Business Information Systems," *Web Engineering*, Lecture Notes in Computer Science, L. Baresi, P. Fraternali and G.-J. Houben, eds., Springer, 2007, pp. 248-253.
- [53] D. Fogli and L. Parasiliti Provenza, "End-User Development of e-Government Services through Meta-modeling," *End-User Development*, Lecture Notes in Computer Science, M. F. Costabile, Y. Dittrich, G. Fischer and A. Piccinno, eds., Springer, 2011, pp. 107-122.
- [54] M. Spahn and V. Wulf, "End-User Development for Individualized Information Management: Analysis of Problem Domains and Solution Approaches," *Enterprise Information Systems*, Lecture Notes in Business Information Processing, J. Filipe and J. Cordeiro, eds., Springer, 2009, pp. 843-857.
- [55] G. Ghiani, F. Paternò and L.D. Spano, "Creating Mashups by Direct Manipulation of Existing Web Applications," *End-User Development*, Lecture Notes in Computer Science, M. F. Costabile, Y. Dittrich, G. Fischer and A. Piccinno, eds., Springer, 2011, pp. 42-52.
- [56] J. Bolmsten and Y. Dittrich, "Infrastructuring When You Don't – End-User Development and Organizational Infrastructure," *End-User Development*, Lecture Notes in Computer Science, M. F. Costabile, Y. Dittrich, G. Fischer and A. Piccinno, eds., Springer, 2011, pp. 139-154.
- [57] M.S. Ackerman, C.A. Halverson, T. Erickson and W.A. Kellogg, eds., *Resources, Co-Evolution and Artifacts: Theory in CSCW*, Springer, 2008.
- [58] A. Dix, "Designing for appropriation," *Proc. 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 2*, British Computer Society, 2007, pp. 27-30.
- [59] R. Kling, H. Rosenbaum and S. Sawyer, *Understanding and Communicating Social Informatics: A Framework for Studying and Teaching the Human Contexts of Information and Communication Technologies*, Information Today Inc., 2005, p. 216.
- [60] A. Hevner and S. Chatterjee, eds., *Design Research in Information Systems*, 22, Springer US, 2010.
- [61] O. Hanseth and C. Ciborra, eds., *Risk, complexity and ICT*, Edward Elgar, 2007.
- [62] A. Akera and W. Aspray, eds., *Using History to Teach Computer Science and Related Disciplines*, Computer Research Association, 2004.
- [63] A. Dunne and F. Raby, *Design Noir: The Secret Life of Electronic Objects*, Springer, 2001.
- [64] L.A. Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions*, Cambridge University Press, 2007.
- [65] L. Nelson, *Socratic Method and Critical Philosophy: Selected Essays*, Dover Publications, 1965.
- [66] J. Orr, *Talking about Machines: An Ethnography of a Modern Job*, Cornell University Press, 1996.
- [67] E.S.K. Yu and J. Mylopoulos, "Understanding "why" in software process modelling, analysis, and design," *Proc. Proceedings of the 16th international conference on Software engineering*, IEEE Computer Society Press, 1994, pp. 159-168.
- [68] H. Barki, G. Pare and C. Sicotte, "Linking IT implementation and acceptance via the construct of psychological ownership of information technology," *Journal of Information Technology*, vol. 23, no. 4, 2008, pp. 269-280.
- [69] M.-K. Stein, R.D. Galliers and M.L. Markus, "Towards an understanding of identity and technology in the workplace," *Journal of Information Technology*, vol. 28, no. 3, 2013, pp. 167-182.
- [70] B. Latour, "Social Theory and the Study of Computerized Work Sites," *Information Technology and Changes in Organizational Work*, W. J. Orlikowski, G. Walsham, M. R. Jones and J. L. Degross, eds., Chapman & Hall, 1996, pp. 295-307.



# Adaptive Difficulty Scales for Parkour Games

Yi-Na Li<sup>1</sup>, Chi Yao<sup>2</sup>, Dong-Jin Li<sup>1</sup>, Kang Zhang<sup>3</sup>

<sup>1</sup> School of Business, Nankai University, China

<sup>2</sup> School of Software and Microelectronics, Peking University, China

<sup>3</sup> School of Software Engineering, Tianjin University, China

**Abstract:** Mobile computer games have become increasingly popular in recent years. A major factor for successful game development is the adequate control of the game's difficulty. This paper discusses justification for dynamic difficulty adaption for Parkour games. It presents an adaptive mechanism for difficulty adjustment in response to the player's run-time performance in the single player mode. The mechanism is based on game content generation techniques, considering constraints for mobile screens. Both the functionality of the game's objects and the player's psychological and behavioral inclinations are taken into consideration. Our preliminary experiment shows that game experiences are significantly enhanced with the adaption mechanism.

**Keywords:** *Mobile games; difficulty adaption; parkour*

## I. INTRODUCTION

Technological advances have made games on smartphones increasingly sophisticated. Customized methodology based on automatic content generation has been introduced into game designs to enhance interactivity and improve responsiveness.

Most mobile games provide the Player versus Environment (PvE) mode for offline scenario. A player completes the missions while progressing the game alone against given obstacles in the game environment. Comparing with Player versus Player (PvP) game mode, PvE mode has no participation from other players who may produce complex variations. Considering the time and labor costs, designing obstacles manually hinders efficient game production while missing rich diversities of obstacle patterns. Repeated game actions and tasks tend to cause boredom to players.

To maximally stimulate players, we introduce rules for automatic content generation which would also reduce the design effort, increase creativity for game content variations, and enhance playability.

Setting difficulty scales statically as in traditional game designs cannot satisfy players' differentiated needs. With such a statically set difficulty scales, each player simply chooses easy, medium and hard, or a few more difficulty scales, before starting. Such a coarse difficulty setting is too simplistic to suit variety of game phases, scenarios and variety types of players. We therefore need to build a dynamic mechanism to

keep the difficulty level just as challenging as attainable by the player.

This paper reports our design and experience with two editions of the Dragon One game, a type of Parkour games, using both a self-adaptive difficulty mechanism and a fixed difficulty mechanism. The former is developed for better user experiences. Various game scenarios and typical features are also taken into consideration. The next section reviews the related work, and Section 3 introduces the mobile Dragon One game. Section 4 analyzes various influences on a game's difficulty. Section 5 describes our difficulty adaptation mechanism, followed by the evaluation results in Section 6. Section 7 concludes the paper.

## II. RELATED WORK

### A. Difficulty Scales of Game Design

Difficulty scaling is one of the most important issues in game designs [2]. Players have fun by continually attempting to overcome obstacles [22] with increasing difficulty. The enjoyment of game stems from the levels of challenge, curiosity and fantasy [8]. Challenge is the key criteria in Game-Flow Model for evaluating game enjoyment [17].

The theory of flow has been introduced into game design by Csikszentmihalyi [5]. This theory describes the mental state when a person is fully immersed in a feeling of energized focus or utterly concentrated in the process of an activity for its own sake. People would experience a distorted sense of time and a lack of self-consciousness. Such a status has become the pursuit of user-centered game design [17]. Simple and repeated tasks would lead to boredom and mindless, while excessive hardness would cause frustration. It is important to set challenging but attainable task flow adaptive to the player's ability.

A good game design would follow an escalating difficulty curve [3] [10]. However, players' perceived difficulties may be different from the predetermined difficulty. Assessment for game playing experiences has both objective and subjective dimensions [9]. Perceived difficulties are essentially determined by players' learning abilities, operation strategies, preference of risks, reaction times and team cooperation. Predetermined difficulties are set by default parameters at each level, such as the maximum speed of protagonists, acceleration abilities, endurances and attritions. A difficulty

increment might be too little for a risk-taking experienced player, too challenging for a risk aversion novice. Therefore, an optimized and adaptable difficulty solution customized for different types of players would promote game experience and replayability.

### B. Procedural Content Generation

Procedural content generation (PCG) has been defined as “the algorithmic creation of game contents with limited or indirect user input” [18]. Just-in-time PCG uses a player’s run-time data during the game, rather than historical data, to generate game contents dynamically. It has become a mainstream of intelligent games with the support of customization. Previous works realized automatic mission, procedure and space generation [12], relying on users’ behavior models based on skills [7], experiences [4] [13] [25] [14], risk preferences [6], and immersion in different difficulty scenarios. Apart from the player’s individual behavior, data from the user community are also collected and analyzed to support automatic generation [21] [23].

Previous research conducted case studies in knowledge learning game [1], puzzle game [16], racing game [19], action adventure game, and etc., proposed generating, adapting, optimizing mechanism and game difficulty balancing mechanism [26][20][11][24] by introducing neural network techniques and theory of incongruently in other disciplines, paving the way to automatic generation tailoring players’ needs.

However, none of the aforementioned studies discusses the interaction and balance mechanisms between the player’s instant behavior and the difficulty level. We consider a self-adaptive difficulty adjustment mechanism, in response to the player’s dynamic performance to be extremely important. Our work introduces dynamic difficulty adjustment in response to the player’s performance, and extends the literature on 2D Parkour games for mobile devices.

### III. DRAGON ONE GAME FOR MOBILE DEVICES

This section presents a simple mobile game Dragon One developed by our team for experiments. Several scenarios of the game are shown in Figure 1.

*Game objective:* The player needs to pass through all the obstacles and collect as many fortunes as possible along the way.

*Narration:* It is a type of Parkour game with a simple story line. A Chinese witch builds a food streamer of thousand layers to satisfy her gluttony. She puts all her preys in the food streamers as her food. One day, her cat, who is kind-hearted in day time, but evil at night, breaks the iron chain of a layer. One of the preys, a cute and innocent dragon, gets a chance to escape. He starts his journey through shots and obstacles to pursue his freedom and rescue other preys.

*Task and manipulation:* The player has to control the dragon’s movement in horizontal directions at a constant speed without delay, to run a long distance and earn as bonuses as possible. If killed by an obstacle, the game would

restart from the same difficulty level, but not the same contents. The player needs to continuously make quick decisions by clicking the left or right arrow.

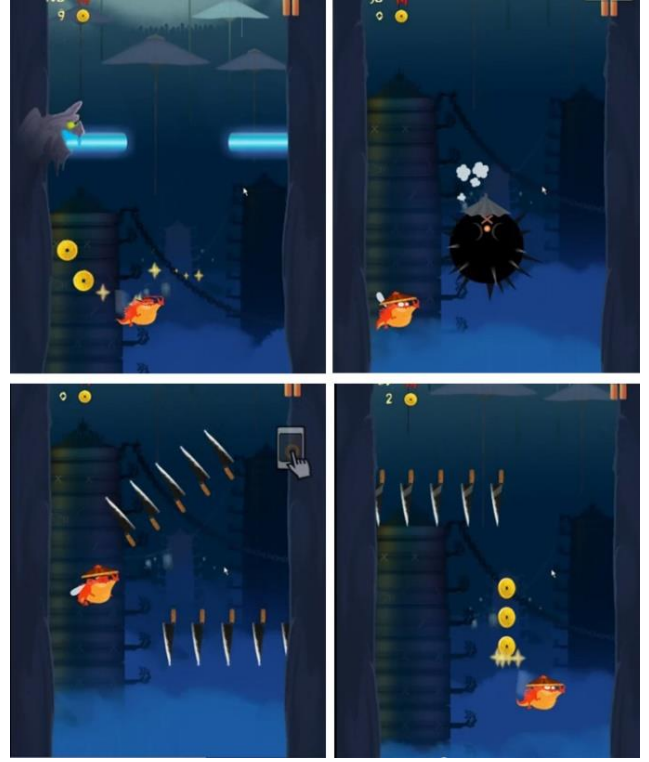


Figure 1. Game interface and scenarios of Dragon One

*Obstacles and bonuses:* Obstacles include knives, bombs and lasers shots. They are lethal once hit. The bonus objects include golden coins and white breads. Collecting as many golden coins as possible for good fortune is one of the game’s objectives. A white bread is an elixir of rebirth when killed by an obstacle. The fortune value, an indication of operation skills, could not be used to exchange for any rebirth chance or extra protection. In spite of this, players would maintain their desire for fortune, trying not to abandon any piece of fortune. The functions of obstacles and bonuses are briefly illustrated in Table 1.

TABLE I. FUNCTIONS OF OBSTACLES AND BONUSES

Type	Item	Function
Obstacle	Flying Knives	Action 1: 5 knives work in one group, reciprocating moving, falling down very fast in vertical direction. Action 2: The knife group revolves when falling down.
	Bomb	Explode with thorns at a fixed height on screen. Two bombs set abreast with a gap through which the dragon can pass.
	Laser Shots	Shoot from left or right side.
Bonus	Golden Coins	Fortune value increases when hitting it.
	White Bread	Get a rebirth chance when hitting it.

*Fixed difficulty experiments:* The game with fixed difficulty randomly invokes 7 manually designed sessions. Each session consists of all the functions in Table 1, lasting about 7 seconds.

The game is reported to be very challenging at the beginning. Players make progress fast after practicing a while, because necessary trainings are simply about repeated patterns. Players get bored very soon after being able to handle the fixed difficulty without much effort. We, therefore, further extended the game with the difficulty levels dynamically adaptable to the player's performance, in order to maximally engage the player and extend the playtime.

#### IV. ANALYSIS OF INFLUENCES ON GAME DIFFICULTY

##### A. A Comparative Study

We extract generally applicable rules to estimate the difficulty in a Parkour game, by comparing *Dragon One* with *Super Mario*, *Man on the Next Layer 100* and the most difficult game *Flappy Bird*. *Super Mario* developed by Nintendo has been popular for almost 30 years. Mario's adventure in Mushroom Kingdom has been considered a masterpiece of Parkour games, providing significant reference of the setting mechanism for tasks, goals, obstacles, bonuses and rebirth functions. However this game is not designed specifically for mobile devices. The *Man on the Next Layer 100* is a 2D action adventure game, developed for mobile devices. Players need exquisite skills to keep on jumping down onto lower floating layers, avoiding falling to the lifeless abyss, or being crushed on the ceiling for sluggish movement. For an ordinary player, this game would last less than 3 minutes. It does not provide a scale of difficulty with step-up sessions. The fixed difficulty sometime irritates players with lower risk preference and lower adaptive ability. Similarly, *Flappy Bird*, a grueling difficulty game commented by game geeks in an online forum, are flocked as touchstone for game intelligence. The first two obstacles could stop more than 95% players. Parkour games tend to control the playing time for each round by setting the difficulty scale. Very brief playing time would facilitate mobile users' fragmented time, but hard to reach the climax of enjoyment.

Game designers tend to follow the "design for segment market" principle. Games are designed for either casual players or hardcore players in a hegemonic manner. Players have to adapt themselves to follow the game rules. The dichotomy of players' categorization ignores the diversity of players and also individual players' prior training. An adaptive mechanism is therefore helpful to change the user-unfriendly impressions.

##### B. Difficulty vs End of Game

Parkour games bear remarkable resemblance between each other in terms of continual operations, penalties and rewards. Except the difference of scenes, storylines and visual traits of items, difficulty setting is the key to game experiences.

Players have to keep on making right decisions in given reaction time. Generally, a Parkour game has high risk with no sunk cost for upgrade or accumulated bonus. The attrition value is as high as 100%. One accidental failure is fatal. It is almost impossible to survive after a wrong decision.

Comparing with the game of lower attrition setting, every moment running a Parkour game is at the door of death. Players have to stay highly concentrated. Accommodating for extremely high risks, low sunk cost makes players not worried about the consequence of failure. The perceived difficulty would not increase when approaching the end of endurance. The potential influences from different stages of endurance thus are excluded.

In addition, the protagonist's speed is constant, without acceleration or buffering actions, which further simplified the game difficulty mechanism. By eliminating possible influences, we can ascribe the difficulty into the reaction time and decision making.

##### C. Objective Influences on Difficulty

The objective difficulty is determined by the physical attributes of game tasks, essentially the speed and complexity. For *Dragon One*, the falling speeds of obstacles are basic criteria for evaluating difficulty. Mixed types of obstacles require different amounts of time for analysis and decision making. When a bonus appears, a concern to balance the risk and fortune further increases the complexity for decision making.

##### D. Subjective Influences on Difficulty

Subjective influences emphasize on players' psychology and behaviors. Players who are highly tempted to purchase fortune could be terribly misled. The floating bonus items suggest seductive routes, which may lead to unpredictable risks. Moreover, players are prone to make more mistakes when they become mindless. A critically difficult task after a few easier and relaxed tasks require players to be on high alert. Those tricky conditions may deviate players' performance from usual.

To summarize, we should employ both objective and subjective influences to estimate and adjust the difficulty scale of a Parkour game. The next section introduces the basic rules for generating game contents and constraint conditions for determining difficulty levels.

#### V. DIFFICULTY ADAPTION MECHANISM

We design the adaption mechanism based on the number of obstacles passed, e.g. the difficulty level increments at every 15 obstacles as shown in Table 1.

In the flowchart in Figure 2, parameter setting and difficulty increment are the key in the adaption mechanism. We assume that the sizes of the items are fixed. Variable parameters include the falling speeds of items and the space between them.

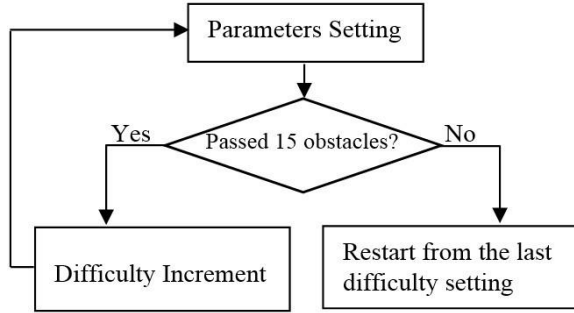


Figure 2. A simple mechanism for difficulty adaption

#### A. Parameter Setting for Objective Difficulty

We generate game contents by randomly arranging obstacles and bonuses using a constraint condition filter. The mobile screen on which our games are implemented measures 4 inches (640×1136 pixels). The dragon's function area is presumed to be a circle, whose diameter is 95 pixels. We defined the diameter of bomb 190 pixels, knife group 260×110 pixels, diameter of the revolving knife group 280 pixels (as shown in Figure 3). The obstacle and bonus items are lined up vertically. The initial falling speed is 384 pixels (1/4 screen) per second. The initial average reaction time on one item is about one second.

We designed an obstacle insertion rule, by inserting an obstacle into a random horizontal position. Players may take advantage of random content generation by staying at either the left or right side of the screen. Without continuing operation, the fun no longer exists. Therefore, we make an obstacle to appear to steer an active action if the dragon has been along the left or right border more than 1.5 seconds.

Expert players tend to control the dragon by flying around 1/5 of the screen height, to pre-judge the approaching danger and optimize his/her strategy according to multiple items. It means that, at each moment of decision making, players process the visual information on the screen as organized patterns, rather than independent individual objects. Therefore, we analyze various composition methods and corresponding solutions, as the determining basis for difficulty control. There are 3 kinds of obstacles, in 5 functions, as listed in Table 1. We abstract the 5 functions into geometric objects, shown in Figure 3. No two consecutive obstacles having the same function would appear in the game. The compositions of the same difficulty functions are excluded in our design. There are potentially 10 possible compositions with every two types of obstacle functions. The top-down reversal compositions obey the same rules.

We define the minimum vertical space value between two consecutive items as interval, denoted  $l$ , i.e. the height of the gateway between two items, from the bottom edge of the above item to the top edge of the item below. Assume that the diameter of the dragon's active area and the diameter of a bomb as  $x_0$  and  $x_2$  respectively, the length of a knife group and the length of a laser as  $x_1$  and  $x_3$  respectively. The  $l$  value could be calculated under the given patterns as Figure 4.

The horizontal distance between items also impacts on the minimum  $l$ . If the horizontal distance is large enough, or the space between items is not the only route for the dragon,  $l$  could be any value. The patterns in Figure 4 represent the most difficulty compositions when the horizontal distance is at minimum while the interval is also at minimum. If a pair of randomly generated obstacles is eligible for the criteria of patterns in Figure 4, a smaller interval value in horizontal direction would produce higher difficulty.

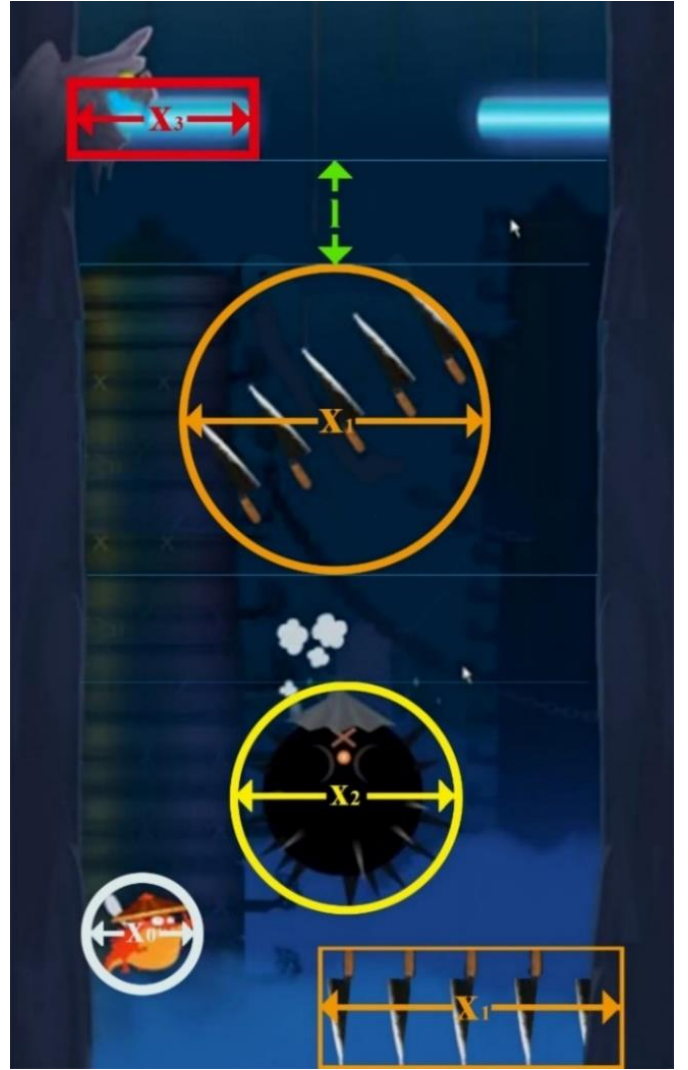


Figure 3. Variables related to obstacles and difficulty

The interval between two items is set between 0-320 pixels. There are rules for the minimum interval values for special cases, which make sure that it is possible for the dragon to get through the toughest obstacle combination. We set the initial interval value at  $4 l_{\min}$ . The interval value decreases when upgrading the difficulty level.



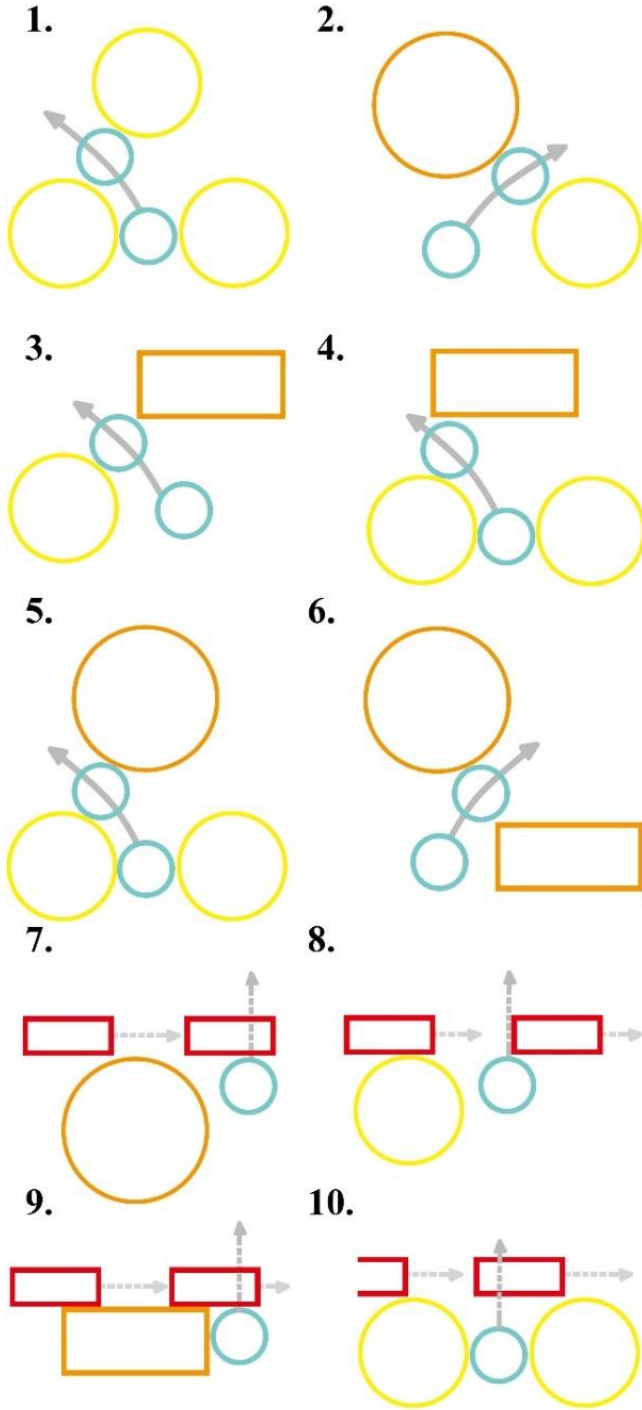


Figure 4. Patterns of obstacles and dragon's movement (blue)

For Pattern 1 in Figure 4,  $l \geq \sqrt{3} \frac{(x_0+x_2)}{2} - x_2$ .

For Patterns 2 and 5, assume the centers of the circles are  $(\alpha_1, \beta_1)$  and  $(\alpha_2, \beta_2)$ ,

$$l \geq \sqrt{\left(\frac{x_1+x_2}{2} + x_0\right)^2 - (\alpha_1 - \alpha_2)^2} - \frac{x_1+x_2}{2}.$$

For Patterns 3, 4, and 6, assume that the lower left corner of the rectangular in Patterns 3 and 4, and upper left corner in Pattern 6 are  $(\alpha, \beta)$ , the center and radius of the circle are  $(a, b)$  and  $r$  ( $r = \frac{x_2}{2}$ ) respectively in Patterns 3 and 4;  $r = \frac{x_1}{2}$  in Pattern 6),

$$l \geq \sqrt{(r + x_0)^2 - (\alpha - a)^2} - r.$$

For Patterns 7-10,  $l \geq 0$ . Players have to find a chance to get through.

#### B. Control for Subjective Difficulty

The level of danger or risk is positively proportional to the level of awards in games and also in life. A long string of bonuses suggests a misleading route. For example, the golden coins arranged along a curve would lead to a corner where it is too late to survive by altering the trajectory. Players need more effort and skills to deal with temptations and potential risks, and foresee the exact time to give up the pursuing of bonuses that may worth maximum rewards. Both golden coins and white breads could be used as baits to elevate subjective difficulty.

Golden coins are usually organized as a line or a matrix. For obstacles, except laser, when the left side of the obstacle  $(p, y_1)$ ,  $x_0 \geq p \geq 0$ , the interval value is 0, a coin line leading to  $(q, y_2)$ ,  $p \geq q \geq 0$ , would be tricky. It works the same way for the right side.

A white bread occurs alone. As a bait, a white bread's position is determinant. An expert player is skilled at planning the route to earn a chance of rebirth. Any white bread at an obviously fatal position would be ignored without any attempt. White breads are therefore placed in reachable positions to raise the player's temptation as traps. The trade-off of risks and gains requires rapid and intuitive decision making with exquisite skills.

Assume the left side of the obstacle to be  $(u, y_1)$ , the interval value  $l$ ,  $0 \leq u \leq x_0$ , the radius of the circle  $r$  (could be  $\frac{1}{2} x_1$  or  $\frac{1}{2} x_2$ ), the horizontal speed of the dragon  $v_1$ , and the vertical speed  $v_2$  (i.e. the falling speed of the items). The horizontal value of the critical position would be  $u + 2r - \frac{lv_1}{v_2}$ . The measurements on the right side are calculated similarly.

When an obstacle and a bonus item are next to each other, if  $l=0$ , the game difficulty is raised to the maximum. The value of  $l$  mediates the difficulty for collecting the fortune.

Bonuses are introduced after the first 30 obstacles have been cleared, and then the rate of bonuses appearing will escalate to average 1 per 2 screen length. The type of subjective difficulties would be random, with coins and white breads as baits at misleading positions.

#### C. Difficulty Scale

We define the cases with maximum falling speed, minimum interval space, highest rate of obstacle compositions as shown in Figure 4, setting the highest frequency of misleading bonuses as the hardest level of difficulty. The difficulty of the game is measured by

$$F(x) = af(v) + bf(l) + cf(h) + df(p),$$

where  $v$  is the falling speed of items,  $l$  is the time interval between two consecutive falling items,  $h$  is the proportion of 10 most difficult compositions in all the appeared compositions, and  $p$  is the frequency of misleading situations,  $a$ ,  $b$ ,  $c$ , and  $d$  are coefficients.

Assume that the difficulty is divided into  $n$  levels and set at 20 initially. Parameter  $n$  could be adjusted according to the player's performance. Upon three times of failure in the same level,  $n$  is automatically set equivalent of 1.1 times of the original value. The amount of levels could be  $20 \times 1.1^n$  ( $20 \leq n \leq 0$ ), i.e. between 20 -120.

For each level up, the difficulty increment is

$$\frac{1}{n} [F(x)_{\max} - F(x)_{\min}].$$

The difficulty increment is therefore more delicate and sensitive to the player's instant performance as desired. After a period of practice, players may reach their learning plateau, the game difficulty remains at the challenging status until players make another break through.

Parkour games usually do not provide tutorials or pretest level for novices due to their simple and intuitive operations. Once the fixed difficulty is no longer challenging enough, the game would be abandoned soon. Therefore, an adaptive difficulty mechanism not only serves as the guidance to the novice, but also stimulates experienced players for upgrading progress.

## VI. PRELIMINARY EVALUATION

We conducted an experiment to evaluate the effects of the difficulty adaptation mechanism on different types of players and report our findings in this section.

### A. Experimental Setup

We invited six volunteers as the experimental subjects whose profiles are shown in Table 2. The subjects are aged from 28 to 58, with the education levels from high school to Ph.D. The skill levels are determined by the subjects' self-evaluation based on their previous experiences in playing similar games. There are five levels, from 1 to 5, representing "never played mobile game before", "beginning player", "average player", "skilled player" and "hardcore player".

TABLE II. SUBJECTS' PROFILES

ID	Gender	Age	Occupation	Skill level
1	Female	30	Ph.D. in Arts, Researcher	3
2	Female	28	Graduate student in Computer Science	3
3	Male	33	Game designer	5
4	Male	30	Ph.D. in Biology, Researcher	4
5	Male	58	Property company manager	1
6	Male	35	Public security guard	2

We handed over a smart phone installed with Dragon One and asked each subject to play the game for at least 15 rounds, and recorded the score in each round.

### B. Results

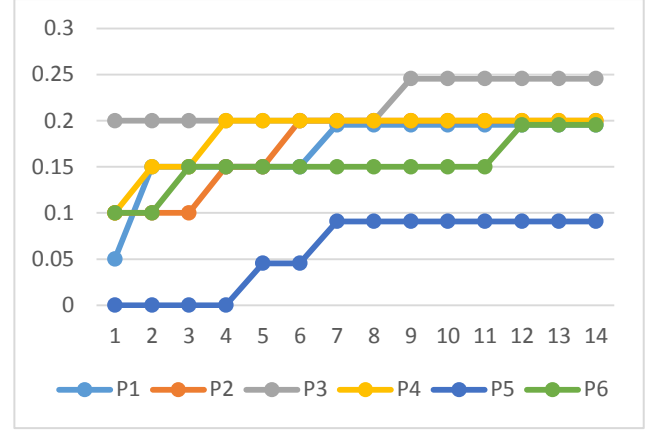


Figure 5. Difficulty levels for the first 15 rounds

We infer the subjects' difficulty curves from their scores, and present them in Figure 5. The horizontal axis indicates the 15 game rounds, and the vertical axis measures the difficulty.

The six players' difficulty curves are significantly different. The hardcore player (P3) got a high score in the first round. His difficulty value raised to 0.2. By the ninth round, his difficulty value reached 0.25. In contrast, the beginning player P5 struggled to pass the first level until the fourth round. In fact, he played for quite a long time at 0.0909 difficulty value after his extraordinary performance. The other players also got appropriate challenges after a few rounds of playing. The difficulty adaptation mechanism has satisfied the disparity of players' expectations.

Game experiences are optimized accordingly with more delicate and sensitive difficulty increments. For example, if player P5 kept on practicing before each upgrade, his difficulty increment at level  $n$  would be  $0.05 \times 1.1^{1-n}$ , far less than the initial increment value of 0.05. The slow learner would not be too frustrated to continue.

It is increasingly difficult to upgrade after reaching the plateau adaptive to the player's skill and ability. Therefore, the difficulty curve would be flat for many rounds to come.

The adaptive difficulty makes the difference in players' incentives. Players in fixed difficulty scenarios are informed of their progress by the increase of the distance and fortune in each round. But in the adaptive difficulty setting, the reference criteria have to be about the upgrade information. If the player's level setting is also customized, the incentive information could be set as the percentile among peer players apart from the player's own experience.

## VII. FUTURE WORK

Galvanic skin response, heart rate and muscle movement data have been used to assist the design of game adaption



mechanisms. Such interaction data could be used in Parkour games to generate customized experience and suit for more complicated playing modes, such as two players in competition or cooperation.

Additionally, the sizes of items in a Parkour game can be varied, as variable sized obstacles require more sophisticated strategies and generate new and exciting game experiences. The increment of item sizes would decrease the available safe areas, and increase the difficulty. The enhanced game difficulty adaption mechanism and its application in marketing research are our future research focus.

#### ACKNOWLEDGEMENT

The third author is partially supported by the National Natural Science Foundation of China (Grant number: 71372099).

#### REFERENCES

- [1] Andersen, E. Optimizing adaptivity in educational games. *Proc. Foundations of Digital Games 2012*, ACM Press, 279-281, 2012.
- [2] Aponte, M., Levieux, G., and Natkin, S., Difficulty in Videogames: An Experimental Validation of a Formal Definition, *Advances in Computer Entertainment Technology*, page 49, ACM Press, 2011.
- [3] Byrne, E., *Game Level Design* (Game Development Series). Charles River Media, December 2004.
- [4] Bakkes, S. C., Spronck, P. H., and Lankveld, G., Player Behavioral Modelling for Video Games. *Entertainment Computing*, 3: 71-79, 2012.
- [5] Csikszentmihalyi, M., *Flow: The Psychology of Optimal Experience*, HarperCollins e-books, 2008.
- [6] Hawkins, G., Nesbitt, K., and Brown, S., Dynamic Difficulty Balancing for Cautious Players and Risk Takers, *International Journal of Computer Games Technology*, Vol. 2012, Article ID 625476, 10 pages, 2012.
- [7] Lankveld, G. van, Spronck, P., Herik, H.J. van den & Rauterberg, G.W.M. Incongruity-based adaptive game balancing. In P. Spronck & H.J. van den Herik (Eds.), *Proc. ACG 2009, LNCS 6048*, Springer-Verlag, 208-220, 2010.
- [8] Malone, T. W., What makes things fun to learn? Heuristics for designing instructional computer games, *SIGSMALL'80: Proc. 3rd ACM SIGSMALL Symposium and the first SIGPC Symposium on Small Systems*, 162-169, ACM Press, 1980.
- [9] Nacke, L. E. Affective Ludology: Scientific Measurement of User Experience in Interactive Entertainment. Karlskrona, Sweden: *Blekinge Institute of Technology*, 2009.
- [10] Natkin, S., Delocque-Fourcaud A., and Novak, E., Video Games and Interactive Media: A Glimpse at New Digital Entertainment. AK Peters Ltd, 2006.
- [11] Nelson, M. J., and Mateas, M. Towards automated game design. *Proc. AI\*IA 2007: Artificial Intelligence and Human-Oriented Computing*, Springer 626-637, 2007.
- [12] Nitsche, M., Ashmore, C., Hankinson, W., Fitzpatrick, R., Kelly, J., and Margenau, K. Designing procedural game spaces: A case study. *Proc. FuturePlay 2006*, 2006.
- [13] Pedersen, C., Togelius, J., and Yannakakis, G. N. Modeling player experience in Super Mario Bros. *Proc. IEEE Symposium on Computational Intelligence and Games*, 132-139, 2009.
- [14] Qin, H., Patrick P. R., Salvendy G., effects of different scenarios of game difficulty on player immersion, *Interacting with Computers*, Vol. 22, No. 3, 230-239, 2010.
- [15] Pedersen, C., Togelius, J., and Yannakakis, G. N. Modeling player experience in Super Mario Bros. *Proc. Computational Intelligence and Games 2009*, IEEE Press, 132-139, 2009.
- [16] Smith, A. M., Andersen, E., Mateas, M., and Popović, Z. A case study of expressively constrainable level design automation tools for a puzzle game. *Proc. Foundations of Digital Games 2012*, ACM Press, 156-163, 2012.
- [17] Sweetser, P., and Wyeth, P., Game flow: a model for evaluating player enjoyment in games. *Computer Entertainment*, 3(3):3, 2005.
- [18] Togelius, J., Kastbjerg, E., Schedl, D., and Yannakakis, G. N. What is procedural content generation? Mario on the borderline. *Proc. Procedural Content Generation in Games 2011*, ACM Press, 2011.
- [19] Togelius, J., Nardi, R. D., and Lucas, S. M. Towards automatic personalised content creation for racing games. *Proc. Computational Intelligence and Games 2007*, IEEE Press 252-259, 2007.
- [20] Togelius, J. and Schmidhuber, J., An experiment in automatic game design. *Proc. Computational Intelligence and Games 2008*, IEEE Press. 111-118, 2008.
- [21] Togelius, J., Yannakakis, G., Stanley, K., and Browne, C., Search-based Procedural Content Generation: A Taxonomy and Survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(99), 2011.
- [22] Vorderer, P., Hartmann, T., and Klimmt, C., Explaining the enjoyment of playing video games: the role of competition. *Proc. 2nd International Conference on Entertainment Computing (ICEC'2003)*, Carnegie Mellon University, 1-9, 2003.
- [23] Weber, B. G., Mateas, M., and Jhala, A. H., Using Data Mining to Model Player Experience. *Proc. FDG Workshop on Evaluating Player Experience*, 1-9, Carnegie Mellon University, 2011.
- [24] Yannakakis, G., and Hallam, J., Real-time game adaptation for optimizing player satisfaction. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(2):121-133, 2009.
- [25] Yannakakis, G., and Togelius, J., Experience-Driven Procedural Content Generation. *IEEE Transactions on Affective Computing*, 2:147-161, 2011.
- [26] Zook, A., and Riedl, M., Generating and Adapting Game Mechanics. *Proc 2014 Foundations of Digital Games Workshop on Procedural Content Generation in Games*, Ft. Lauderdale, Florida, 2014.

# “Magic Mirror in my Hand, what is the Sentiment in the Lens?”: an Action Unit based Approach for Mining Sentiments from Multimedia Contents

Luca Casaburi<sup>1</sup>, Francesco Colace<sup>1</sup>, Massimo De Santo<sup>1</sup>, Luca Greco<sup>2</sup>

<sup>1</sup>DIEM - Department of Information Engineering, Electrical Engineering and Applied Mathematics

<sup>2</sup>DIIN - Department of Industrial Engineering

Università degli Studi di Salerno

Fisciano (Salerno) - Italy

{lcasaburi, fcolace, desanto, lgreco}@unisa.it

**Abstract** — Emotions are an increasingly important factor in Human-Computer Interaction. So, extracting emotions from multimedia contents is becoming one of the most challenging research topics in Computer Science. Facial expressions, posture, gestures, speech, emotive changes of physical parameter (e.g. body temperature, blush and changes in the tone of the voice) can reflect changes in the user's emotional state. All this kind of parameters can be detected and interpreted by a computer leading to the so-called “affective computing”. Through affective computing, client's posture, gestures, and facial expressions could be used, along with words, for a more accurate evaluation of their psychological state. In this paper an approach for the extraction of emotions from images will be introduced. The proposed framework involves the adoption of action units' extraction from facial expression according to the Ekman theory. The proposed approach has been tested on standard datasets and the results are interesting and promising.

**Keywords** — *Affective Computing, Ekman Theory, Emotional Intelligent*

## I. INTRODUCTION

Affective computing is a kind of human-computer interaction where a device has the ability to detect and appropriately respond to its user's emotions and other stimuli. A computing device with such a capacity could gather cues about user emotion from a variety of sources. Facial expressions, posture, gestures, speech, the force or rhythm of key strokes, temperature changes of the hand on a mouse can all reflect changes in the user's emotional state, and these can all be detected and interpreted by a computer. Affective computing gets its name from the field of psychology (where “affect” is, basically, a synonym for “emotion”) and could offer benefits in an almost limitless range of applications: e-learning, e-health, e-therapy, entertainment, marketing [1][2][3][4][5][9][10][11][12][13][14][58].

The problem of automatic affective recognition through non-verbal communication (facial expression, gestures, movement and posture of the body and hands) is becoming more and more an attractive research topic in recent days.

Several studies [7] have shown that a face, especially a facial expression, can be a powerful communication channel to convey emotions and opinions related to experiences or common situations. A facial expression can be defined as a visible manifestation of the emotional state, cognitive activity, intention, personality and psychology of a person [3]. It is well known that facial expressions contribute strongly to the effect of a multimedia message, more than vocal and verbal part [11].

For mining affective states from multimedia contents the adoption of the Ekman Model is an effective approach [7] [8]. This model can mine six different affective states: happiness, anger, sadness, disgust, fear and surprise. Some studies have focused on enriching this model by introducing particular states such as attention [15], fatigue [16] and pain [17]. In general, the detection of affective states from multimedia contents follows two main approaches: the direct recognition of discrete basic affects (template matching) or the recognition of affects by the inference from movement of facial muscles according to the Facial Action Coding System Coding (FACS) [8]. FACS classifies the facial movement as Action Units (AUs) describing the facial expressions as a combination of AUs.

The first approach requires the execution of two main steps: encoding the face through features (landmarks or filtered images) and classification of facial expression. Many papers deal with this approach such as [18] that shows how to represent the facial expressions in a space of face. The face is encoded as a landmark (58 points) and the classification is performed with a probabilistic recognition algorithm based on the manifold subspace of aligned face appearances. The adoption of a space of faces allows describing the sequences of facial expression [19][20]. Zhang et al. [21] analyzes the space of facial expressions to compare two classification systems (geometric-based (face is encoded by a landmark) and Gabor-based (face is encoded by Gabor features)) and performs the classification with two-layer preceptor network. They show that the best results are obtained with a network of 5-7 hidden preceptors to represent the space of expression. In this way, the facial expression analysis can be performed on static images [21][23] or video sequence [22][24][57]. Cohen et al. [22] proposes a new architecture of HMM to segment and recognize facial expression and affects from video flow, while Lee et al.

[24] proposes a method using probabilistic manifold appearance. Wang et al. [27] describes an automatic system that performs face recognition and affect recognition of grey-scale images of face by making a classification on a space of faces and facial expressions. This system can learn and recognize if a new face is in the image and which facial expression is represented among basic affects. In H. Deng's paper [25] is shown how to choose the Gabor features with PCA method and then LDA is used to identify the basic affects. Bartlett et al. [26] proposes an extraction system of facial expressions from video that chooses the Gabor features with AdaBoost Algorithm and then affects are classified by a SVM. Garbas et al. [31] extracts features from the face through a LBP filter and chooses the most representative features using Real-AdaBoost algorithm. Finally the faces are classified as positive or negative by a binary classifier.

The recognition of affects by the inference from movement of facial muscles according to the Facial Action Coding System Coding (FACS) requires three steps: feature extraction, AUs recognition and basic affect classification. Parts of the face, such as eyebrows, eyes, nose and lips, are analyzed and encoded in sets of points [28][29] or as texture features [3] [17][30] to detect AUs. Cohn et al. [28] introduces a method to detect the AUs starting from eyebrows, classifying their movements as spontaneous or voluntary by the use of a Relevance Vector Machine approach. After the detection of the AUs it classifies their affective class by a probabilistic decision function. The problem is also addressed in the case of rotation of the head [29]. Automated Facial Image System (AFA) [32] analyzes video in real-time to detect the sentiments. In this case, the face is encoded with a 2D mask which is used to interrogate a SVM to detect the associated affect. P. Robinson et al. [33] have developed a system that analyzes real-time video streams to detect the presence of one of the following moods: concordant, discordant, focused, interested, thinking and unsure. The face is encoded by 24 points and the features used are the distances between these points to identify different situations (open mouth, head movements, position of the eyebrows) the expressions encoded by FACS are recognized by a chain of HMM for each possible action and the computation of the probability of each state is obtained by the use of a Bayesian Network.

In this paper a new method is proposed for analyzing facial expressions and recognizing emotion from multimedia contents, in particular images and videos. This method uses AUs detection to recognize basic emotions [18] and implements a new technique for extracting feature points from the face, including an original method for measuring emotion. The classic prototypes have been extended introducing the concept of combinations of AUs: when a combination occurs, a bonus or a penalty is assigned to the measure of emotions. In this way, a more detailed recognition model can be obtained.

The paper is organized as follows: the proposed approach is discussed in the next section. In the second section results of test on CK+ dataset [33], for image analysis, and on MMI Facial Expression [34], eNTERFACE '05 [55] and Cam3D [51] datasets for video analysis are presented. The obtained results are discussed in last section.

## II. THE PROPOSED FRAMEWORK

As previously said, in this paper a system of Facial Expression Analysis, based on the Facial Action Coding System (FACS), is proposed. The proposed framework is organized in three fundamental modules:

- **Features Detection Module:** a face skeleton composed by feature points is obtained from a RGB image.
- **AUs Detection Module:** the probability that a specific action unit has been performed is here calculated. The action units (AUs) are obtained from the position of the feature points in the face skeleton. A vector of pairs (AU, probability) is built as result of this module.
- **Affect Detection and classification:** recognition is carried out with Ekman's prototype. Detected affects are classified according to the Ekman's categories: happy, sad, angry, fear, disgust and surprise

In the next paragraphs the modules will be described.

### A. Features Detection Module

According to Eckman's theory, the feature points of interest are depicted in Fig. 1. The features detection process consists of the following steps (Fig. 2):

- **Face Detection:** the Region Of Interest (ROI) of the image, containing a face, is detected.
- **ROI Selection:** the ROIs of eye, eyebrows, mouth and nose are extracted from the face's ROI.
- **Eye Feature Detection:** feature points of the eyes (points 8 and 9) are identified. These points are useful also to detect feature points of eyebrows and the orientation of the head (roll).
- **Nose Feature Detection:** the feature points of the nose (point 10) are identified. These points are used for detecting the orientation of the head (yaw and pitch).
- **Eyebrow Feature Detection:** the feature points of eyebrows (points 4, 5, 6 and 7) are identified.
- **Mouth Feature Detection:** the feature points of the mouth (points 0, 1, 2 and 3) are identified.

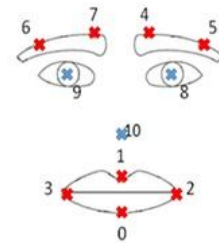


Figure 1 Feature points for emotion extraction.

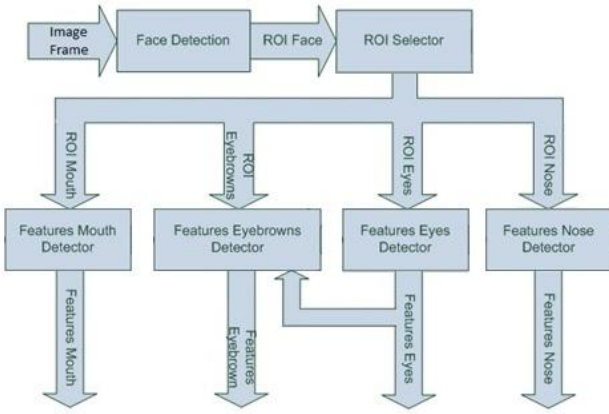


Figure 2 Feature detection process.

### B. Face Detection

The problem of face detection is a specific case of object detection and in literature there are various solutions: the SIFT (Scale Invariant Feature Transform) algorithm [36], SURF (Speeded Up Robust Features) algorithm [37], Haar Cascade, also known as the algorithm of Viola-Jones [38], SEMB-LBP Cascade (Statistically Effective Multi Block Local Binary Pattern Cascade) [39] and SURF Cascade (Speeded Up Robust Features Cascade) [46]. The solutions based on the SIFT and SURF algorithms include a first phase of features extraction and then a second phase of classification of the features that is typically performed using SVM. They have excellent characteristics being:

- Scale invariant;
- Rotation invariant;
- Symmetric invariant;
- Partially invariant to brightness changes;
- Highly repeatable.

Haar Cascade, SEMB-LBP Cascade and SURF Cascade are based on the idea of the cascade classifier [40] and more in particular on the AdaBoost algorithm [41].

According to many studies [42, 43, 44, 45, 46], face detection based on SIFT or SURF obtains higher performance, but are too slow for real-time applications. For our aims we selected four classifiers based on the OpenCV framework [53]: Stump-based Haar Cascade, Tree-based Haar Cascade, SEMB-LBP classifier and SURF Cascade. We compared the four classifiers using the following datasets: IMM Face DB [47], CMU-MIT Face Test Set [48], Caltech Faces 1999 [49], Caltech 10,000 Web Faces [50], Cam3D [51]. The results show that the SURF Cascade has a false positive rate close to zero and has an excellent hit rate for high-quality images and is the best on video with moving subjects. Instead, the Haar Cascade stump-based has the highest hit rate for images both with low and with high resolution.

In addition to precision evaluation, for each algorithm processing time has been calculated using the Caltech dataset. The tests were performed on a personal computer with the

following specifications: CPU: Intel I3-2328M 2.20GHz, RAM: 4GB DDR3, Video Card: NVIDIA GeForce GT 635M, HD: 400GB SATA2, SO: Windows 8 x64. The results are shown in table 4.

Caltech with 450 images		
Classifier	Total time (sec)	Time on 1 image (sec)
Haar Cascade 1	537,17	1,1937
Haar Cascade 2	334,10	0,7424
LBP Cascade	38,43	0,0854
SURF Cascade	36,04	0.0809

Table 2 Test results of the processing times

In conclusion, for the proposed system we chose to adopt the SURF cascade because it shows good performance for face detection on video streams, it is very fast and has a false positive rate close to zero.

### C. ROI Selection

In this module the ROI of eyes, eyebrows, nose and mouth are extracted. The image of the face is divided with a 24x24 grid, each grid cell outlines a part of the face. The set of certain cells defines a search area where the parts of the face may be present. The search areas are the ROIs used in the feature point detection.

#### 1) Eye Feature Detection

Eye Feature Detector receives the ROIs of eyes as input and returns the feature points as output. This module carries out the following phases:

- Eye detection: the position of the eyes is detected using a Haar Cascade classifier. The output is a rectangle that circumscribes the eye by defining its location and size.
- Feature Point Detection: the center of the rectangle is located and corresponds to the center of the pupil.

#### 2) Nose Feature Detection

The module of Nose Feature Detection is similar to Eye Feature Detection, but returns the position of the tip of the nose as a feature point. The performed steps are:

- Nose Detection. The nose is detected by a Haar Cascade classifier: a rectangle is identified and it surrounds the tip of the nose.
- Feature Point Detection. The center of the rectangle is identified as feature point of the nose.

#### 3) Eyebrow Feature Detection

Eyebrow Feature Detection involves the segmentation process of the image to obtain a binarized image of eyebrows. The segmentation algorithm performs these steps:

- Extraction of the red channel;
- Image equalizer. The equalization technique allows to obtain a uniform histogram by redistributing grey levels.
- Thresholding. The binarized image B is derived from the equalized image Ceq in the following way:

$$B(x, y) = \begin{cases} 1 & \text{if } C_{eq}(x, y) > \theta \\ 0 & \text{otherwise} \end{cases}$$

Where

$$\begin{aligned} \theta &= \bar{C} + m \cdot \alpha \cdot \sigma \\ m &= \pm 1 \\ \bar{C} &= \frac{1}{N} \sum_{x,y} C_{exp}(x, y) \\ \sigma &= \sqrt{\frac{1}{N} \sum_{x,y} (C_{exp}(x, y))^2 - \bar{C}^2} \quad \alpha \in [0, 1] \end{aligned}$$

x and y are the coordinates of pixels of the image.

The binarized image is enhanced through dilation with an elliptical kernel. Then the feature points are obtained by taking the projections of the ends of the eyes to the top limit of the eyebrows.

#### 4) Mouth Feature Detection

Mouth feature detection is similar to the eyebrow feature detection. Segmentation algorithm, to obtain the binarized image, includes the following steps:

- Mouth detection. Haar Cascade classifier is used to locate the precise position of the mouth in the ROI.
- Extraction of green channel of the image.
- Calculation of the cumulative probability histogram. The cumulative histogram (CH) is obtained from the histogram of the image (H) in the following way:

$$CH(i) = CH(i-1) + H(i) \text{ with } i = 1, \dots, 255$$

$$CH(0) = H(0)$$

- Thresholding. The binarized image B is obtained from the equalized image C in the following way:

$$B(x, y) = \begin{cases} 1 & \text{if } CH(C(x, y)) - i < \theta \\ 0 & \text{otherwise} \end{cases}$$

where

$$i = 0, \dots, 255; x = 0, \dots, \dim_x(C); y = 0, \dots, \dim_y(C)$$

and x and y are the coordinates of pixels of the image.

The binarized image is improved through dilation with elliptical kernel. Then the Canny algorithm is applied to detect the contours of the mouth. The right, left, top and bottom extremes of these contours are the feature points.

#### D. AU Detection

The feature points can represent the face skeleton and a particular facial expression. The facial expression is described

by only 8 feature points of the eyebrows and mouth. The remaining 3 feature points of eyes and nose, describing the rotation of the head, are used in mathematical calculations performed by the AU detector.

In a neutral expression feature points of eyebrows and mouth are in a well-defined region, while they move out if an AU is performed. AU detector calculates the distance of the points from the neutral region and in this way recognizes the performed AU. The distance is defined as a normalized distance respect to the distance of the pupils, of the feature point from eye line and normal line. The eye line is the segment, which passes through the feature points of eyes, and normal line is the normal of eye line, which passes through the centre of the eye line (fig. 3). Then 3 feature points are used to define the eyeline and normal-line and to calculate the distance of the points from these lines.

The regions of neutral state are identified by an upper and lower limit, which are calculated in the following way:

$$\begin{aligned} position_{new} &= position_{neutral} + \gamma(dif_{UpDown}, dif_{EyesNose}) \\ \gamma(dif_{UpDown}, dif_{EyesNose}) &= \alpha(dif_{EyesNose}) + \beta(dif_{UpDown}) \end{aligned}$$

The “neutral position” distance can be considered constant and it is determined empirically by analyzing different subjects in neutral poses. The values are reported in Table 3.  $\alpha$  and  $\beta$  are two variables and depend on the rotation of the head (yaw and pitch). The following parameters are used:

$$\begin{aligned} dif_{EyesNose} &> 0 \quad \text{head turnd to the right} \\ &< 0 \quad \text{head turned to the left} \\ dif_{UpDown} &> 0 \quad \text{head turnd downward} \\ &< 0 \quad \text{head turned upward} \end{aligned}$$

Where

$$\begin{aligned} dif_{EyesNose} &= (x_{nose} - x_{left\_eye}) - (x_{right\_eye} - x_{nose}) \\ dif_{UpDown} &= (y_{eye\_model\_position} - y_{eye}) - (x_{right\_eye} - x_{nose}) \end{aligned}$$

$\alpha$  and  $\beta$  are functions that vary according to the AU. These functions have been obtained by inferring mathematical models empirically through a dataset of images designed specifically for this purpose.

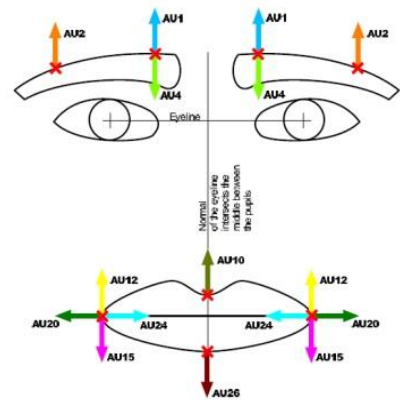


Figure 3 AU Detection



**Table 5 Bonus and penalty**

AU	Lower limit	Upper limit	Reference line
1	0.238	0.392	Eyeline
2	0.278	0.472	Eyeline
4	0.228	0.188	Eyeline
10	0.902	0.825	Eyeline
12	1.086	0.822	Eyeline
15	1.056	1.111	Eyeline
20	0.556	0.583	Normal line
24	0.415	0.276	Normal line
26	1.203	1.284	Eyeline

**Table 3 Thresholds for the neutral regions**

#### E. Affect Detection and Classification

For the affect detection, Ekman's prototypes have been modified. These variants calculate a measure, between 0 and 1, that identifies if a particular affect is detected (Table 4). If particular combinations of AU arise, the result of these adapted prototypes is amended by adding or subtracting a score. This technique allows increasing the difference of the measures of affect from the higher.

Working on the CK+ dataset, we found that in the case of a particular emotion the relative AUs of prototypes occur at their high intensity while the other ones are at their low intensity. For example, if a person smiles, AU12 at high intensity and AU10 and AU20 at low intensity are detected, then the subject could be happy, scared and disgusted.

With the combinations of AUs, in this case a bonus is given to the emotion of happiness and penalties are given to fear and disgust highlighting the difference. A bonus or a penalty is added to calculated measure with adapted prototypes. This bonus or penalty is obtained according to the combinations of AUs that occur on the face (Table 5).

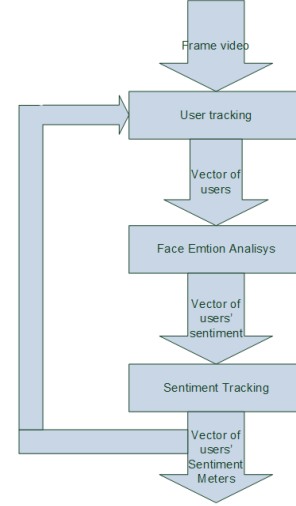
Affect	Original definition	Adapted Definition
Fear	1+2+4+5+20+25	$(1L+1R+2L+2R+20L+20R)/6$
Surprise	1+2+5+26	$(1L+1R+2L+2R+26)/5$
Anger	4+5+7+24	$(4L+4R+24L+24R)/4$
Sad	1+4+15	$(4L+4R+15L+15R)/4$
Disgust	4+9+10+17	$(4L+4R+10)/3$
Happy	6+12+25	$(12L+12R)/2$

**Table 4 Original definition and Adapted definition of the prototypes**

Combination	Bonus	Penalty
AU4L – AU4R	/	surprise -0.1 fear – 0.1 happy – 0.3
AU1L – AU1R – AU2L – AU2R	surprise – 0.2	disgust – 0.3 anger – 0.3 sadness – 0.2
AU24L – AU24R (low probability)	disgust – 0.2	/
AU24L – AU24R (high probability)	anger – 0.2	/
AU12L – AU12R – AU10 (low probability)	happy – 0.2	disgust – 0.1

#### F. Video analysis

The scheme of the Fig. 2 is able to perform the analysis on an image, but it can be easily extended to the affect detection in video. In this case also the user tracking and the affect tracking problems have to be considered. Video analysis module, shown in fig. 4, is designed to analyze off-line and real time video. It consists of three modules:



**Figure 4 Architecture of the Video Analysis module**

1. User tracking: it receives as input a new frame and the history of the user present in the scene. This history is stored in a vector of users: for each user the position at the last frame and the affect measures are stored. This module analyzes if there are any others in the scene, if they are new or were already present. The output is the position of the user in the scene. The problem of user tracking is resolved by the concept of the minimum Euclidean distance from the last position and current position of the users' face. Every detected user's face is compared to the vector of users' faces of previous frame to infer if such a face appeared in the previous frame. We make the assumption that the face can be subjected to fewer movements between two consecutive frames, and then two faces having a slightly different position in two consecutive frames can be the same. The faces without a previous position are considered as new users in the scene and, then, their location and size are stored in the vector of users. Not detected faces in the current frame, but detected in the previous frame, can be false positive or users that leave the scene. So, for each face detected in a frame, the tracking is obtained by the determination of the minimum among its Euclidean Distances with the other faces detected in the previous frame.
2. Face Emotion Analysis: it analyzes the faces of all users and calculate the probabilities of the six basic affects. In this step emotion detection process (Fig. 2) is applied on all users' face.



3. Affect Tracking: it calculates the measures of all affect of the users considering the changes of mood in the time. The problem of the affect tracking is resolved with an "affect meter". For each basic emotion there is an affect meter that shows the measure of affect found in the previous frame. This measure is represented as a real value  $A \in [0; 10]$ . This measure is increased if the calculated measure for the current single frame is greater than the average of the measures of the last five frames.

### III. EXPERIMENTAL RESULTS

The image analysis was carried out by implementing the approach presented in the previous section, so we created a module that receives an input images and returns as output the position of the faces in the frame and the relative affective states. Image analysis was tested using the dataset Extended Cohn-Kanade (CK +). Three tests were performed. The first tests the capabilities of the systems in identifying the AUs on the eyebrows and the mouth, the second one detects the basic sentiment (happiness, surprise, anger, disgust, fear and sadness). The third test identifies the status of the detected affects (positive or negative). In particular, the system recognizes as positive sentiments happiness and surprise, while as negative anger, disgust, fear and sadness. The results are shown in Table 6, 7 and 8. Tests show very satisfactory results compared to those presented in literature [56]. Analyzing the obtained results we can say that the sentiments having similar prototype exhibit lower performance and, according to the involved AUs, can be grouped into:

- anger, sadness and disgust with AU4;
- fear and sadness with AU1 and AU2;
- happy.

Happiness emotion has no AU in common with the other emotions and so the performances are better. According to this new grouping approach, better results are obtained by rerunning tests as shown in the table 9.

AU	Precision %	Recall %
1	81,81	96,11
2	92,94	88,76
4	82,65	84,37
10	90,00	69,23
12	69,33	66,67
15	73,08	46,34
20	75,00	64,28
24	75,61	59,61
26	81,18	66,35

Table 6 Test results on AUs.

Affect	Precision %	Recall %
Anger	51%	48%
Disgust	55%	56%
Fear	67%	95%
Happy	79%	75%
Sadness	59%	48%
Surprise	80%	87%

Table 7 Test results of image analysis on base sentiment (dataset CK+)

Affect	Precision %	Recall %
Negative	82%	87%
Positive	79%	76%

Table 8 Test results of image analysis on positive negative sentiment (dataset CK+).

Affect	Precision %	Recall %
anger/sadness/disgust	82	87
happy	79	76
surprise	79	76

Table 9 Test results on emotions' groups on CK+ dataset.

In the case of video analysis, we have performed two different tests on MMI, Cam3D and eINTERFACE'05 datasets. In the first test the system has to detect Ekman's sentiments (happy, surprise, anger, disgust, sadness, fear). In the second test, basic emotions are organized into three main categories:

- Positive emotions: happy, surprise;
- Negative emotions: anger, disgust, fear, sadness;
- Neutral emotion.

Results are shown in tables 10 and 11.

	MMI		eINTERFACE 05'		Cam3D	
	Pre %	Rec %	Pre %	Rec %	Pre %	Rec %
Happy	80,55	60,42	67,44	47,54	67,86	100,00
Fear	62,50	37,50	45,58	57,30	-	-
Anger	46,15	40,00	39,06	48,27	-	-
Disgust	40,00	53,33	43,25	84,54	75,00	60,00
Sadness	41,93	66,67	67,90	83,90	100,00	100,00
Surprise	57,14	95,23	42,32	55,48	60,00	50,00
Neutral	-	-	-	-	50,00	50,00
Average	54,71	58,86	50,93	62,84	70,57	72,00

Table 10 Test results on the MMI, eNterface 05' and Cam3D datasets on base sentiments

	MMI		eINTERFACE 05'		Cam3D	
	Pre %	Rec %	Pre %	Rec %	Pre %	Rec %
Positive	77,78	45,90	70,75	52,86	56,67	94,44
Negative	51,31	84,78	41,06	78,60	33,33	50,00
Neutral	-	-	-	-	88,00	56,41
Average	64,54	65,34	59,40	65,73	59,33	66,95

Table 11 Test results on the MMI, eNterface 05' and Cam3D datasets on positive and negative sentiments

As depicted in the previous tables, the proposed method shows good result in comparison with the other approaches. For a better characterization of the proposed method experimentation on a real dataset has been conducted. In particular, a new dataset, named UniSA dataset, has been built shooting the facial expressions of 12 users who watched three videos. Each video arouses certain affective status:

- the first video depicts funny sketches of animals and inspires in the users happiness and surprise sentiments (positive affective);

- the second video is a public service announcement on road safety and inspires sadness and anger sentiments(negative affective);
- the third video is a scene from a horror movie and inspires fear sentiments (negative affective).

At the end of the shooting users filled a questionnaire declaring their sentiments during the views.



Figure 5 Frames extracted from the dataset UniSA.

The obtained results are in tables 12.

	N° video	UniSA Dataset	
		Pre %	Rec %
Happy/Surprise	11	54,54	75,00
Sadness/Anger	9	66,67	54,64
Fear	9	33,33	100,00
Average	29	51,51	76,55

Table 12 – Base Sentiments Retrieval: Test Results

Also in this case the obtained results are interesting. The proposed approach suffers in the detection of Fear sentiment. The reasons are in the difficulties to discriminate this state from the Sadness/Anger and in the impulsive nature of this kind of sentiment.

#### IV. CONCLUSIONS

In this paper a novel approach to the detection and classification of a sentiment inside multimedia contents has been introduced. This technique has been based on the definition of a head tracking strategy and the relative extraction of points for the definition of the action units defined in the Ekman's model. The face detection and the extraction of the points of interest have been obtained by the use of image processing techniques that have been improved or adapted for our aims. The definition of the sentiment has been obtained by the use of the Ekman's theory. The proposed approach has been tested on the main standard datasets and the results are really interesting. The future works aim to apply the proposed approach on synchronous video and collect the sentiment of the user during the view of some contents or during some activities.

#### REFERENCES

- [1] Shen, L., Wang, M., & Shen, R. (2009). Affective e-Learning: Using "Emotional" Data to Improve Learning in Pervasive Learning Environment. *Educational Technology & Society*, 12 (2), 176–189.
- [2] Sun Duo, Lu Xue Song, An E-learning System based on Affective Computing, *Physics Procedia*, Volume 24, Part C, 2012, Pages 1893–1898
- [3] Zisook, M., Hernandez, J., Goodwin, M.S., and Picard, R. W., "Enabling Visual Exploration of Long-term Physiological Data," In *Proceedings of the 2013 IEEE Conference on Visual Analytics Science and Technology*, Atlanta, Georgia, USA, October 2013
- [4] McDuff, D., Kaliouby, R., Kodra, E., Picard, R. W., "Measuring Voter's Candidate Preference Based on Affective Responses to Election Debates", To appear in *The 5th biannual Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII 2013)*, 2-5 September 2013
- [5] A. Luneski, E. K., P.D. Bamidis, "Affective Medicine: a review of Affective Computing efforts in Medical Informatics", *MIM*, 49(3), 207–218, 2010
- [6] M.S. Bartlett, Gwen Littlewort, M. Frank, C. Lainscsek, Ian Fasel, and J. Movellan. Recognizing facial expression: machine learning and application to spontaneous behaviour. In *Computer Vision and Pattern Recognition*, 2005. *CVPR 2005*. IEEE
- [7] P. Ekman and W. Friesen. *Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions*, 2007
- [8] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, 1978.
- [9] Gwen Littlewort, M.S. Bartlett, Ian Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. In *Computer Vision and Pattern Recognition Workshop*, 2004. *CVPRW '04*. Conference on, 2004.
- [10] A. Verma and L.K. Sharma. A comprehensive survey on human facial expression detection. *International Journal of Image Processing*, 7(2), 2013.
- [11] Martin Wollmer, Felix Weninger, Tobias Knaup, Bjorn Schuller, Congkai Sun, Kenji Sagae, and Louis-Philippe Morency. Youtube movie reviews: Sentiment analysis in an audio-visual context. *IEEE Intelligent Systems*, 28(3):46-53, 2013.
- [12] L. Barsalou. Grounded cognition. *Ann. Rev. Psychogy*, 59, 2008.
- [13] Jesse Chandler and Norbert Schwarz. How extending your middle finger affects your perception of others: Learned movements inuence concept accessibility. *Journal of Experimental Social Psychology*, 45:123-128, 2009.
- [14] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39-58, 2009.
- [15] El Kaliouby, R. and Robinson, P. Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures. In *Proc. Int'l Conf. Computer Vision & Pattern Recognition*, 3, 2004.
- [16] Gu, H. and Ji, Q. An automated face reader for fatigue detection. In *Proc. Int'l Conf. Face & Gesture Recognition*, 111-116, 2004.
- [17] Bartlett, M.S., Littlewort, G., Frank, M.G., Lainscsek, C., Fasel, I. and Movellan, J. Fully automatic facial action recognition in spontaneous behavior. In *Proc. Conf. Automatic Face & Gesture Recognition*, 223-230, 2006.
- [18] Ya Chang, Changbo Hu, and Matthew Turk. Probabilistic expression analysis on manifolds. In *Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition*, *CVPR'04*, pages 520-527, Washington, DC, USA, 2004. IEEE Computer Society.
- [19] I. Shalif, "The Emotions and the Dimensions of Discrimination Among Them in Daily Life", Ph.D. Thesis, Psychology Dept., Bar-Ilan Univ., Ramat-Gan, Israel, 1991.
- [20] K. Schmidt and J. Cohn, "Dynamics of Facial Expression: Normative Characteristics and Individual Difference", *Intl. Conf. on Multimedia and Expo*, 2001.

- [21] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison Between Geometry-based and Gabor-waveletsbased Facial Expression Recognition Using Multi-layer Perceptron", Third IEEE Intel. Conf. On Automatic Face and Gesture Recognition, 1998.
- [22] I. Cohen, N. Sebe, A. Garg, L.S. Chen, and T.S. Huang, "Facial Expression Recognition From Video Sequences: Temporal and Static Modeling", Computer Vision and Image Understanding, 2003.
- [23] A. Elad, R. Kimmel, "On bending invariant signatures for surfaces", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, Issue: 10, Oct. 2003, pp. 1285 – 1295
- [24] K. Lee, J. Ho, M.H. Yang, and D. Kriegman, "Videobased Face Recognition Using Probabilistic Appearance Manifolds", Conference on Computer Vision and Pattern Recognition, 2003.
- [25] H. Deng, L. Jin, L. Zhen, J. Huang. A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA, International Journal of Information Technology Vol. 11 No. 11, 86-96 2005
- [26] G. Littlewort, M. Stewart Bartlett, I. Fasel, J. Susskind, J. Movellan. Dynamics of Facial Expression Extracted Automatically from Video. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04)
- [27] H. Wang, N. Ahuja. Facial Expression Decomposition. Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03), 2 003 IEEE
- [28] M. F. Valstar, M. Pantic, Z. Ambadar and J. F. Cohn. Spontaneous vs. Posed Facial Behavior: Automatic Analysis of Brow Actions. ICMi'06, November 2–4, 2006, Banff, Alberta, Canada.
- [29] J. F. Cohn, L. I. Reed, Z. Ambadar, J. Xiao and T. Moriyama. Automatic Analysis and Recognition of Brow Actions and Head Motion in Spontaneous Facial Behavior. 2004 IEEE International Conference on Systems, Man and Cybernetics.
- [30] J. Whitehill, C. W. Omlin. Haar Features for FACS AU Recognition. Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06), 2006
- [31] J. Garbas, T. Ruf, M. Unfried and A. Dieckmann. Towards Robust Real-time Valence Recognition from Facial Expressions for Market Research Applications. 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction.
- [32] J. F. Cohn, S. Lucey, J. Saragih, P. Lucey e F. De la Torre. Automated Facial Expression Recognition System.
- [33] R. El Kaliouby and P. Robinson. Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures, CVPRW '04 Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 10
- [34] M. F. Valstar and M. Pantic. Induced disgust, happiness and surprise: an addition to the MMI facial expression database. In Proceedings of Int'l Conf. Language Resources and Evaluation, Workshop on EMOTION, pages 65-70, Malta, May 2010.
- [35] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pages 94-101, 2010.
- [36] Lowe, David G., Object recognition from local scale-invariant features, The Proceedings of the Seventh IEEE International Conference on (Volume:2 ). Pages: 1150 - 1157 vol.2., 1999
- [37] Herbert Bay, Tinne Tuytelaars e Luc Van Gool, SURF: Speeded Up Robust Features, Computer Vision – ECCV 2006. Lecture Notes in Computer Science Volume 3951, 2006, pp 404-417.
- [38] Paul Viola, Michael Jones, Rapid Object Detection using a Boosted Cascade of Simple Features, Conference on Computer Vision and Pattern Recognition (CVPR), 2001, pp. 511-518.
- [39] Shengcai Liao, Xiangxin Zhu, Zhen Lei, Lun Zhang and Stan Z. Li, Learning Multi-scale Block Local Binary Patterns for Face Recognition, International Conference on Biometrics (ICB), 2007, pp. 828-837.
- [40] J. Gama and P. Brazdil, Cascade Generalization, Machine Learning December 2000, Volume 41, Issue 3, pp 315-343.
- [41] Robert E. Schapire, Yoram Singer, Improved Boosting Algorithms Using Confidence-rated Predictions Machine Learning. December 1999, Volume 37, Issue 3, pp 297-336.
- [42] M. N. Dailey S. Tongphub, N. Thongsak, Rapid detection of many object instances, Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science Volume 5807, 2009, pp 434-444
- [43] A. Schmidt and A. Kasinski, The Performance of the Haar Cascade Classifiers Applied to the Face and Eyes Detection, Computer Recognition Systems 2, Advances in Soft Computing Volume 45, 2007, pp 816-823
- [44] P. Dreuw, P. Steingrube, H. Hanselmann and H. Ney, SURF-Face: Face Recognition Under Viewpoint Consistency Constraints, BMVC 2009.
- [45] R. Rani, S. K. Grewal and K. Panwar, Object Recognition: Performance evaluation using SIFT and SURF, International Journal of Computer Applications (0975 – 8887) Volume 75 – No.3, August 2013.
- [46] N. Younus Khan, B. McCane and G. Wyvill, SIFT and SURF Performance Evaluation Against Various Image Deformations on Benchmark Dataset, Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference.
- [47] "The IMM Face Database - An Annotated Dataset of 240 Face Images", Michael M. Nordström, Mads Larsen, Janusz Sierakowski, Mikkel B. Stegmann. Informatics and Mathematical Modelling, Technical University of Denmark, DTU - 2004
- [48] CMU-MIT Frontal Face Dataset: [http://vasc.ri.cmu.edu/idb/html/face/frontal\\_images/](http://vasc.ri.cmu.edu/idb/html/face/frontal_images/)
- [49] Caltech Archive: <http://www.vision.caltech.edu/html-files/archive.html>
- [50] Caltech 10, 000 Web Faces [http://www.vision.caltech.edu/Image\\_Datasets/Caltech\\_10K\\_WebFaces/](http://www.vision.caltech.edu/Image_Datasets/Caltech_10K_WebFaces/)
- [51] Marwa Mahmoud, Tadas Baltrušaitis, Peter Robinson, Laurel Riek, 3D corpus of spontaneous complex mental states, Affective Computing & Intelligent Interaction 2011
- [52] PCSDK Documentation: [http://software.intel.com/sites/landingpage/perceptual\\_computing/documentation/html/](http://software.intel.com/sites/landingpage/perceptual_computing/documentation/html/)
- [53] OpenCV Documentation: <http://docs.opencv.org/>
- [54] GENKI FaceTracer database: [http://mplab.ucsd.edu/wordpress/?page\\_id=398](http://mplab.ucsd.edu/wordpress/?page_id=398)
- [55] O. Martin, I. Kotsia, B. Macq and I. Pitas : "The eNTERFACE'05 Audio-Visual Emotion Database" Proceedings of the First IEEE Workshop on Multimedia Database Management, Atlanta, April 2006.
- [56] Michel F. Valstar Timur R. Almaev. "Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition". In Humaine Association Conference on Affective Computing and Intelligent Interaction, 2013.
- [57] Colace, F., Foggia, P., Percannella, G., A probabilistic framework for TV-news stories detection and classification, IEEE International Conference on Multimedia and Expo, ICME 2005, 2005, pp. 1350-1353
- [58] Francesco Colace, Massimo De Santo, Luca Greco (2013). A probabilistic approach to Tweets' Sentiment Classification. In: Proceedings of 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction Ginevra 3-6 Settembre 2013 IEEE Vol.1, Pag.37-42

# Ranking Highlight Level of Movie Clips: A Template Based Adaptive Kernel SVM Method

Zheng Wang<sup>1</sup>, Gaojun Ren<sup>1</sup>, Meijun Sun<sup>2,\*</sup>, Jinchang Ren<sup>3</sup> and Jesse J. Jin<sup>1</sup>

<sup>1</sup> School of Computer Software, Tianjin University, Tianjin, China

<sup>2</sup> School of Computer Science and Technology, Tianjin University, Tianjin, China

<sup>3</sup> Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, United Kingdom  
wzheng@tju.edu.cn, gaojunren@tju.edu.cn, sunmeijun@tju.edu.cn, jinchang.ren@strath.ac.uk, jinsheng@tju.edu.cn

**Abstract**—This paper looks into a new direction in movie clips analysis –model based ranking of highlight level. A movie clip, containing a short story, is composed of several continuous shots, which is much simpler than the whole movie. As a result, clip based analysis provides a feasible way for movie analysis and interpretation. In this paper, clip-based ranking of highlight level is proposed, where the challenging problem in detecting and recognizing events within clips is not required. Due to the lack of publicly available datasets, we firstly construct a database of movie clips, where each clip is associated with manually derived highlight level as ground truth. From each clip a number of effective visual cues are then extracted. To bridge the gap between low-level features and highlight level semantics, a holistic method of highlight ranking model is introduced. According to the distance between testing clips and selected templates, appropriate kernel function of Support Vector Machine (SVM) is adaptively selected. Promising results are reported in automatic ranking of movie highlight levels.

**Keywords**—video analysis; highlight level; movie clip; template based method; adaptive kernel SVM

## I. INTRODUCTION

Nowadays, we can easily access to thousands of new video or movie resources from the Internet. To draw eyes of movie consumers, conventional film directors need to put great efforts to produce movie trailers by picking up the most vivid and representative highlighting clips of a movie. The whole process is very challenging and trivial and has raised a question as whether we can automatically determine and extract clips. This is the question we attempt to address in this paper.

Generally, the existing affective models for video highlights analysis can be summarized into two categories [2]: i.e. categorical affective content analysis and dimensional affective content analysis. In categorical affective content analysis, emotions commonly belong to a few pre-defined basic categories, such as "fear", "anger", "sad" or "surprise" [3]. Dimensional affective content analysis applies the well-known psychological Arousal-Valence emotion space (A-V space) proposed by Alan [4], which is characterized by the dimensions of arousal (intensity of affect) and valence (type of affect) providing a solid basis to represent the video affective content. The highlight level addressed in this paper is referred to the degree that draws audience's attention. It is not about labeling a video clip with one emotion type, nor about picking up the video highlights by choosing emotion spaces.

In general, considering the complicated structure and variations in rich content, related works in highlights extraction from generic movies remain rare. Due to the relative simple structure and clear semantics contained, highlights extraction from sports videos and music videos have been intensively investigated[24][25]. Existing work has mainly focused on event-based approaches, where modeling from features to events is required[25]. Consequently, the difficulty in extracting highlights has been converted to another challenging problem, i.e. event detection and recognition.

In order to identify the highlights contents from movies, classification tool is employed to distinguish affective contents from others. Support vector machine (SVM) [1], as an efficient classification tool, has been widely applied in many research fields. In David [8] several applications using SVMs in text categorization, computer vision, and bioinformatics are summarized. In addition, SVM is also applied in the fields of medical diagnosis [9], financial engineering [10], and information processing [11].

Most of the researches are mainly focused on single kernel function based SVM, which has inevitably limited its performance. In [7], it is found that the performance of SVM is greatly affected by the choice of a kernel function. To overcome this problem, mixing or combining multiple kernels by certain mathematical operators instead of using a single one for higher accuracy of SVMs is introduced [6]. Lu [12] proposes to optimize the combined kernel function by Particle Swarm Optimization (PSO) based on the large margin learning theory of SVM. However, there is few and sparse research investigating into incorporating with individual good performance of kernels of SVMs.

The novel contribution of this paper is the clip with user vote database and new conception of template based adaptive kernel SVM framework, which incorporates the good performance of individual kernels in SVMs without creating new combined kernels. The rest of this paper is organized as follows. Section II illustrates the overall framework of the proposed approach. Feature extraction and feature selection are presented in Section III and Section IV, respectively. Section V discusses how the novel template based adaptive kernel SVM works. The user survey experiment, data set establishment, and the designed experiments are described in Section VI. Finally, some concluding remarks are drawn in Section VII along with a brief discussion of possible future improvements.

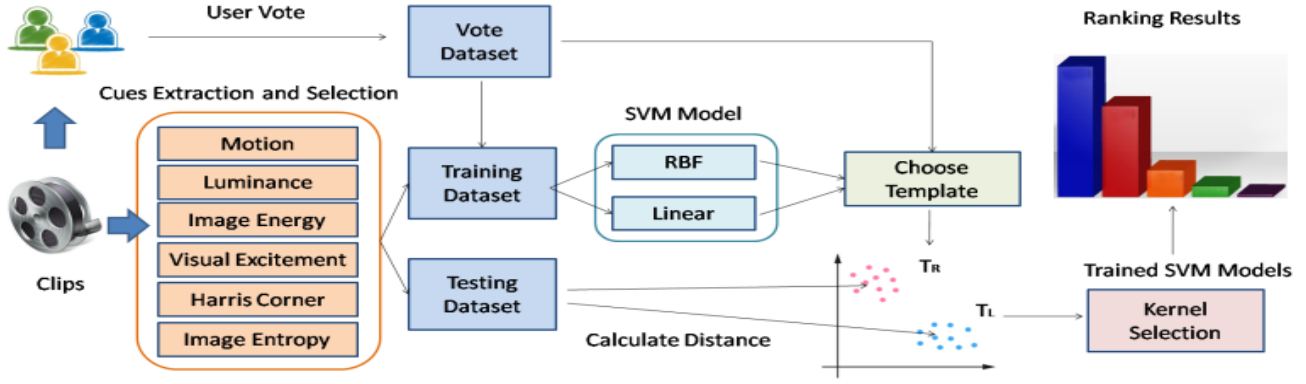


Figure 1. Framework of the proposed system.

## II. OVERALL METHODOLOGY

Figure. 1 illustrates the framework of proposed solution. With movie clips as input, the whole procedure is divided into three parts: clip with user vote database, visual cues extraction and template based adaptive kernel SVM classification.

### A. Clip With User Vote Database

Many movies are manually segmented into clips to ensure each clip contains a relative complete story in several continuous shots. The movie split tool [26] is shown in Figure. 2. Each movie clip has been voted to the audience's feeling whether the clip is exciting or not. The related votes are regarded as the ground truth to the associated movie clips.

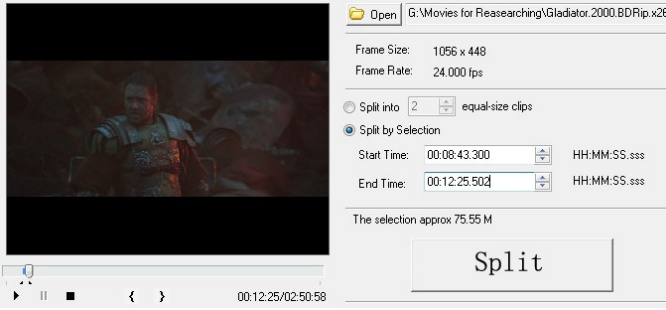


Figure 2. Movie split interface: example of a movie to be split according to the set up start and end time.

### B. Cues Extraction and Selection

The feature cues employed in this paper present the essential visual factors of movie clips and can be regarded as indicators between the highlight space and the low-level feature space. Audio events have strong hints to movie affective content and many works[3][13] have been done to analyze audio features. Although the results could be further improved with the audio analysis, this paper does not take audio features into account, but focus on the analysis of visual cues.

The size difference of clips results in different length of feature cues which makes it difficult to train the model. A time warping method [23] is adopted in this paper to overcome this problem; Finally the principal component analysis method is applied for dimensionality reduction.

### C. Template Based Adaptive Kernel SVMs

To evaluate the highlight levels of clips, instead of using single kernel or creating a new kernel this paper proposes a template based adaptive kernel SVMs by applying single RBF kernel and single linear kernel SVM on the dataset, and select top 10 best result clips respectively as templates.

## III. CUES EXTRACTION

As shown in Figure. 3, we extract six visual cues from movie clips and discuss in detail below.

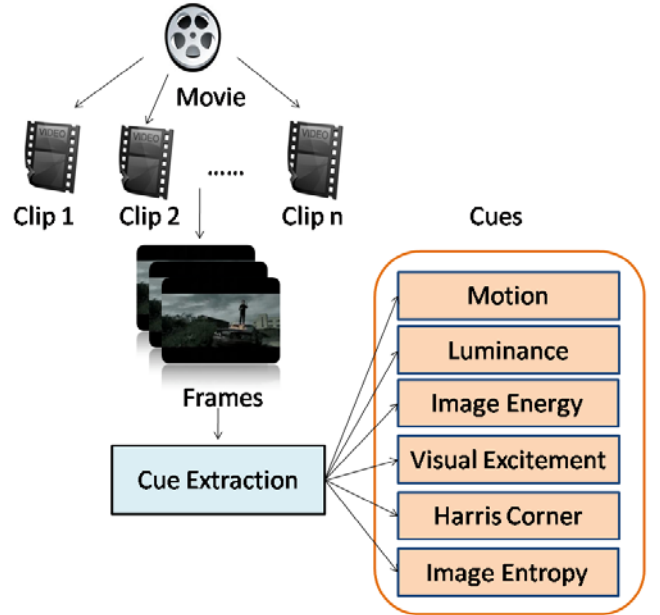


Figure 3. Cues extraction.

### A. Motion cue

Motions in a video include both the object motion and camera motion, which play important roles in highlights analysis. Figure. 4 shows an example that a clip which are more attractive to users appear to have high peak value and tremendous changed curves (the blue) than the other (the red). Normally these movie clips with remarkable motion features are more attractive to viewers.

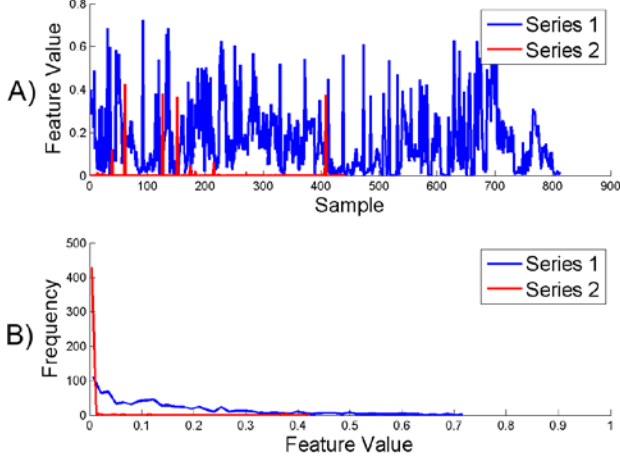


Figure 4. A) Motion features time curve of two clips with different number of user votes; B) Histogram of feature values

Let the motion feature at frame  $k$  be  $m(k)$  with the average magnitudes of  $N$  motion vectors in the frame with its width  $w$ . In the concept of optical flow, motion vectors are processed in 8 directions which are divided by  $2\pi$ .

$$m(k) = \sum_{i=1}^N \frac{|\vec{v}_i| \cdot |p_{ix} - q_{ix}|}{8 \cdot w \cdot N} \quad (1)$$

Where  $|\vec{v}_i|$  denote the length of the  $i$ th optical flow vector and  $p_{ix}, q_{ix}$  are  $x$  values of two points in  $\vec{v}_i$ .

#### B. Luminance cue

From the cinematographic perspective, lighting techniques play an extremely important role in movies. Generally two major aesthetic lighting techniques are frequently employed [16]: Low-key lighting and high-key lighting. Low-key lighting, or chiaroscuro lighting, characterized by a contrast between light and shadow areas, is quite usual in scenes like horror movies to driven atmosphere. On the contrary high-key lighting, or flat lighting, deemphasizing the light/dark contrast, is usually used to express the atmosphere of cheerful, warm or magnificent scenes.

Luminance cue simulates the lighting in cinematography well. For instance, the luminance curve waves left up and down when the low-key lighting is applied in a movie clip and the luminance curve waves smoothly when the high-key lighting is applied. As Figure. 5 shows, the clip with high level of highlight shows a curve (the blue) with high peak value and drastic changes. On the contrary, the curve of a low voted clip is smoother (the red).

Luminance feature  $l(k)$  at frame  $k$  is obtained in the hue, saturation and intensity (HSI) color space.

$$l(k) = \frac{h+s+i}{3} \quad (2)$$

The parameters  $h, s, i$  are the average value of hue, saturation and intensity.

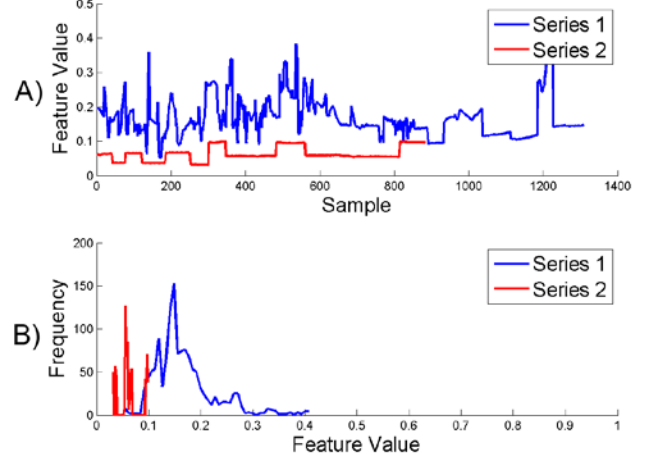


Figure 5. A) Luminance features time curve of two clips with different number of user votes; B) Histogram of feature values

#### C. Image Energy

Image energy describes the distribution of the gray levels in the image and texture coarseness, which in some way reflects the style of the video frame, like cartoon style with simple texture or realistically styles with complex texture. Movie clips are often more attractive to viewers than others, if they are full of beautiful scenery and inviting views and whose average image energy is lower than others. As verified in Figure. 6, the image energy of high voted movie clips (the blue) is lower than a less voted clip (the red). The image energy introduced in this paper is obtained by computing the sum of the squares of each element in gray-level co-occurrence matrix (GLCM) proposed by Haralick [18] in the 1970s.

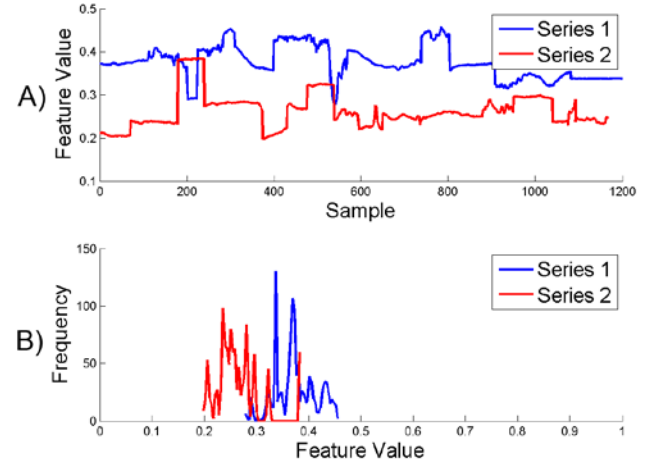


Figure 6. A) Energy features time curve of two clips with different number of user votes; B) Histogram of feature values

#### D. Visual Excitement

This feature is proposed by Wang [3] which reflecting the relation between the low-level feature and visual excitement. As a measure of visual excitement, this feature works for clips with different highlight levels.



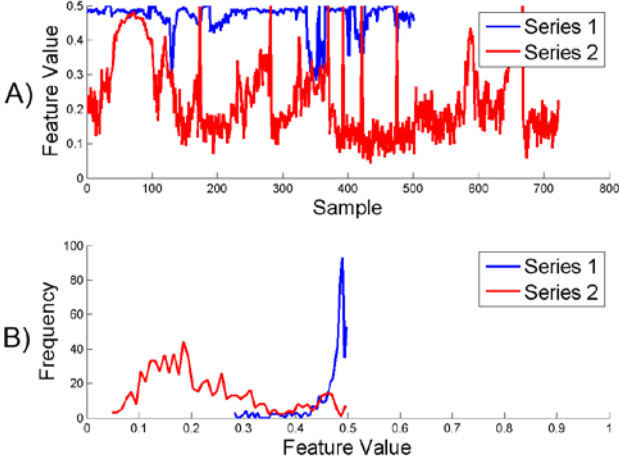


Figure 7. A) Visual Excitement features time curve of two clips with different number of user votes; B) Histogram of feature values

Figure. 7 shows that the movie clip with higher votes have higher visual excitement value in average (the blue) than those with lower votes (the red), which can be concluded both from the time series curves and the histogram.

#### E. Harris Corner

Normally, movie clips with rich content can be regarded as a scene with more interest points than others; these clips usually attract the eyes of the viewer's more easily. The Harris Corner detection [19] is suitable for such a case. Figure. 8 shows that clips with higher votes have an average higher number of interest points (the blue) than those with lower votes (the red).

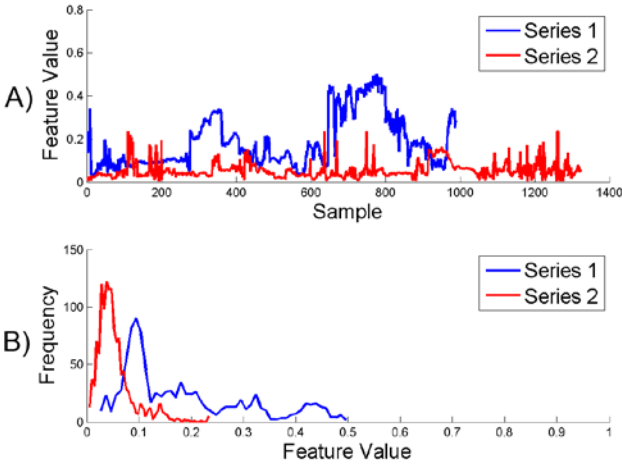


Figure 8. A) Harris Corner features time curve of two clips with different number of user votes; B) Histogram of feature values

#### F. Image Entropy

Two dimensional image entropy is a measurement of the disorder or randomness in video frame [17]. It describes the spatial characteristics of the image gray. Movie clips with some abnormal activities attract users' attention than those with normal activities as shown in Figure. 9 that the movie clips

with high votes have higher image entropy (the blue). 2D image entropy  $e(k)$  at frame  $k$  with width  $w$  and height  $h$  can be described by

$$e(k) = - \sum_{i=0}^{255} \sum_{j=0}^{255} P_{ij} \ln P_{ij} \quad (3)$$

$$P_{ij} = \frac{f(i,j)}{wh} \quad (4)$$

The  $f(i,j)$  is the number of pairs denoted by  $(i,j)$ , which represents the pixel's gray value and its neighborhood gray value are  $i,j$  respectively.

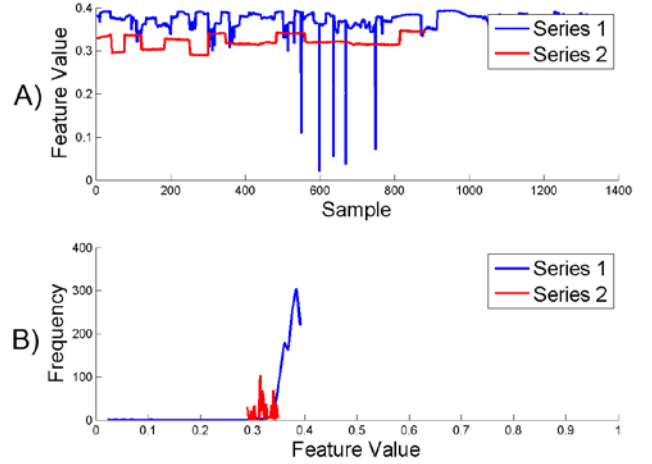


Figure 9. A) Image entropy features time curve of two clips with different number of user votes; B) Histogram of feature values

### IV. FEATURE SELECTION

This section elaborates the feature selection stage of the proposed framework. The purpose of this stage is dimensionally reduction and feature length aligning.

Given an input movie clip  $V$ , we denote each of these features by  $f_i$  and the feature set by  $F$  below, where  $k$  is the number of feature cues and equals six in this paper.

$$F = \{f_i; i = 1, 2, \dots, k\} \quad (5)$$

Due to different numbers of frames contained in these clips, the size of the raw feature  $f_i$  acquired directly from a movie clip differs from each other. This requires the features to be resized into the same length. The feature selection method is formulated in formula (6)

$$F' = \text{PCA}(T(f_1), T(f_2), \dots, T(f_k)) \quad (6)$$

The  $F'$  is the final integrated feature vector after selection, the function  $\text{PCA}()$  [21] and  $T()$  represent the principal components analysis and time warping method respectively. This proposed method converts different size of feature series to a given size and reduces data redundancy.

Figure. 10 shows how the time warping method with hyperbolic tangent curve works. For instance, consider a 25s clip with 1000 frames, mapping the 1000 dimensions feature to the  $x$  axis. The 1000 dimensions range from -500 to +500 on the  $x$  axis will be mapped to a result ranges from -25 to +25 on the  $y$  axis. This mapping keeps the details of the middle part dimensions and reduces the noised of both ends at the same

time, which is suitable for movie clips since the middle part of a movie clip is more likely to have the effective content.

#### Cues with Different Dimensions

#### Cues with Same Dimension

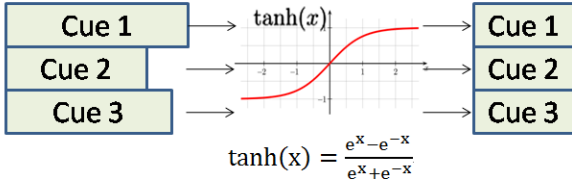


Figure 10. Time warping method with the hyperbolic tangent curve.

Each feature time series  $f_i$  is warped to the size of 50, and then we obtain a total 300 dimension integrated feature of every clip. Finally a PCA process is applied to get a 100 dimensions integrated feature for the following stage.

#### V. TEMPLATE BASED ADAPTIVE KERNEL SVM

In the database, we have obtained the votes  $v$  for each video clip, which stands for the evaluation value of affection, and the relationship between the low-level feature and the affection value is given as follows:

$$W = \langle F', v \rangle \quad (7)$$

Where  $F'$  is the integrated feature from Equation 6. With the relationship  $W$ , we propose SVM based framework which adaptively select RBF or Linear kernel based on the comparison results between testing clip and templates, as shown in Figure. 11.

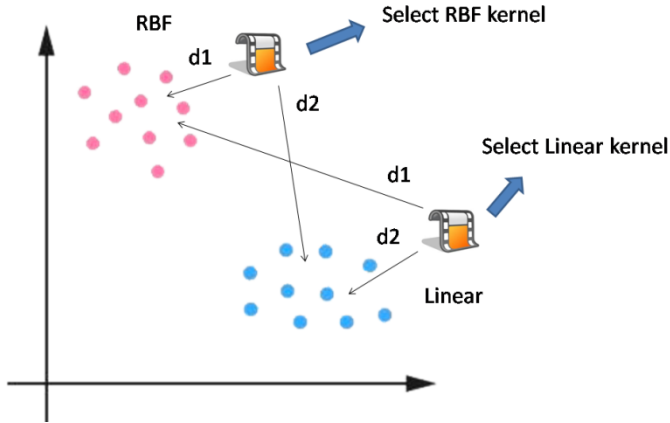


Figure 11. Template based adaptive kernel SVM

The whole method can be described as following steps:

- 1) *Papre training and testing dataset from  $W$*
- 2) *Training two SVM model with Linear and RBF kernel, denoted as  $M_L$  and  $M_R$*
- 3) *Based on the testing results, select top 10 clips' features set for each model and take them as templates sets, denoted as  $T_L$  and  $T_R$*
- 4) *For a testing clip feature  $F'$ , calculate the distance  $d_L$  and  $d_R$  to the two templates sets respectively. Then  $d_L =$*

*$Distance(F', T_L)$  and  $d_R = Distance(F', T_R)$ , where the  $Distance()$  function is the Hamming distance method[22].*

- 5) *If  $(d_L > d_R)$  Select  $M_L$  as predicting model for the clip. else select  $M_R$*

#### VI. EXPERIMENTS

This section introduces some experiments to evaluate the proposed framework and show the advantage of the template based adaptive kernel SVM method. For the convenience of comparative assessment, a user experiment was conducted to establish the ground truth which each movie clip is labeled with user votes as the measurement of highlight level.

##### A. Proposed database

The user experiment is set up as follows. The audiences are 70 undergraduate students with 30 females and 40 males who have received detailed instructions of how the experiment is conducted. During the experiment, each user is asked to watch these 350 movie clips on a computer screen and vote for those clips whether he/she enjoys it. As a result, votes of movie clips range from 0 (nobody votes for it) to 70 (everybody votes for it).

The clips mentioned above are manually chosen from a total of 20 movies listed in Table 1 with four major movie genres including action, horror, war and disaster. Each genre contains 5-6 movies and there are about 15-18 clips are chosen from each movie. The movie clips varies from a minimum of 40 seconds to a maximum of 2 minutes in length. Figure. 12 shows the key frames of these movies.



Figure 12. Key-frames from the testing database

TABLE I. MOVIES IN THE DATABASE

No.	Movie Names	Genres
1	Red (2010)	Action/Comedy
2	Mission: Impossible III(2006)	Action/Thriller
3	Live Free or Die Hard (2007)	Action/Thriller
4	Gladiator (2000)	Action/Drama
5	Terminator 2: Judgment Day (1991)	Action/Sci-Fi
6	The Day After Tomorrow (2004)	Disaster/Sci-Fi
7	Twister (1996)	Disaster/Drama
8	Titanic (1997)	Disaster/Romance
9	The Perfect Storm (2000)	Disaster/Drama
10	2012 (2009)	Disaster/Family
11	The Silence of the Lambs (1991)	Horror/Crime
12	Silent Hill (2006)	Horror/Thriller

13	The Shining(1997)	Horror
14	Final Destination (2000)	Horror/Thriller
15	Alien: Resurrection (1997)	Horror/Sci-Fi
16	Black Hawk Down (2001)	War/History
17	The Lord of the Rings: The Return of the King (2003)	War/Action
18	Troy (2004)	War/Romance
19	Brave heart (1995)	War/History
20	Saving Private Ryan (1998)	War/History

The criterion of choosing clip is selecting those clips whose highlight level rang from low to high averagely. Finally, there are 596 clips. Figure. 13 shows the sorted user vote's histogram, it can be observed that the database contains clips with different highlight levels. The clip database will be published very soon.

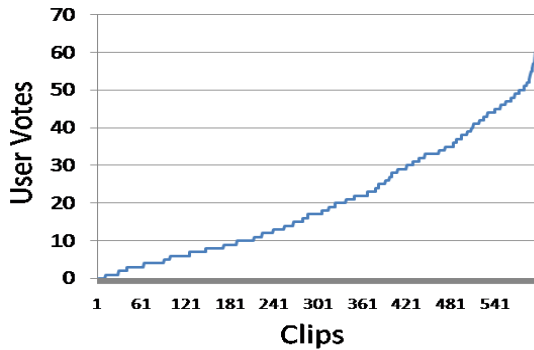


Figure 13. Statistic histogram of user votes which have been sorted.

### B. Experiments on the proposed database

To show the advantage of the proposed method in this paper, experiments with single kernel SVMs and proposed method are compared.

To train the single kernel SVM, 426 clips are selected randomly from database, half of them are training data and others are testing data. According to the results, the top 10 best result clips are selected as templates. Then the other 170 clips are tested by using the RBF kernel SVM, linear kernel SVM and the proposed method. Table 2 shows the statistic error results of linear kernel SVM, RBF kernel SVM and proposed method. The average error between predicted value and the ground truth is employed as evaluation criterion.

TABLE II. COMPARISON RESULTS OF THREE METHODS

Error Range	Linear		RBF		Proposed	
	Clips Amount	Average Error	Clips Amount	Average Error	Clips Amount	Average Error
(0, 10]	80	4.509	86	5.033	114	4.365
(10,20]	49	15.137	52	14.559	38	14.756
(20,30]	27	24.537	30	23.165	17	22.923
(30,40]	9	33.310	1	31.371	0	--
(40,50]	2	41.472	1	41.378	1	40.014
(50,60]	2	55.796	0	--	0	--
(60,70]	0	--	0	--	0	--
(70,80]	1	86.172	0	--	0	--
Total	170	13.797	170	11.514	170	8.753

As shown in Table II, it is obvious to see that proposed performs better than linear kernel and RBF kernel SVM in

most cases. When the error value lies in the ranges from 0 to 20, the results shows that the proposed method result contains more clips than the other two methods, which indicates the higher performance of proposed method.

A snapshot of the results comparison between the ground truth and the proposed method is shown in Figure 14. The result of the proposed method is very close to the ground truth.

### C. Other Case Study

To demonstrate the framework proposed in this paper, half part (about 55 minutes in length) of the movie Wanted (2008) with a total of 110 minutes in length is recruited as an example of movie highlights summarization.

TABLE III. MOVIE TRAILER WITH 53s LENGTH CONTAINING 10 SHOTS GENERATED FOR THE MOVIE WANTED (2008)

No.	Snap Shot	Description
1		A Gun is flying from a man towards the hero.
2		The hero is running after a man in the train station.
3		The hero is trying to calm down the scared passengers. The scene is chaotic as people shout loudly.
4		A train is running through a gorge and suddenly the train stops and falls down from pathway.
5		The hero is seeking for his enemy in a train with a gun in his hand. Passengers are scared.
6		The hero is firing around with a man's body as shield. Many enemies die.
7		A train is running through a city with beautiful night urban landscape.
8		A man is driving a car trying to catch up the train, shooting at the train at the time.
9		Someone gets shot and falls into the lake, bleeding.
10		The hero gets back to the base for a final battle.

First of all, the selected movie is sliced into clips and re-connected by the shot boundary detection method[20]. Those movie clips whose length are less than 3 seconds are neglected automatically. Second, the mentioned integrated feature cues are extracted from each of the clips. Finally, for each feature vector, the predicted result is obtained with the template based

adaptive kernel SVM model. The top 10 clips with highest predicted results make up the final movie highlights summarization, as shown in Table III.

## VII. CONCLUSIONS AND FUTURE WORK

In this paper, we built a clip database with user votes for their highlight level, and then a novel template based adaptive kernel SVM framework is introduced to ranking movie clip highlight level by using six visual feature cues. These features are combined together via time warping and PCA to represent the characteristics of a given movie clip. Experimental results have demonstrated that the proposed framework is effective in ranking movie clips and producing movie highlight recommendation.

Further studies will be conducted to enlarge the proposed clip database and improve the performance of proposed algorithm by using deep learning method.

## VIII. ACKNOWLEDGEMENT

The authors wish to acknowledge the support from the National Natural Science Foundation, China, under the grant 61003201 and 61202165, and a joint project funded by Royal Society of Edinburgh and NSFC 61211130125.

## IX. REFERENCES

- [1] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, B. Scholkopf. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Networks*, 12 (2) (2001), pp. 181–202. 2001.
- [2] S. Zhang, Q. Huang, S. Jiang, W. Gao, Q. Tian. Affective visualization and retrieval for music video. *IEEE Trans. Multimedia*. Vol. 12, No. 6, pp.510-522. October 2010.
- [3] H. L. Wang, L. F. Cheong. Affective understanding in film. *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 6, pp. 689-704. Jun, 2006.
- [4] A. Hanjalic, L. Q. Xu. Affective video content representation and modeling. *IEEE Trans. Multimedia*, vol. 7, no. 1. Pp. 143-154. Feb, 2005
- [5] O. Chapelle, V.N. Vapnik. Choosing multiple parameters for support vector machines. *Machine Learn.*, 46 (1–3) (2002), pp. 131–159. 2002.
- [6] Mehmet Gönen, EthemAlpaydin. Multiple Kernel Learning Algorithms. *Journal of Machine Learning Research* 12 (2011) 2211-2268. 2011
- [7] M. Hussain, S.K. Wajid, A. Elzaat, M. Berbar. A Comparison of SVM Kernel Functions for Breast Cancer Detection. 2011 Eighth International Conference on Computer Graphics, Imaging and Visualization (CGIV), pp.145-150. 2011.
- [8] V. David, A. Sánchez. Advanced support vector machines and kernel methods. *Neurocomputing*, 55 (1–2) (2003), pp. 5–20. 2003.
- [9] D. Conforti, R. Guido. Kernel based support vector machine via semidefinite programming: application to medical diagnosis. *Computers & Operations Research*, 37 (2010), pp. 1389–1394. 2010.
- [10] Van Gestel et al. Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Trans. Neural Networks*, 12 (4) (2001), pp. 809–821 Special Issue on Neural Networks in Financial Engineering. 2001.
- [11] Vapnik et al.. Support vector method for function approximation, regression estimation, and signal processing. *The Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA (1997), pp. 281–287. 1997.
- [12] Ming-Zhu Lu, Chen, C.L.P. Jian-Bing Huo. Optimization of combined kernel function for SVM by Particle Swarm Optimization. 2009 International Conference on Machine Learning and Cybernetics. 12-15 July 2009.
- [13] M. Xu, L.-T. Chia, J. Jin. Affective content analysis in comedy and horror videos by audio emotional event detection. *IEEE international conference on multimedia and expo*, vol. 61, no. 2, pp.2-5. July, 2005.
- [14] A. Ekin, A. M. Tekalp, R. Mehrotra. Automatic soccer video analysis and summarization. *IEEE Trans. on image processing*, vol. 12, no. 7, pp. 796-807. July 2003.
- [15] J.W. Hsieh, S.L. Yu, Y.S. Chen. Motion-based video retrieval by trajectory matching. *IEEE Trans. Circuits and Systems for Video Technology*, 16 (3) (2006), pp. 396–409. 2006.
- [16] D. Bordwell and K. Thompson. *Film art: An Introduction*, 7th ed. new York: McGraw-hill, 2004.
- [17] Md. Haidar Sharif, ChabaneDjeraba. An entropy approach for abnormal activities detection in video streams. *Pattern Recognition*, Volume 45, Issue 7, Pages 2543-2561. July 2012.
- [18] R. M. Haralick, K. Shanmugam, and I. Dinstein, Textural Features for Image Classification, *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. SMC-3, No.6, pp.610-621, November 1973.
- [19] Li Yi-bo, Li Jun-jun. Harris Corner Detection Algorithm Based on Improved Contourlet Transform. *Procedia Engineering*, Volume 15, 2011, Pages 2239-2243. 2011.
- [20] Jinhui Yuan, Huiyi Wang, et al. A Formal Study of Shot Boundary Detection. *Circuits and Systems for Video Technology*, IEEE Transactions on. Volume: 17, Issue: 2, Page(s): 168 – 186. Feb. 2007.
- [21] Sahouria, E; Zakhori, A. Content analysis of video using principal components. *Circuits and Systems for Video Technology*, IEEE Transactions on. Volume: 9, Issue: 8, pp. 1290-1298. DEC 1999
- [22] Lei Zhang, Yongdong Zhang, et al. Binary Code Ranking with Weighted Hamming Distance. 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1586-1593. June 2013.
- [23] Chan, F.K.-P. Haar wavelets for efficient similarity search of time-series: with and without time warping. *Knowledge and Data Engineering*, IEEE Transactions on. Volume:15, Issue: 3. pp. 686-705. May, 2003.
- [24] Shiliang Zhang, Qingming Huang, Shuqiang Jiang, Wen Gao, and Qi Tian, "Affective Visualization and Retrieval for music video," *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 510-522, Oct 2010.
- [25] Guangyu Zhu, Changsheng Xu, "Event Tactic Analysis Based on Broadcast Sports Video," *IEEE Transactions on Multimedia*, pp. 49-67, 2009.
- [26] <http://www.boilsoft.com>

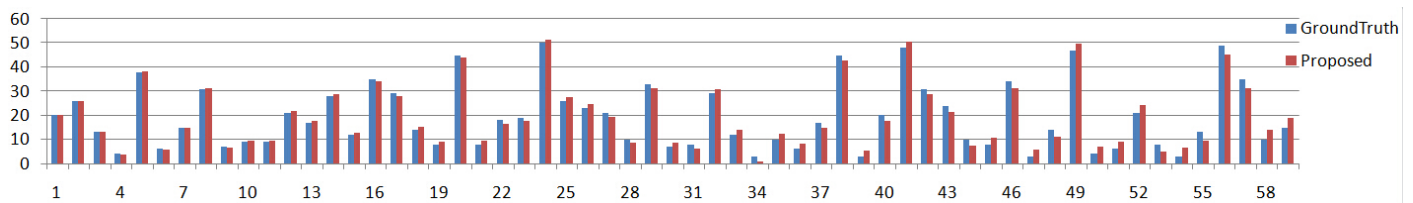


Figure 14. Result comparison between ground truth and the proposed method.



# Relevance Measures for the Creation of Groups in an Annotation System

Danilo Avola, Paolo Bottoni, Amjad Hawash  
Department of Computer Science  
Sapienza University of Rome  
Via Salaria 113, 00198, Rome, Italy  
(avola, bottoni, hawash)@di.uniroma1.it

## Abstract

*The MADCOW annotation system supports a notion of group, facilitating focused annotations with respect to a domain. In previous work, we adopted ontologies to represent knowledge about domains, thus allowing more refined annotations to a group, and discussed how the use of ontologies facilitates the formulation of semantically significant queries for retrieving annotations on specific topics. We now expand on previous results and study two new types of measures to identify matches between users' interests and groups: **Degree Centrality**, developed for social networks to assess the quality of concepts in an ontology, and **URL concordance**, indicating the similarity of interests among users who annotate the same pages.*

**Keywords:** Web annotation, Matching, Class Match Measure, Degree Centrality Measure.

## 1. Introduction

Collaborative activities require that some common terminology be established among agents possibly coming from different backgrounds and focusing on different, though related, tasks. In general, even in cross-disciplinary endeavours, collaboration among users is focused on some domain for which specialised terminologies may have been defined. In the last 15 years, ontologies have become the modeling technique of choice for characterising the concepts relevant to a domain, so that every project requires, albeit implicitly, agreement on the content of some *domain* [13], or even *local* [10], ontology.

From a different perspective, and notably based on technologies relative to the so-called Web 2.0, forms of addition of personalised content to publicly available documents have become ubiquitous, ranging from simple tags to on-the-fly linking of existing material, to the production of new texts or sketches, while allowing the automatic capture of contextual information. In particular, the MADCOW sys-

tem [4] provides an integrated access to all such techniques, and to the results of annotation processes, in the form either of local popups showing existing annotations or of dynamically generated pages incorporating all the added material.

Annotations produced by a user while perusing a document, or embedded in the document itself, as well as the documents generated from their integration, can then become the basis for collaborative activities, on which other users can integrate information, question or revise existing annotations, and establish links with other information sources [3]. Such a progressive addition of new information differs from collaborative construction of documents, as it allows other users to read the development of any ongoing discussion, rather than only its final result, while all the time preserving the integrity of the original document, which might be owned by third parties.

In this open setting, clusters of annotations (and authors) on specific topics emerge; at the same time, repeated collaborative activities lead to the formation of groups of annotators with common interests, or who cooperate to realise some tasks in some specific domain. However, some problems arise in the creation of groups, if one needs to find fellow annotators or to look for groups relevant to his or her interests. Indeed, manual investigation of the topics of candidate groups or of existing annotations becomes rapidly unwieldy as the number of groups or of annotators increases.

In [1, 2, 14] we proposed automatic groups-users matching as an alternative to manual search of groups to join (or of users to invite in a group): Group owners (users) are now presented with ranked lists of relevant users (groups). Experimental tests showed that automatic suggestions significantly reduced the time spent in looking for proper choices [2]. At the basis of the matching process is the association of groups with ontologies whose concepts (used as terms) are representative of the group objectives [6]. Terms can be recognised as significant by users, and searched for and manipulated by suitable services [7, 12, 24, 26], while annotations can be endowed with a set of tags to represent the intention behind them.

In this paper, we improve the groups-users matching process by integrating it with information about the structure of the ontology itself and by considering the actual contents on which the annotation was performed. In particular, we overcome some limitations of the Class Match Measure (**CMM**) from previous works [1, 2], which, although providing meaningful results, depends on a single term-ontology matching, and does not consider the relations among the ontology concepts that match the terms provided by the user. Hence, we turn to social networks analysis where centrality measures consider the roles played in the network topology by a given node or set of nodes. Graph Degree Centrality (**DC**) [11, 8] is used here, in a slightly modified version, so that the system computes the centrality of the maximum subgraph that could be formed from the provided terms against a given ontology.

We also propose a measure based on the annotated URLs, following the observation that different users annotating the same specialised Web sites usually share the same interests. Discovering that fellow annotators already belong in some group could encourage others to become members. On the other hand, group owners can look for users who systematically annotate websites which are targets of annotations for many group members.

We report on a pilot test comparing the use of **CMM** and **DC** to assess whether consideration of an ontology topology improves the quality of matching. We also compare the ontology-based and URL-based matching processes.

After considering related work in Section 2, we introduce relevance measures and discuss their use in Section 3, also sketching the storage of ontologies. Section 4 presents an applicative scenario, with experimental results discussed in Section 5, and Section 6 concludes the paper.

## 2. Related work

Ontologies were proposed by Paralic and Kostjal as representational schemes for domain knowledge to enhance the retrieval process, in comparison with the vector and the latent semantic indexing models [21]. In [5] a retrieval agent is described, providing access to information from multiple domains based on domain ontologies and users' interests.

Both [23] and [9] apply similarity measures and ontologies to match job-seekers and job offers based on descriptions of the skills respectively owned and required.

Sentence similarity based on semantic nets and corpus statistics is treated in [17], with reference to very short – one sentence long – texts, and taking into account implied semantics and word order information. In [22], similarity relevance is extended to include the relations between completely different resources. They propose a methodology to measure semantic relevance between resources based on ontological representations, taking into account different

meaningful relations.

Centrality measures have been used to assess the relevance of an ontology to a collection of terms. In [20], group centrality is used to aggregate journals into disciplines in a co-citation network. In [27] a topological centrality measure is proposed, and compared with other centrality measures, to discover communities in complex networks and to construct the backbone network.

In [15], the benefits of applying Social Network Analysis to ontologies and the Semantic Web are illustrated. They discuss how different notions of centrality describe the core content and structure of an ontology, and illustrate the insights provided by centrality measures such as degree, betweenness and eigenvector on two ontologies (SWRC<sup>1</sup> and SUMO<sup>2</sup>) which are different in purpose, scope and size.

In his work, Freeman [11] defined **DC** for a whole graph by considering the in-degree and out-degree for each node in the graph. We are using the same definition here, but with an extension to consider two types of node relation: Part-of and Similarity. For terms not involved in the subgraph, we use the simple definition of **DC**. We here adopt **DC** to assess the relevance of existing domains (represented by sets of terms) for a given group. The same measure is also used to match domains with users (represented by their annotations) in order to recommend groups to users and users to groups. This work represents a continuation of the works presented in [1, 14, 2] where ontology-based groups-users matching was introduced and its mathematical basis given.

## 3. Relevance Measures

We introduce some definitions preliminary to the description of the role of ontologies for MADCOW groups: (1) **Domain**: a unique name designating the area of knowledge to which an ontology refers. (2) **Concept**: the name associated with a node. (3) **Lexemes**: lexicalisations of some concept. (4) **Terms**: lexicalisations provided by a group owner to characterise the intent of group creation. (5) **Tags**: lexicalisations provided by a user to characterise an annotation content. After describing the creation of the ontology repository used in our experiments, we illustrate the use and implementation of two independent relevance measures for proposing group-user associations.

### 3.1. Ontologies Repository

A repository of 40 ontologies was created extracting data from the BabelNet Ontology<sup>3</sup>. BabelNet is a multilingual

<sup>1</sup><http://ontobroker.semanticweb.org/ontologies/swrc-onto-2001-12-11.owl>

<sup>2</sup><http://www.ontologyportal.org/>

<sup>3</sup><http://babelnet.org>



encyclopedia dictionary, with lexicographic and encyclopedic coverage of terms, and a semantic network connecting concepts and named entities in a very large (more than 9 million entries) network of semantic relations, called Babel synsets. Each synset represents a concept and contains all the synonyms which express that concept in a range of different languages. Synsets are obtained from the automatic integration of WordNet and Wikipedia [18] and a JAVA API helps in the data extraction process [19].

A Java application for ontology extraction was developed, which takes a domain (which will provide the ontology name) as argument, starts identifying the top BabelNet synset for that domain and then visits the related synsets recursively, increasingly adding concepts to the ontology under construction. In the current implementation, we stop recursion at level 3<sup>4</sup>. For every visited synset, its ID, URL, lexemes, MAP (the set of related synsets), and the types of relationships with these synsets are extracted and saved in MySQL tables. In the experimental tests in our previous works, 6 ontologies generated by OntoLearn [25] were used, structured as trees with concepts as nodes and edges labelled by the IS-A relationship. Here, we build ontologies as directed graphs with 4 different concepts relationships: *Subclass*, *Superclass*, *Part-of* and *Similarity*. Fig. 1 shows a fragment of an Entity-Relation diagram for the ontology repository, while the following schematises the JAVA code used to create the repository. For simplicity, we do not show here the structural features of the *Ontology* class corresponding to those shown in Fig. 1.

```
Ontology ontology = new Ontology(name);
Concept concept = createConcept(name, new BabConn());
ontology.addConcept(concept);
ontology.buildOnto(concept.getLinkedConcepts(), 3);
saveInDataBase(ontology);

Concept createConcept(String title, BabConn babelNet){
    ID = babelNet.getSynsetID(title);
    URL = babelNet.getSynsetURL(ID);
    lexemes = babelNet.getLexemes(ID);
    linkedConcepts = babelNet.getLinkedConcepts(ID); }

class Ontology{
    ....
    public buildOnto(List<Concept> linkedCpts, int level){
        if(level==0) return;
        else {
            List<Concepts> leaves = new ArrayList();
            for(Iterator it = linkedCpts.iterator(); it.hasNext()){
                Concept concept = (Concept)(it.next());
                this.addConcept(concept);
                leaves.add(concept.getLinkedConcepts());
            }
            this.buildOnto(leaves, --level); } }
```

<sup>4</sup>Matching computation time increases with the existence of more ontologies and more levels within each ontology. However, executing the matching itself is not a frequent process. Future work will include a study for minimizing matching computation time.

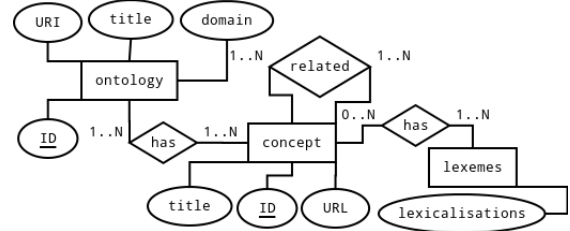


Figure 1: A fragment of the Repository Schema.

The construction of instances of *Concept* exploits a *babelNet* object to provide access to BabelNet, retrieving all the data related to a concept, according to the schema in Fig. 1. After a new ontology is created and an initial concept added, it gets populated by recursively following the concepts linked to the concepts in BabelNet corresponding to those already included in the ontology. The first invocation of *buildOnto* has as argument the collection of concepts linked to the first one. The recursive invocations follow the links up to the indicated level of depth, here fixed to 3. The member function *getLinkedConcepts* retrieves the concepts (if found) having relations: *Subclass*, *Superclass*, *Part-of* and *Similarity* with the concept being processed. Finally, *saveInDataBase* saves the whole ontology in the database according to the relations and properties present in Fig. 1.

### 3.2. Ontology-based measures

We start by describing the two distinct measures used in MADCOW: Class Match and Degree Centrality.

#### 3.2.1 Class Match Measure

The Class Match Measure **CMM** (exact and partial) evaluates the coverage of an ontology for a given set of lexemes.

**Definition 1** (Class Match Measure). *Let  $C(O)$  be the set of concepts for an ontology  $O$ ,  $L_O(c)$  the set of lexemes for a concept  $c \in C(O)$ , and  $T$  a set of lexical items. Let  $token : Lex \times Lex \rightarrow \mathbb{B}$  be a function such that  $token(x, y)$  is *true* iff  $y$  is a subword of  $x$  and *false* otherwise, where  $Lex$  is the set of all possible lexicalisations. Then a Class Match Measure for  $O$  and  $T$  is a function  $CMM(O, T) = \alpha E(O, T) + \beta P(O, T)$ , with  $\alpha \geq \beta$ ,  $\alpha + \beta = 1$ , and<sup>5</sup>:*

$$I(c, t) = \begin{cases} 1 & \text{if } \exists l \in L_O(c)[l = t] \\ 0 & \text{otherwise} \end{cases}$$

$$E(O, T) = \sum_{c \in C(O)} \sum_{t \in T} I(c, t)$$

$$J(c, t) = \begin{cases} 1 & \text{if } \exists l \in L_O(c)[token(l, t)] \\ 0 & \text{otherwise} \end{cases}$$

$$P(O, T) = \sum_{c \in C(O)} \sum_{t \in T} J(c, t)$$

<sup>5</sup>We have empirically determined  $\alpha = 0.6, \beta = 0.4$

$E(O, T)$  and  $P(O, T)$  are called *exact* and *partial matches* between  $O$  and  $T$ , respectively.

### 3.2.2 Degree Centrality Measure

According to [16, 20, 27], the Degree Centrality Measure (DC) quantifies the importance of a concept in an ontology with respect to its number of connections, viewing the ontology as a directed graph. We adapt this definition to provide a measure of the relevance of an ontology for a group, based on the notion of maximal connected subgraph formed by ontology concepts/lexemes matched to group terms<sup>6</sup>. Fig. 2 depicts an ontology as a directed graph, highlighting its maximal connected subgraph.

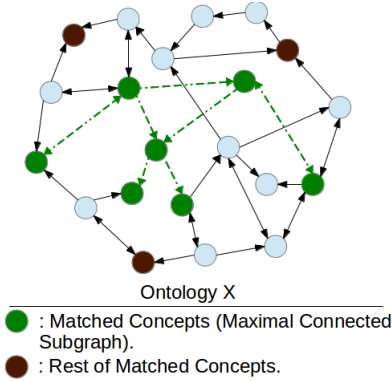


Figure 2: Ontology as a directed graph.

**Definition 2** (Matched Concepts). Let  $O$ ,  $C(O)$ ,  $L_O(c)$  and  $T$  be as in Definition 1; then the set  $M(O, T)$  of matched concepts is calculated as  $M(O, T) = \bigcup_{t \in T} \bigcup_{l \in L_O(c)} \bigcup_{c \in C(O)} H(c, l, t)$ , where:

$$H(c, l, t) = \begin{cases} \{c\} & \text{if } \exists l \in L_O(c) [l = t] \\ \emptyset & \text{otherwise} \end{cases}$$

Given  $M(O, T)$  the graph  $MG(O, T)$ , has the elements in  $M(O, T)$  as nodes and for each pair of nodes  $c_1, c_2$  an edge if the corresponding concepts are linked in  $O$ . Then, for each maximal connected subgraph  $MCSG^i(O, T)$  of  $MG(O, T)$ ,  $MCSG_n^i(O, T)$  is its set of nodes.

**Definition 3** (Degree Centrality of Maximal Connected Subgraph). The maximal degree centrality for  $MCSG^i(O, T)$  is defined as  $DC_m(O, T) = \max(\frac{\sum_{c_h \in MCSG_n^i(O, T)} MD-degree(c_h)}{(n-1)(n-2)})$ , where  $n$  is the number of concepts in  $MCSG$ ,  $MD = \max\{degree(c) \mid c \in MCSG_n^i(O, T)\}$  is the maximum value for the degree function, and  $degree(c) = sub(c) + sup(c) + part\_of(c) + sim(c)$ , gives the number of concepts with at least one relation of type Subclass, Superclass, Part-of or Similarity with a concept  $c$ , respectively.

<sup>6</sup>In case of several maximal subgraphs, we select the highest DC value.

**Definition 4** (Outer Degree Centrality). Given  $M(O, T)$  and  $MCSG^i(O, T)$  as in Definition 2, let  $Q^i(O, T) = M(O, T) \setminus MCSG_n^i(O, T)$ . Then  $DC_o(O, T) = \max(\sum_{c \in Q^i(O, T)} degree(c))$  is the outer degree centrality of  $Q^i(O, T)$ .

Finally we define the *degree centrality* of  $M(O, T)$  as  $DC(O, T) = DC_m(O, T) + DC_o(O, T)$ .

Assuming that ontologies are represented by domains, groups by sets of terms chosen by their owners and users by sets of tags used to adorn annotations, we define three different relevance measures based on **CMM** and **DC**.

1. Group-Domain Relevance: This measure supports groups' owners in identifying appropriate domains by evaluating **CMM** or **DC**, with  $O$  the ontology associated with a domain and  $T$  the set of terms  $Tms(g)$  the owner indicates as characterising group  $g$ .
2. Domain-Users Relevance: This measure supports groups' owners in identifying potential members. Here,  $O$  is the ontology describing the domain to which the group is associated and  $T$  is the set  $Tgs_p(u)$  of tags adorning *public* annotations of a user  $u$ .
3. User-Domains Relevance: This measure supports users in identifying relevant groups according to their interests. The concepts in  $O$  are matched with the set of tags  $Tgs_a(u) = Tgs_p(u) \cup Tgs_v(u)$  adorning *public* and *private* annotations of a user  $u$ .

In all these cases, domains or users are then ranked according to the relevance provided by the adopted measure.

The following pseudocode<sup>7</sup> describes the process by which we rank the relevance of ontologies from a given set for a group  $G$  represented by its terms for both **CMM** and **DC**. We assume that `ontologies` is an array whose elements maintain the information on the different ontologies<sup>8</sup>.

```
ln = ontologies.length; ranksCMM = new array[ln] of int;
relevanceCMM = new array[ln] of float;
ranksDC = new array[ln] of int;
relevanceDC = new array[ln] of float;

for(i = 0; i < ln; i++){
    relevanceCMM[i] = matchCMM(G.terms, ontologies[i]);
    relevanceDC[i] = matchDC(G.terms, ontologies[i]);
}
ranksCMM = computeRanking(relevanceCMM);
ranksDC = computeRanking(relevanceDC);

float matchCMM(gTerms[], ontology) {
    exactMatch = partialMatch = 0;
    oTerms = ontology.getTerms();
    for(i = 0; i < gTerms.length; i++){
        for(j = 0; j < oTerms.length; j++){
            if(gTerms[i] == oTerms[j])
                exactMatch++;
            elseif(isTokenIn(gTerms[i], oTerms[j]))
                partialMatch++;
        }
    }
    return exactMatch + partialMatch;
}
```

<sup>7</sup>Some of the functions were not coded due to paper size limit.

<sup>8</sup>The code uses nested loops to facilitate understanding. Loops are actually implemented through suitable SQL statements.

```

        partialMatch++;
    }
    return exactMatch*0.6 + partialMatch*0.4;
}
float matchDC(gTerms[], ontology) {
    matched = checkMatch(gTerms[], ontology);
    connected = checkMaxConnected(matched, ontology);
    DC = match(matched, ontology);
    rest = matched - connected;
    for(j=0; j<rest.length(); j++)
        DC += degree(rest[j], ontology);
    return DC;
}
float degree(concept, ontology) {
    return sub(concept, ontology) +
        sup(concept, ontology) +
        part_of(concept, ontology) +
        sim(concept, ontology);
}

```

Analogous code is used for the other cases.

### 3.2.3 Example

The following example illustrates the process of matching a group with two ontologies. Given the fragments of the *Animal* and *Plant* ontologies in Fig. 3 where each node (concept) has a sample of lexemes close to it (the white rectangle)<sup>9</sup>, let us suppose we have a group  $G$  with terms (Living Thing, Organism, System, Body, Grows, Operations, Adult). We need to check the relevance of the two ontologies for the group using **CMM** and **DC**.

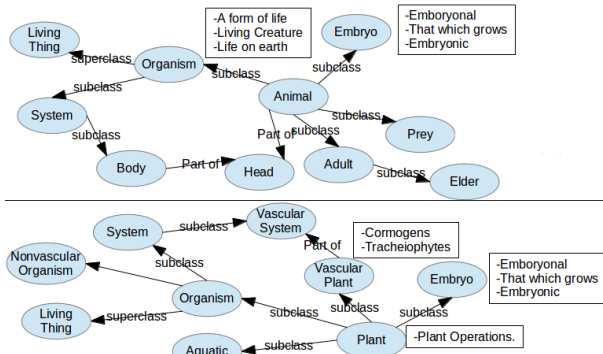


Figure 3: Fragments of *Animal* and *Plant* ontologies.

In particular, we obtain:  $CMM(G, Animal) = (1+1+1+0+0+1)*0.6 + (0+0+0+0+1+0+0)*0.4 = 3.4$  and  $CMM(G, Plant) = (1+1+1+0+0+0+0)*0.6 + (0+1+1+0+1+1+0)*0.4 = 3.4$ .

Based on **CMM**, we can rank the two ontologies:

$$Relevance(G, Animal) = 3.4, Rank(Animal) = 1, Relevance(G, Plant) = 3.4, Rank(Plant) = 1,$$

For the **DC** measure we obtain:

$$DC(G, Animal) = \frac{(3-1)+(3-2)+(3-2)}{(4-1)(4-2)} + 2 = 2\frac{2}{3} \text{ and } DC(G, Plant) = \frac{(4-2)+(4-2)}{(3-1)(3-2)} = 2.5.$$

with ranking:

<sup>9</sup>Both fragments are taken from their corresponding ontologies after being synthesised from BabelNet.

$$Relevance(G, Animal) = 2\frac{2}{3}, Rank(Animal) = 1, Relevance(G, Plant) = 2.5, Rank(Plant) = 2$$

Although the terms for  $G$  are more related to the *Animal* ontology, we observe that the **CMM** equally ranked the two ontologies while the **DC** ranked *Animal* before *Plant*.

### 3.3. URL-based measure

This measure is based on the number of URLs annotated by both group members and external users, under the assumption that such URLs represent shared interests, suggesting potential group members. In [1], we gave a formal definition of the measure and an example of calculation. In this work, we include the **URL** measure in the experimental test to study possible improvements on the results of ontology-based measures.

## 4. Working Scenario

A university uses MADCOW as a coordination tool for its faculty members and students. Taya is a teacher interested in several subjects related to Computer Science, and she would like to direct her interests towards a new field related to robotics. She creates a group titled “Introduction to Robotics” and browses the existing domains by clicking Refer to Domain link, but finds out that no immediately suitable domain exists. Hence, she enters relevant terms such as “robotics”, “intelligence”, “direction”, “motion” to represent the group intent and asks the system to suggest related ontologies for this group by clicking on “Search”. The system looks for the most suitable ontologies available and generates a ranked list by executing **CMM** and **DC** measures. Taya examines the generated lists and picks the “AI” ontology and clicks on “Link” to refer her group to this domain. The concepts in the “AI” ontology become available for users to tag annotations (see Fig. 4).

Class Match Measure			Degree Centrality		
	Ontology	Rank		Ontology	Rank
<input type="radio"/>	music	7	<input type="radio"/>	computer science	6
<input checked="" type="radio"/>	ai	8	<input type="radio"/>	ai	6
<input type="radio"/>	chemistry	9	<input type="radio"/>	fashion	7
Link			Link		

Figure 4: Ranked list (partial) of matched domains (and ontologies) with terms provided by Taya (**CMM** and **DC** Measures).

Taya invites some users whom she knows might be interested in her group and they start submitting annotations. After a while, Taya clicks on Suggest Members to obtain a

list of potentially interested users. The system performs the ontology-based matchings as well as the **URL-based** one and shows Taya three ranked lists of users from the various types of matches, from which she selects and invites users.

Conversely, Daniel has created several annotations related to different subjects and would like to find interesting groups. From his portal menu, he chooses “Suggest Groups”. The system matches all his public and private annotations to all the available domains, presenting him with a ranked list of these domains and their associated groups.

## 5. Experimental Tests

We conducted a pilot test to assess the adequacy of the ontology- and URL-based measures. 20 Participants (university students familiar with MADCOW and employees) were divided into 3 disjoint sets (7, 7, and 6 respectively). Participants from the first set created 11 MADCOW groups and assigned domains for them, then manually invited participants from the second set to join the groups (21 invitations were created). All participants in this set accepted the invitations and submitted annotations to the joined groups (17 annotations). Participants in the third set submitted private and public annotations (28 annotations), all adorned with suitable tags. We asked participants to annotate a set of similar websites to check **URL** matching (for this we used 10 different websites). Group owners requested the system to suggest members and invitations were sent (18 invitations). Participants of the third set requested the system to suggest proper groups and sent membership requests (15 requests). After the test, we asked all participants to fill an online questionnaire about the three measures, the results being reported in Tables 1 and 2 for owners and users, respectively, (where *G. Asso.* stands for group-domain association, *M.* for member, *Presence* for presence of irrelevant ontologies or members, and *Absence* for absence of relevant ontologies or members). Figure 5 shows data for the corresponding ontological measures.

Results show a preference for suggestions obtained using **DC** over those using **CMM**, consistently with the hypothesis that considering relations between concepts in an ontology can favour the process of identifying suitable groups or users. Although partial match in **CMM** retrieves more results, this could be a source of confusion to some users, since some of these might be irrelevant.

Table 1: OWNERS’ ASSESSMENT OF MEASURES.

		<i>Adequate</i>	<i>Presence</i>	<i>Absence</i>
<b>CMM</b>	<i>G. Asso.</i>	66%	18%	16%
	<i>M. Suggestion</i>	76%	22%	2%
<b>DC</b>	<i>G. Asso.</i>	82%	2%	16%
	<i>M. Suggestion</i>	92%	3%	5%
<b>URL</b>	<i>Needed (M. Sugg.)</i>	80%		
	<i>Not needed (M. Sugg.)</i>	20%		

Table 2: MEMBERS’ ASSESSMENT OF MEASURES.

	<i>Adequate</i>	<i>Presence</i>	<i>Absence</i>
<b>CMM G. Suggestion</b>	56%	16%	28%
<b>DC G. Suggestion</b>	86%	9%	5%
<b>URL G. Suggestion</b>	<i>Needed</i>	72%	
	<i>Not needed</i>	28%	

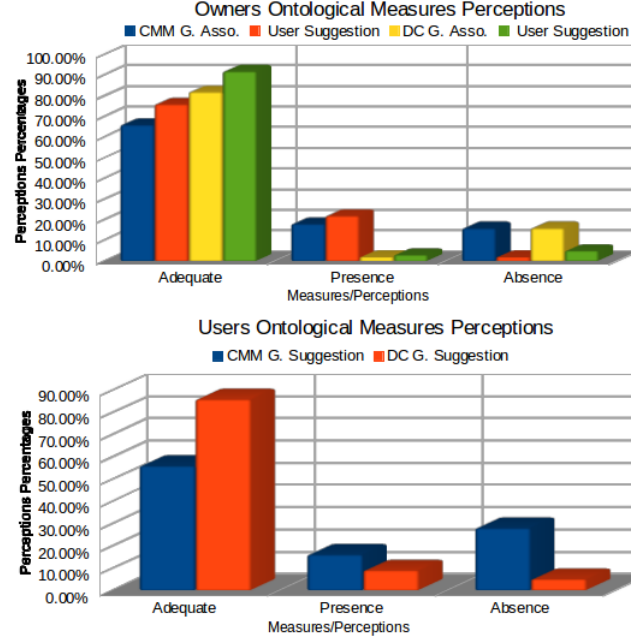


Figure 5: Owners’ and Members’ assessment of ontology-based measures.

To be sure of participants’ credibility in filling questionnaires, and to ensure correctness of relevance calculations and ranking, we stored (in the background) the ranked results as well as the choices of the participants in order to calculate the percentage of participants who chose the 1st and the 2nd top ranked results (Table 3). Values in the table appear to be consistent with the data in Tables 1 and 2.

Table 3: 1ST AND 2ND SELECTIONS.

	<i>Asso.</i>	<i>Users Sugg.</i>	<i>Groups Sugg.</i>
<b>CMM</b>	52%, 24%	20%, 30%	45%, 15%
<b>DC</b>	75%, 25%	66%, 16%	75%, 14%
<b>URL</b>		80%, 10%	50%, 16%

Table 4 presents the average times (in seconds) for computing the measures involved in the tests. The reason for the higher cost of **CMM** lies in the use of partial matching.

Table 4: MATCHING OPERATIONS AVERAGE TIMES.

	<b>CMM</b>	<b>DC</b>	<b>URL</b>
<i>Domain Asso.</i>	1.13	0.85	—
<i>Sugg. Users</i>	2.15	1.43	1.25
<i>Sugg. Groups</i>	2.35	0.73	1.36

## 6. Conclusions and future work

We have introduced ontologies to represent domain knowledge relevant to the formation of groups in the MADCOW annotation system and realised tools to find matches between groups and potentially interested users. Tools for facilitating the retrieval of interested users (or of interesting groups) are provided based on matches between tags freely used by submitters of annotations and lexemes contained in the different ontologies integrated in the MADCOW system. Experimental results show that **DC** is preferred over **CMM** for suggestions. In general, **DC** improved the suggestion process due to its revised list of results compared with **CMM**, minimising users confusion in selecting the most appropriate items. Moreover, the evaluation of **CMM** consumes more time than **DC**, which makes the latter faster in generating the suggestion lists. Experimental results also indicate that reference to URLs can further improve results provided by ontology-based measures.

As future work, we plan to proceed with deeper experimental investigation and to explore possible enhancements to the adopted measures. An interesting candidate measure is based on the notion of Clustered Concepts, considering the number and size of clusters of matched concepts, as well as their relative closeness, based on the idea that the ontology with closest concepts is the most relevant.

## References

- [1] D. Avola, P. Bottoni, and A. Hawash. Using ontologies for users-groups matching in an annotation system. In *Proc. CSIT 2013*, pages 38–44, 2013.
- [2] D. Avola, P. Bottoni, and A. Hawash. Users-groups matching in an annotation system: Ontological and URL relevance measures. In *Proc. CSIT 2014*, pages 100–109, 2014.
- [3] D. Avola, P. Bottoni, S. Levialdi, and E. Panizzi. Annotation threads in MADCOW 2.0. In *Proc. HC'10*, pages 159–166. ACM, 2010.
- [4] P. Bottoni, S. Levialdi, N. Pambuffetti, E. Panizzi, and R. Trinchese. Storing and retrieving multimedia web notes. *Int. J. Comput. Sci. Eng.*, 2(5/6):341–358, 2006.
- [5] R. Braga, C. Werner, and M. Mattoso. Using ontologies for domain information retrieval. In *Proc. DEXA 2000*, pages 836–840, 2000.
- [6] C. Brewster, K. O'Hara, S. Fuller, Y. Wilks, E. Franconi, M. A. Musen, J. Ellman, and S. B. Shum. Knowledge representation with ontologies: The present and future. *IEEE Intelligent Systems*, pages 72–81, January 2004.
- [7] B. Chandrasekaran, J. R. Josephson, and V. Benjamins. What are ontologies, and why do we need them? *IEEE Intelligent Systems*, 14(1):20–26, 1999.
- [8] E. Estrada and J. A. Rodríguez-Velázquez. Subgraph centrality in complex networks. *Phys. Rev. E*, 71:056103–1/9, 2005.
- [9] M. Fazel-Zarandi and M. Fox. Reasoning about skills and competencies. In *CProc. PRO-VE 2010*, volume 336 of *IFIP AICT*, pages 372–379. Springer, 2010.
- [10] F. Fonseca, M. Egenhofer, C. Davis, and K. Borges. Ontologies and knowledge sharing in urban GIS. *CEUS*, 24:232–251, 2000.
- [11] L. C. Freeman. Centered graphs and the structure of ego networks. *Mathematical Social Sciences*, 3(3):291–304, 1982.
- [12] R. Gil, A. Borges, and L. Contreras. Shared ontologies to increase systems interoperability in university institutions. In *Proc. IMCSIT 2007*, pages 799–808, 2007.
- [13] N. Guarino. Formal ontology and information systems. In *Proc. FOIS'98*, pages 3–15. IOS Press, 1998.
- [14] A. W. Hawash. Introducing groups to an annotation system. In *CHIItaly (Doctoral Consortium)*, volume 1065 of *CEUR Workshop Proceedings*, pages 43–54, 2013.
- [15] B. Hoser, A. Hotho, R. Jüschke, C. Schmitz, and G. Stumme. Semantic network analysis of ontologies. In *The Semantic Web: Research and Applications*, volume 4011 of *LNCS*, pages 514–529. 2006.
- [16] A. Landherr, B. Friedl, and J. Heidemann. A critical review of centrality measures in social networks. *Business & Information Systems Engineering*, 2(6):371–385, 2010.
- [17] Y. Li, D. Mclean, Z. Bandar, J. O'Shea, and K. Crockett. Sentence similarity based on semantic nets and corpus statistics. *IEEE KDE*, 18(8):1138–1150, 2006.
- [18] R. Navigli and S. P. Ponzetto. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250, 2012.
- [19] R. Navigli and S. P. Ponzetto. Multilingual WSD with just a few lines of code: the BabelNet API. In *Proc. ACL 2012 (System Demonstrations)*, pages 67–72, 2012.
- [20] C. Ni, C. R. Sugimoto, and J. Jiang. Venue-author-coupling: A measure for identifying disciplines through author communities. *JASIST*, 64(2):265–279, 2013.
- [21] J. Paralic and I. Kostial. Ontology-based information retrieval. In *Proc. IIS 2003*, pages 23–28, 2003.
- [22] S. K. Rhee, J. Lee, and M.-W. Park. Ontology-based semantic relevance measure. In *Proc. SWW 2.0*, volume 294 of *CEUR Workshop Proceedings*, 2007.
- [23] H. Tangmunarunkit, S. Decker, and C. Kesselman. Ontology-based resource matching in the grid - the grid meets the semantic web. In *Proc. ISWC 2003*, volume 2870 of *LNCS*, pages 706–721, 2003.
- [24] D. Vallet, M. Fernández, and P. Castells. An ontology-based information retrieval model. In *Proc. ESWC 2005*, pages 455–470. Springer, 2005.
- [25] P. Velardi, S. Faralli, and R. Navigli. OntoLearn reloaded: A graph-based algorithm for taxonomy induction. *Computational Linguistics*, 39(3), 2013.
- [26] Y. Zhang, W. Vasconcelos, and D. Sleeman. Ontosearch: An ontology search engine. In *Research and Development in Intelligent Systems XXI*, pages 58–69. Springer, 2005.
- [27] H. Zhuge and J. Zhang. Topological centrality and its e-science applications. *JASIST*, 61(9):1824–1841, 2010.

# Evolutionary Crowdsourcing Approach for Quality Control

Duncan Yung and Shi-Kuo Chang

*Department of Computer Science, University of Pittsburgh*  
{duncanyung, chang}@cs.pitt.edu

**Abstract**—Crowdsourcing is widely used for solving simple tasks (e.g. tagging images) and recently, some researchers [9][10] propose new crowdsourcing models to handle complex tasks (e.g. articling writing). In both type of crowdsourcing models (for simple and complex tasks), voting is a technique that is widely used for quality control [9]. However, we argue that voting is actually a technique that selects a high quality answer from a set of answers. It does not directly enhance answer quality. In this paper, we propose a new crowdsourcing approach that can incrementally improve answer quality. The new approach is based upon two principles - evolutionary computing and slow intelligence, which helps the crowdsourcing system to propagate knowledge among workers and incrementally improve the answer quality. We perform explicitly 2 experimental case studies to show the effectiveness of the new approach. The case study results show that the new approach can incrementally improve answer quality and produce high quality answers for non-trivial tasks.

**Keywords:** Crowdsourcing, Evolutionary Computing, Slow Intelligence Approach, Quality Control.

## I. INTRODUCTION

Crowdsourcing is the practice of solving problems by combining contributions from a large group or the public. It is widely used for solving simple tasks (e.g. tagging images) and recently, some researchers [9][10] propose new crowdsourcing models to handle complex tasks (e.g. articling writing). In both type of crowdsourcing models (for simple and complex tasks), voting is a technique that is widely used for quality control [12][9][13]. For example, a crowdsourcing system may ask workers to vote for good answers, and discards poor ones. In the other words, the voting approach actually selects high quality answers from a set of answers that are provided by a group of independent workers. Hence, the idea of propagating knowledge so as to incrementally improve the answer quality is lacking.

Evolution is a kind of slow intelligence. It is the process of incrementally improving quality or achieving a goal in a given environment, and it is widely used in many different applications [8][14][11][18]. Theoretically, it is so powerful that it has the ability to tackle any search space provided that initialization and variation operators are available [6].

In this paper, we propose a new crowdsourcing approach that help to propagate knowledge among workers and as a result, improve answer quality of crowdsourcing task through evolution. The challenges of building such approach lie in two aspects: 1) how can the crowdsourcing system model

the process of evolution; 2) how to make the approach to be easily incorporated in existing crowdsourcing models?

In order to address these challenges, we propose a new approach based upon two principles - slow intelligence and evolutionary computing. Under the context of crowdsourcing, the notion of slow intelligence suggests that quality of answer can be improved if time is allowed for more workers to work on the task. However, we argue that solely allowing more workers to work on a task independently is not enough for tremendously improving the answer quality as knowledge is not accumulating. Because of that, we propose a novel mechanism—the evolutionary feedback strategy—that allows workers to propagate their knowledge to others without direct communication (to address challenge 1).

The evolutionary feedback strategy models the behavior of creature evolution. It gradually eliminates poor quality answers and only propagates high quality answers to new workers. Based on the high quality answers, new workers can generate new answers with higher quality. Therefore, the answer quality can be improved incrementally without direct communication between workers. Furthermore, evolution is a process that, once started, does not need any external intervention. As a result, users only have to define the task and objective for the evolution process to start. They do not need to even define the answer quality function or decide on how to partition the task, which makes the approach to be easily adopted by any existing crowdsourcing models (to address challenge 2).

The rest of the paper is structured as follows: Section II describes related work and background. An overview of the new crowdsourcing approach is provided in Section III. Sections IV describes formally the two computation cycles for the new crowdsourcing approach. We present 2 detailed case studies in Section V and VI to show the effectiveness of our approach. Section VII shares with you our experiences of this work and Section VIII concludes this paper.

## II. RELATED WORK AND BACKGROUND

The work related to this paper falls under 3 categories: slow intelligence system, evolutionary computing, and crowdsourcing.



### A. Slow Intelligence System

The concept of slow intelligence suggests that the relationship between quality of output and time is a natural trade-off. An artificial system that follows the slow intelligence principle can be called an Artificial Slow Intelligence System (ASIS) [3]. An Artificial Slow Intelligence System consists of 6 important characteristics that are useful for our crowdsourcing model. They are:

- 1) **Enumeration:** Given a problem, the system can improve solution quality as well as the confidence of correctness by enumerating different possible solutions.
- 2) **Adaptation:** Solutions are adapted and improved according to the system requirement.
- 3) **Elimination:** Poor solutions are eliminated so that only the appropriate ones are further considered.
- 4) **Concentration:** If there exist too many valid solutions accumulated in the system, the system selects and stores only the high quality ones, so as to maintain system efficiency.
- 5) **Propagation:** The system is aware of its environment and constantly exchanges information with the environment.
- 6) **Multiple Decision Cycles:** The multiple decision cycles enable the SIS to both cope with the environment and meet long-term goals.

The concept of slow intelligence is used by some existing applications. Wang et al [16] propose a slow intelligence approach for searching appropriate feature selection algorithms; Dong [4] argues that human intelligence can be modeled in the slow intelligence framework; Colace et al [5] propose a network management tool based on slow intelligence principle and ontology based techniques.

### B. Evolutionary Computing

The idea behind evolutionary computing is that given a population of individuals, the environmental pressure causes natural selection and hereby the fitness of the population is growing [6]. It is used to solve many real world problems. Hu et al [8] use genetic algorithms to solve the problem of air traffic control in multi-runway systems; Szlapczynski et al [14] apply evolutionary algorithms and some assumptions of game theory to solve ship encounter situations; Lima et al [11] propose a hybrid algorithm which combines support vector regression with evolutionary strategy for predictive models in the environmental sciences; Zhang et al [18] develop an approach and prototype for selecting optimal material constituent compositions. There are more developments and applications of evolutionary computing are not covered in this related work.

### C. Crowdsourcing

Crowdsourcing is the practice of solving problems by combining contributions from a large group or the public.

This process is often used to subdivide tedious work. For example, the crowd may be invited to experiment with a new technology, carry out a design task, refine or carry out the steps of an algorithm, or help capture, systematize, or analyze large amounts of data.

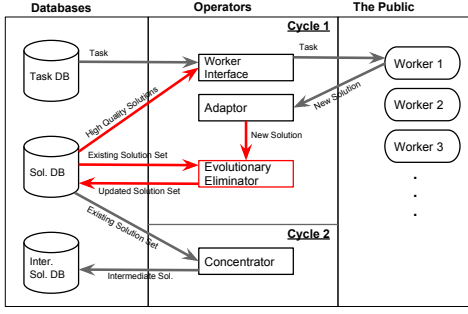
Crowdsourcing is widely used for completing tasks and improve computing efficiency. Franklin et al [7] propose CrowdDB which uses human input via crowdsourcing to process queries that neither database systems and search engines can answer. Wang et al [15] propose a hybrid human-machine approach that finds records that refer to the same entity. The approach uses machine to coarsely pass over all data, and crowdsourcing to verify only mostly likely matching pairs. Yan et al [17] present hybrid image search queries that use machine to generate candidate query results and crowdsourcing platform to validate the results. Bernstein et al [2] present a word processing interfaces that enables writers to request crowdsourcing workers to edit their documents on demand. Parameswaran et al [12] consider the problem of filtering data items based on a set of properties that can be verified by humans in a crowdsourcing platform. They develop deterministic and probabilistic algorithms to optimize the expected cost and error, while our work is focusing on improving the solution quality without extra monetary cost.

Kittur et al [9] propose a Map-reduce style crowdsourcing framework for accomplishing complex and interdependent tasks. In their framework, users have to explicitly design their partition, map, reduce, and quality control tasks and strategies. Kulkarni et al [10] propose Turkomatic, which is a tool that recruit workers to decompose and solve tasks. Turkomatic also allows requester to intervene in the middle of the task.

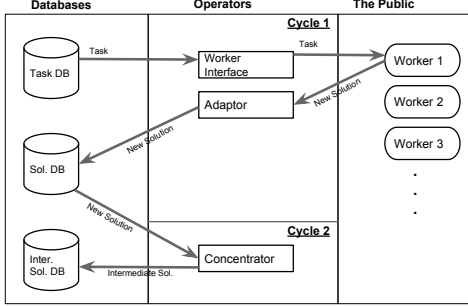
Although there exist a lot of works about crowdsourcing, this work is the first one that proposes using the notion of slow intelligence [3] and evolutionary computing to enhance the answer quality of crowdsourcing tasks.

## III. OVERVIEW

Fig 1(a) is the system architecture of our new crowdsourcing approach. The main system contains 2 cycles. Cycle 1 is mainly for solution collection and cycle 2 is for solution computation and filtering. Cycle 1 contains mainly 2 phases: adaptation and elimination phase (IV-A and IV-B). The adaptor collects solutions from workers and transfers them the evolutionary eliminator; The evolutionary eliminator eliminates existing solutions and updates the solution database. For cycle 2, it only contains the concentration phase (Section IV-D) which helps the system to find out the high quality solutions.



(a). Crowdsourcing System using Evol. Approach



(b). Traditional Crowdsourcing System

Fig. 1. Two different Crowdsourcing Approaches for Quality Control

Comparing to the traditional crowdsourcing approach (Fig 1(b)), our new approach uses an evolutionary eliminator to eliminate solution and offer references to new workers (highlighted in red in Fig 1(a)), which is a form of knowledge propagation.

#### IV. EVOLUTIONARY CROWDSOURCING APPROACH

In this section, we formally define the two cycles of the new approach using SIS operator ( $+adap=$ ,  $->elim-$ , and  $>conc=$  are SIS operators which are not traditional mathematics symbols).

##### Definition 1: [Cycle 1-Adaptation and Elimination]

$$guard[1,2] \ P \ +adap= (P_{sol}, P_{sol}^{FB}) \ ->elim- \ S_{sol}$$

where

- $P$  is a crowdsourced problem,
- $(P_{sol}, P_{sol}^{FB})$  is a pair of solution and feedbacks from a worker,
- $S_{sol}$  is a set of qualified solutions, and
- $guard[1,2]$  is a guard that determines which cycle to switch to.

The system presents the problem to workers. Then, the system waits for workers to submit their solutions  $P_{sol}$  and feedback  $P_{sol}^{FB}$  set. After that, the poor quality solutions are eliminated and only high quality ones  $S_{sol}$  are retained. Finally, the guard object  $guard[1,2]$  determines which cycle the system

will switch to.

##### Definition 2: [Cycle 2-Concentration]

$$guard[1,terminate] \ S_{sol} \ >conc= \ S_{sol}^{inter}$$

where

- $S_{sol}$  is the set of solutions after elimination,
- $S_{sol}^{inter}$  is set of solution for  $P$  that accumulates solution with score higher than threshold, and
- $guard[1,terminate]$  means if time is up or budget becomes zero, terminate the system; otherwise, switch back to cycle 1.

Solutions in the solution set  $S_{sol}$  are used to compute intermediate solutions. Then, the guard object  $guard[1,terminate]$  determines which cycle the system will switch to.

Below is the detailed description of each phase in cycle 1 and 2.

##### A. Cycle 1-Adaptation

In this phase, the system presents the problem  $P$ , objectives, and a set of existing solutions  $S_{sol}$  to workers.  $S_{sol}$  is empty for the first worker. Then, the system waits for solutions  $P_{sol}$  and feedbacks  $P_{sol}^{FB}$  (e.g. rating of existing solutions) from workers, where size of  $|P_{sol}| = 1$  and  $|P_{sol}^{FB}|$  is the number of existing solutions. (We will illustrate the purpose of collecting feedbacks in Section IV-B.) However, not all solutions from workers are kept in  $S_{sol}$ . In later cycles, poor quality subtask solutions will be removed from  $S_{sol}$  based on our proposed **Evolutionary Strategies** (in Section IV-B).

##### B. Cycle 1-Elimination

Given  $P_{sol}$  and  $P_{sol}^{FB}$  as input, the system updates the current solution set  $S_{sol}$  based on the evolutionary feedback strategy that we propose in Section IV-C. The updated set of current solution is offered to new workers. If the requester wants to get an intermediate solution from  $S_{sol}$ , the guard ( $guard[1,2]$ ) allows the system to proceed to cycle 2; otherwise, the system stays in cycle 1.

##### C. Evolutionary Feedback Strategy

The evolutionary feedback strategy is inspired by the process of natural selection, which gradually eliminates poor genes that do not fit the environment. Here, the goal of our evolutionary feedback strategy is to preserve solutions that are useful for new workers to come up with higher quality solutions.

In short, the system computes a **Feedback Score**  $r_j$  and a new **Impact Score**  $Q_j^{new}$  for each existing solution in  $S_{sol}$  after the system receives an solution set  $(P_{sol}, P_{sol}^{FB})$  from a worker. The new impact score (Definition 3) takes

into account the current impact score, current feedback, and the historical feedbacks. It also allows users to fine-tune the formulation by changing the parameter  $\beta$ . More details are presented below.

After the system receives  $P_{sol}^{FB}$ , the system updates the current impact score  $Q_j^{curr}$  of each existing solution  $Sol_j$  in  $S_{sol}$  to the new impact score  $Q_j^{new}$  (The impact score of the new solution is not updated using Definition 3.).

We define the impact score and feedback score to be:

Let,

- $r_j$  be the feedback score of  $Sol_j \in S_{sol}$ ,
- $Q_j^{curr}$  be the current impact score of  $Sol_j \in S_{sol}$ , and
- $Q_j^{new}$  be the new impact score of  $Sol_j \in S_{sol}$ .

**Definition 3: [Impact Score]** is the weighted sum of feedback score  $r_j$  and current impact score  $Q_j^{curr}$  and it is only for existing solutions. It is defined as:

$$Q_j^{new} = \beta r_j + (1 - \beta) Q_j^{curr}$$

,where  $\beta \in [0, 1]$  is a user input parameter.

**Definition 4: [Feedback Score]** is the diminished average of historical feedbacks  $FB_{Sol_j}^k, \forall k = 1 \dots m$  and the current feedback  $FB_{Sol_j}^{m+1}$ . It is defined as:

$$r_j = \frac{\sum_{k=1}^{m+1} FB_{Sol_j}^k \times k}{\sum_{k=1}^{m+1} k}$$

,where  $m$  is the number of historical feedback.

In Definition 4, we assume that  $FB_{Sol_j}^i$  appears before  $FB_{Sol_j}^k$  if  $i < k$ . Definition 4 intends to take into account historical feedbacks as the most recent feedback may be an outlier which does not reflect the quality of the solution. However, the effect of historical feedbacks should be diminished with time. That is because the older the feedback, the less it can reflect the quality of the solution.

After that, the system eliminates all solutions in  $S_{sol}$  that have score  $Q_j^{new} < \tau$ , where  $\tau$  be a user defined threshold. The system can also use a top-k strategy, which retains  $k$  solution(s) with the k-highest score(s)  $Q_j^{new}$ .

Since a new solution was not rated and it does not have  $Q_j^{curr}$ , Definition 3 is not applicable for computing its impact score. A way to compute its impact score  $Q_j^{new}$  is needed. We propose to compute the impact score based on the worker's rating on existing solutions, the quality of existing solutions and an improvement factor  $I_f$ .

**Definition 5: [Initial Impact Score]** is the impact score of a

new solution. It is defined as:

$$Q_j^{new} = \begin{cases} \min \left[ \left[ \sum_{j=1}^k \frac{FB_{Sol_j}^{m+1} \times Q_j^{curr}}{\sum_{j=1}^k FB_{Sol_j}^{m+1}} \right] \times I_f, S_{max} \right] & \text{when } k > 0 \\ S_{max} \times \frac{3}{4} & \text{when } k=0 \end{cases}$$

,where  $k$  is the number of task being rated,  $FB_{Sol_j}^{m+1}$  is the current feedback of existing solution  $j$ ,  $I_f \in [1, 2]$  is the improvement factor, and  $Q_j^{curr}$  is the current quality of the task being rated.

As workers are offered existing solutions when answering the task, we propose to take into account the quality of existing solutions and how much the worker learns from each existing solutions when we compute the initial impact score of a new solution. Also, it is expected that a normal worker will offer a solution with higher quality than existing solutions. Therefore, we define  $Q_j^{new}$  to be  $\sum_{j=1}^k \frac{FB_{Sol_j}^{m+1} \times Q_j^{curr}}{\sum_{j=1}^k FB_{Sol_j}^{m+1}}$  multiplying by  $I_f$ , where  $I_f$  is a constant that represents the expected improvement of the new solution.

#### D. Cycle 2-Concentration

Cycle 2 is a cycle that generates intermediate solution from a set of existing solution. A typical crowdsourcing model has only 1 solution. However, for our new crowdsourcing model, there are more than 1 solution. Hence, we suggest requesters to look for an intermediate solutions that are submitted recently by workers as recent solutions tend to have higher quality in the evolutionary approach. After that, the system is terminated if time is up or budget is used up; otherwise, the system is switched back to cycle 1.

### V. CASE STUDY 1- 2-DAY TRAVEL PLAN

#### A. Details

A tourist wants to visit Baltimore (Maryland, USA) for 2 days. Based on that, we constructed two crowdsourcing tasks using our evolutionary approach and the independent approach. For the evolutionary approach, we present existing answers for workers to refer to and ask them to rate the offered existing answers. We calculate the impact score and feedback score based on Definition 3, 4, and 5 and we set  $\beta$ ,  $I_f$ , and  $\tau$  to be 0.5, 1.7, and 6 respectively. For the independent approach, we just ask workers to do the task without offering any references. In the experiment, we use the same among of budget (US\$1/answer) for both approaches and we provide below objectives to workers in both approaches:

- 1) maximize the number of sightseeing spots that she can visit,
- 2) maximize the total duration at sightseeing spots,
- 3) minimize total expenditure,
- 4) and minimize transportation time.

## B. Results

We posted the tasks to Amazon Mechanical Turk [1] for 3 days and we collected 17 answers from workers for each approach. After that, we posted the 34 (17/approach) answers to Amazon Mechanical Turk [1] for workers to rate and we collected 15 set of answers (each set has 17 rates.).

### Answer Quality

Below, we compare the two highest score plans from the two approaches. Sightseeing spots are in bold letter and details and information of sightseeing spots are underlined.

#### 1) Evolutionary Approach $Ans_E$ :

**Day 1 Plan:** Enjoy history? Politics? Art? Architecture? Why not explore all four in one day? Begin your day in Baltimore's historic Mount Vernon neighborhood. The first stop is the **Washington Monument and Museum**, home of the oldest monument honoring George Washington. This historic gem was designed by Robert Mills, architect of the Washington Monument in the District of Columbia. Next, walk a few blocks to the **Walters Art Museum**, where you'll spend hours admiring collections from 19th Century American and European Masterpieces to Greek sculptures. After spending the afternoon sightseeing in the Mount Vernon district, take the Charm City Circulator (a free rapid transit bus) to Downtown Baltimore and visit **Emerson Bromo-Seltzer Tower**, an architecturally unique historic clock-tower now used as studio space. Finally, continue on to the last stop, **Baltimore City Hall**. The building features fantastic Second-Empire, Baroque style architecture and is in the middle of Downtown Baltimore and within walking distance of the **Inner Harbor**.

**Day 2 Plan:** The second day can be spent exploring Baltimore's Inner Harbor. This district has a plethora of different tourist attractions and venues, ranging from the **National Aquarium** to the **Maryland Science Center** to the **Port Discovery Children's museum**. In a single day, you can visit exhibits which allow experiencing Atlantic Coral Reefs, tropical rain forests, and even distant planets. After a full day of exploring land, sea and air the **Inner Harbor** also provides ample opportunities for sampling local fare. Restaurants like, Nick's Fish House & Grill and Mezze, are a famished foodie's delight. Most are family-friendly and won't bust your budget.

#### 2) Independent Approach $Ans_I$ :

**Day 1 Plan:** Baltimore's Inner Harbor is a gem that cannot be missed. There is no better place to begin your two-day adventure. The region boasts several spectacular (and family friendly) attractions, like the **National Aquarium**, **Maryland Science Center**, and **American Visionary Art Museum**. All of these attractions are easily accessible and a short (bus or water taxi) ride away. You can finish up your day with a delicious meal at the famed **Thames Street Oyster House**.

**Day 2 Plan:** No trip to Baltimore is complete without visiting Camden Yards. This sports complex actually contains several venues (**Oriole Park**, **M&T bank Stadium**, the **Sports Legends Museum**, **Babe Ruth Birthplace and Museum**, and **Baltimore Grand Prix**) and is in close proximity to the Inner



(a) Evol. Approach Day 1 (b) Independent Approach Day 1

(c) Evol. Approach Day 2 (d) Independent Approach Day 2

Fig. 2. Sightseeing Spots and Routes

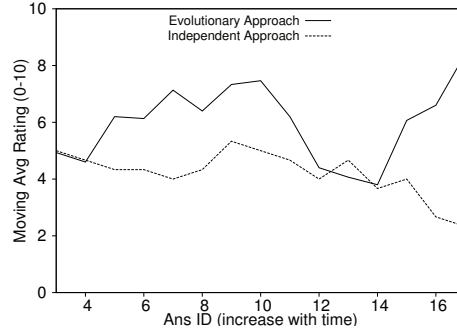


Fig. 3. Moving Average Rating (Case Study 1)

Harbor. Enjoy a day at the ballpark and take in a museum tour in Baltimore's most famous sports area.

Figure 2 shows the sightseeing spots and routes in above plans. The total number of sightseeing spots in both plans are 9 and the length of each route is around 4 miles. Hence, in term of number of sightseeing spots and transportation time, both plans are very similar. We believe that it is because most of the workers offered popular sightseeing spots in Baltimore. However, it is clear that the plan produced by the evolutionary approach contains a lot more details and information (underlined words) about sightseeing spots than the independent approach. Therefore, this case study shows that the evolutionary approach produces better quality solutions while using the same amount of budget (the average rating of  $Ans_E$  8.4 versus 5.3 for  $Ans_I$ ,  $t(14) = 10, p < 0.00001$ ).

## Trend of Answer Quality

Figure 3 shows the moving average ratings of 3 consecutive answers. We can see that the moving average rating of the evolutionary approach has an increasing trend between answer ID 4 and 10. After that, the moving average rating drops until answer ID 14. That is because the appearance of a few poor performance workers. However, the moving average rating continues to go up after answer ID 14. On the other hand, we cannot see any increasing trend in the moving average rating of the independent approach. It is always between 3 to 5.3. From this case study, we can see that the evolutionary approach can improve the answer quality although the existence of poor workers still affect the answer quality (the average rating of the 17<sup>th</sup> answer 8.4 versus the rating of the 7<sup>th</sup> 6.9 answer,  $t(14) = 4.841, p = 0.000262$ ; the average rating of the 7<sup>th</sup> answer 6.9 versus the rating of the 1<sup>st</sup> answer 4.4,  $t(14) = 5.158, p = 0.000145$ ).

## VI. CASE STUDY 2-OUTLINE FOR AN ARTICLE

### A. Details

In this case study, we repeat the quality control experiment, which requires workers to generate an outline for an article on Gulf of Mexico oil spill, in [9]. The only difference is that we only require workers to generate a 4-paragraph outline in this case study. We compare three approaches-1)the evolutionary approach, 2)the independent approach, and 3)the map-reduce approach in the quality control experiment of [9] (Quality Control Section *P.46 – 47*). The evolutionary approach and the independent approach are the same as in Section V, except we offer workers with top-2 existing answers instead of setting  $\tau$  to 6; the map-reduce approach asks 5 workers to independently create 5 outlines. Then, the 5 outlines are randomly assigned to 5 different sets of three outlines (outlines could be in more than one set). Each set is given to a different worker, who is asked to create a new outline for the article using elements from the 3 outlines in his or her set.  $\$1/Hit$  is used for all three approaches.

### B. Results

We posted the task to Amazon Mechanical Turk [1] for 2 days and we collected 10 answers from workers for each approach. After that, we posted 30 answers to Amazon Mechanical Turk [1] for another set of workers to rate and we collected 15 set of answers (each set has 30 rates.).

### Answer Quality

Below, we compare the three highest score plans of the three approaches.

#### 1) Evolutionary Approach ( $Ans_E$ ):

- Paragraph 1: Introduction
  - 1) WHAT happened? (oil spill)
  - 2) WHERE and WHEN did it happen? (Gulf Coast, provide specific city and largest city nearest)

- 3) HOW did it happen? (detailed, concise explanation)
- 4) WHO or what entity is taking responsibility?
- 5) Does the local or federal government have anything to say on the matter?

#### • Paragraph 2: Ramifications

- 1) WHO is affected by the oil spill? (What people, businesses, industries, etc?)
- 2) HOW are they currently affected? (Are they entirely displaced? Is this merely a setback? How bad is it?)
- 3) WHAT will need to happen in order to return this population to its former status quo? (Details about clean-up efforts?)
- 4) WHEN do those in charge of the restoration expect life will return to normal for the affected population?

#### • Paragraph 3: Preventative measures

- 1) HOW MANY oil spills have affected the region in the past twenty years? (Give brief history)
- 2) WHAT, if anything, is currently being proposed to avoid another oil spill?
- 3) If no preventative solutions are on the table, WHY?
- 4) If preventative measures are being suggested, by WHOM? HOW do they plan to implement these measures?
- 5) How does the affected population feel about preventative measures being taken? The local and federal government?

#### • Paragraph 4: Conclusion

- 1) HOW will this incident and other recent spills affect the oil industry as a whole?
- 2) WHAT solutions are other oil companies or other regions considering to address the problem of spills IN GENERAL?
- 3) Closing sentences, 1-2. Include any unique insights on the situation that could be derived from the information already provided.
- 4) If no insights can be gleaned, end with a quote from someone affected or someone hoping to rectify the situation.

#### 2) Independent Approach ( $Ans_I$ ):

##### • Paragraph 1: Introduction to the Gulf of Mexico

- 1) Physical description of the site
  - a) Depth of shipping lanes
  - b) Weather patterns
- 2) Ecosystem of the site
  - a) Wildlife on the coast and in the ocean
- 3) Industrial history of the site
  - a) Overview of shipping industry in the Gulf
  - b) Growth over time of industry in the regions
  - c) Major industry other than oil in the Gulf

##### • Paragraph 2: Overview of oil industry in and around Mexico

- 1) History
  - a) When did oil start being shipped in the Gulf
  - b) What companies do the shipping, and what is their nationality
- 2) Current statistics about oil transport in the region
  - a) How much oil is transported through the Gulf each years
  - b) Where is the oil shipped from

- c) Where is the oil shipped to
- d) What is the economic impact of the oil transport in and around Mexico
- Paragraph 3: The Oil Spill
  - 1) What happened and when
  - 2) Aftermath, and how is it today
- Paragraph 4: Conclusions
  - 1) What is being done
  - 2) What needs to be done
  - 3) How to prevent a repeat

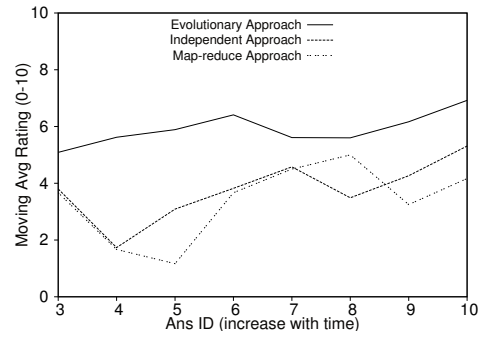


Fig. 4. Moving Average Rating (Case Study 2)

### 3) Map-reduce Approach ( $Ans_{MR}$ ):

- Paragraph 1: Introduction
  - 1) The Deep water Horizon was an offshore oil drilling rig which exploded on the night of April 20, 2010 while working on a well on the sea floor in the Gulf of Mexico
  - 2) Briefly mention how much damage took place, how much oil was spilled and how wide of an area was affected
- Paragraph 2: Explosion
  - 1) Workers had started building a cement casing to reinforce the well when methane gas shot up from the well ignited and exploded
  - 2) At least mention who was hurt
- Paragraph 3: Stopping the leak
  - 1) Closing the blowout preventer
  - 2) Drilling a relief well
  - 3) Capping the well
- Paragraph 4: Containing the spill and protecting the shore
  - 1) Use of booms
  - 2) Oil-eating microbes
  - 3) Dispersant
  - 4) Skimmer ships
  - 5) Burning oil slicks
- Paragraph 5: Aftermath
  - 1) The effects of the oil spill of wildlife, tourism, fisheries, the economy of states affected and other lingering environmental impacts
  - 2) The negative public perception of British Petroleum
  - 3) In conclusion you can discuss the environmental impacts still experienced today
  - 4) Whether this disaster has affected the future of drilling for oil in the ocean

Although answers of the evolutionary approach ( $Ans_E$ ) and the independent approach ( $Ans_I$ ) are rated the same by all workers (Both got average score 8.5 and their standard deviations are 1.6 and 2 for  $Ans_E$  and  $Ans_I$  respectively.),  $Ans_E$  is more related to Gulf of Mexico oil spill. We can see that paragraph 1 and 2 of the  $Ans_I$  are about some facts of Gulf of Mexico; paragraph 3 and 4 of  $Ans_I$  are about the oil spill event, but they are very brief. Thus, the focus of  $Ans_I$  is not totally on Gulf of Mexico oil spill. On the other hand, all paragraphs of  $Ans_E$  are about the oil spill event and detailed thoughts are provided. In order to support this argument, we posted  $Ans_E$  and  $Ans_I$  to Amazon Mechanical Turk [1] and

we reminded workers to focus more on the contents when comparing them. For this time, the average rating of  $Ans_E$  and  $Ans_I$  are 8.6 and 4.8 respectively ( $t(14) = 15.46, p < 0.00001$ ). There are 14 out of 15 workers agree that  $Ans_E$  should be rated higher than  $Ans_I$  and their reasons for rating  $Ans_I$  lower support our argument. Therefore, we conclude that  $Ans_E$  is actually a better outline than  $Ans_I$ .

Both the evolutionary approach and the independent approach offer 4-paragraph outlines while the answer of the map-reduce approach ( $Ans_{MR}$ ) is a 5-paragraph outline, which is slightly derailed from the 4-paragraph article objective. Hence, the average rating of  $Ans_{MR}$  is just 5.13. However,  $Ans_{MR}$  does contain relevant information for writing a 4-paragraph article on Gulf of Mexico oil spill. More work is needed for the requester to refine the outline so as to write a 4-paragraph article. Therefore, we conclude that  $Ans_E$  is better than  $Ans_{MR}$  in this case (average rating of  $Ans_E=8.5$  versus 5.13 for  $Ans_{MR}$ ,  $t(14) = 8.836, p < 0.00001$ ).

Also, we believe that we can regard the map-reduce approach as a kind of evolutionary approach, but with only 1 evolution cycle. In order to support such claim, we asked for feedback from the anonymous worker who offered  $Ans_{MR}$ . In her feedback, she mentions that she used two offered outlines and her memory of the event to write the new outline, which is the same as a 1-time evolution of the outline. Hence, it is no wonder that the evolutionary approach provides a better answer than the map-reduce approach as the evolutionary approach goes through more than 1 evolution cycle.

**Feedback from worker who offers  $Ans_{MR}$ :** I used the two outlines which were provided; there was a third supposed outline but it didn't have any information in it. .... I did add one item which was not already mentioned. I added a line about the need to comment on who was hurt in the explosion. That just came from my memory though, not an outside reference. I remembered that some people had died, so I thought that should be researched and included in the final report. ....

### Trend of Answer Quality

Figure 4 shows the moving average ratings of 3 consecutive answers of all approaches. We can see that the moving average rating of the evolutionary approach increases with time. On the contrary, there is no such trend in the moving average



ratings of the other two approaches. This shows that the evolutionary approach gradually improves the answer quality while the other two approaches do not. (the average rating of the 9<sup>th</sup> answer 8.5 versus the rating of the 5<sup>th</sup> answer 7.2,  $t(14) = 3.87, p = 0.0017$ ; the average rating of the 5<sup>th</sup> answer 7.2 versus the rating of the 1<sup>st</sup> answer 3.33,  $t(14) = 9.62, p < 0.00001$ .) Hence, we can conclude that the evolutionary approach can gradually improve answer quality and is an effective quality control method for [9].

## VII. DISCUSSION

In this section, we would like to share our experiences that we learn from this work.

### A. Limitations of Evolutionary Approach

In above case studies, we demonstrated that the evolutionary approach can incrementally improves answer quality. Here, we want to talk about the limitations of evolutionary approach. Firstly, the evolutionary approach is not useful for a task with exact answer or which is very easy. For example, if a requester asks workers to sum up all numbers in a picture, it is likely that all workers will give the same answer. Secondly, the number of workers involved in the task cannot be too small as evolution usually takes many "generations" to take effect. Finally, the evolutionary approach is also affected by the randomness of the performance of workers. The appearance of a few poor performance workers can affect the improvement progress. The valley (between Ans ID 10 and 14) in Figure 3 demonstrated this phenomenon.

### B. Why is it not a good idea to put all budget on a worker?

It is too risky to put all budget on a worker for an answer due to the randomness of crowdsourcing workers. If a worker offers a poor solution, which is not unusual in crowdsourcing, the requester has to issue another hit for a better solution using extra budget.

### C. How much to pay?

At the very early stage of our case studies, we tried to offer \$0.2/Hit for workers. After 3 days, we got around 7 answers for each approach. Nevertheless, the answer quality was extremely poor. Most of the answers that we got were spam, derailed, or very brief. As we expect workers to spend around 10 to 15 minutes on each hit, we believe that \$0.2/Hit is totally not a reasonable price (\$0.8 – \$1.2/hr). After that, we increased the offer to \$1/Hit (\$4 – \$6/hr), and as expected, we obtained higher quality answers though poor quality answers still occurred occasionally. Therefore, we believe that offering a reasonable hourly wage is the first step to get reliable answers.

## VIII. CONCLUSION

In this paper, we propose a new crowdsourcing approach that incrementally improves answer quality. We also proposed evolutionary feedback strategy which allows knowledge to be propagated among workers. Our new approach only requires requesters to define the task and objective and enter some

simple parameters so it can be easily incorporated into any existing crowdsourcing systems. Finally, we compare our approach with independent approach and map-reduce approach in [9] by performing two case studies-the 2-day travel planning and writing article outline. The result shows that our new approach can incrementally improve answer quality and produce high quality answers.

## REFERENCES

- [1] <https://www.mturk.com/>.
- [2] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell, and K. Panovich. Soylent: A word processor with a crowd inside. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 313–322, New York, NY, USA, 2010. ACM.
- [3] S.-K. Chang. Editorial: a general framework for slow intelligence systems. *International Journal of Software Engineering and Knowledge Engineering*, 20(1):1–15, 2010.
- [4] F. Colace, M. De Santo, and S. Ferrandino. Snmp-si: A network management tool based on slow intelligence system approach. In T.-h. Kim, T. Vasilakos, K. Sakurai, Y. Xiao, G. Zhao, and D. Izak, editors, *Communication and Networking*, volume 120 of *Communications in Computer and Information Science*, pages 83–92. Springer Berlin Heidelberg, 2010.
- [5] T. Dong. Modeling human intelligence as a slow intelligence system. In *DMS*, pages 41–46. Knowledge Systems Institute, 2010.
- [6] A. E. Eiben and M. Schoenauer. Evolutionary computing. *Information Processing Letters*, 82(1):1–6, 2002.
- [7] M. J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, and R. Xin. Crowddb: Answering queries with crowdsourcing. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, pages 61–72, New York, NY, USA, 2011. ACM.
- [8] X.-B. Hu and E. Di Paolo. An efficient genetic algorithm with uniform crossover for air traffic control. *Comput. Oper. Res.*, 36(1):245–259, Jan. 2009.
- [9] A. Kittur, B. Smus, S. Khamkar, and R. E. Kraut. Crowdforge: Crowdsourcing complex work. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 43–52, New York, NY, USA, 2011. ACM.
- [10] A. Kulkarni, M. Can, and B. Hartmann. Collaboratively crowdsourcing workflows with turkomatic. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 1003–1012, New York, NY, USA, 2012. ACM.
- [11] A. R. Lima, A. J. Cannon, and W. W. Hsieh. Nonlinear regression in environmental sciences by support vector machines combined with evolutionary strategy. *Comput. Geosci.*, 50:136–144, Jan. 2013.
- [12] A. G. Parameswaran, H. Garcia-Molina, H. Park, N. Polyzotis, A. Ramesh, and J. Widom. Crowdscreen: Algorithms for filtering data with humans. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD '12, pages 361–372, New York, NY, USA, 2012. ACM.
- [13] A. Sorokin and D. Forsyth. Utility data annotation with amazon mechanical turk. In *Computer Vision and Pattern Recognition Workshops*, 2008. CVPRW '08. *IEEE Computer Society Conference on*, pages 1–8, June 2008.
- [14] R. Szlapczynski and J. Szlapczynska. On evolutionary computing in multi-ship trajectory planning. *Applied Intelligence*, 37(2):155–174, Sept. 2012.
- [15] J. Wang, T. Kraska, M. J. Franklin, and J. Feng. Crowder: Crowdsourcing entity resolution. *Proc. VLDB Endow.*, 5(11):1483–1494, July 2012.
- [16] Y. Wang and S.-K. Chang. High dimensional feature selection via a slow intelligence approach. In *DMS'11*, pages 10–15, 2011.
- [17] T. Yan, V. Kumar, and D. Ganesan. Crowdsearch: Exploiting crowds for accurate real-time image search on mobile phones. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, MobiSys '10, pages 77–90, New York, NY, USA, 2010. ACM.
- [18] X.-J. Zhang, K.-Z. Chen, and X.-A. Feng. Material selection using an improved genetic algorithm for material design of components made of a multiphase material. *Materials and Design*, 29(5):972 – 981, 2008.

# Linked Open Graph: browsing multiple SPARQL entry points to build your own LOD views

Pierfrancesco Bellini, Paolo Nesi, Alessandro Venturi

Distributed Systems and Internet Technology Lab, DISIT Lab, Department of Information Engineering  
University of Florence, Florence, Italy, tel: +39-055-4796523, fax: +39-055-4797363  
<http://www.disit.dinfo.unifi.it>, [paolo.nesi@unifi.it](mailto:paolo.nesi@unifi.it), [pierfrancesco.bellini@unifi.it](mailto:pierfrancesco.bellini@unifi.it), <http://log.disit.org>

**Abstract** -- A number of accessible RDF stores are populating the linked open data world. The navigation on data reticular relationships is becoming every day more relevant. Several knowledge base present relevant links to common vocabularies while many others are going to be discovered increasing the reasoning capabilities of our knowledge base applications. In this paper, the Linked Open Graph, LOG, is presented. It is a web tool for collaborative browsing and navigation on multiple SPARQL entry points. The paper presented an overview of major problems to be addressed, a comparison with the state of the arts tools, and some details about the LOG graph computation to cope with high complexity of large Linked Open Data graphs. The LOG.disit.org tool is also presented by means of a set of examples involving multiple RDF stores and putting in evidence the new provided features and advantages using dbPedia, Getty, Europeana, Geonames, etc. The LOG tool is free to be used, and it has been adopted, developed and/or improved in multiple projects: such as ECLAP for social media cultural heritage, Sii-Mobility for smart city, and ICARO for cloud ontology analysis, OSIM for competence / knowledge mining and analysis.

**Keywords** LOD, LOD browsing, knowledge base browsing, SPARQL entry points.

## I. INTRODUCTION

The large publication of OD (open data) has opened the path for the information sharing. Most of the OD are published by governmental organizations, in file formats such as: html, xml, csv, shp, etc., and typically provide information that may present links to web resources. These links are typically coded as un-typed hyperlinks, URLs (Uniform Resource Locators). In 2006, Tim Berners-Lee published the LD (Linked Data) principles [1], as a model to stimulate the process of making accessible and sharing data as digital resources on the web and from them establishing links with semantically connected sources via URI (Uniform Resource Identifiers) [2]. On this wave, the data publication moved towards the diffusion of LD, opening the path for the construction of LD repositories and thus for creating a globally connected and distributed data space with integrated semantics. LD are based on documents formalized in RDF (Resource Description Framework) [3]. LD are mainly designed to be accessed and reused by machines. An RDF link leads to a triple putting in relationship two entities. For example, *Carl knows Paolo*, this consists of a subject, a predicate and an object or data value, which in turn are represented with URI. Thus, LD as triples can be accessible via specific LD Browsers, which allows to follow URI from one

data set to the model definition and/or to another dataset. Predicates, as “*knows*”, may be specified by using well-known vocabulary, such as the FOAF (Friend Of A Friend, [4]) that defines aspects and characteristics of people and their relations, and many others as mentioned in the sequel. A vocabulary defines the common characteristics of things belonging to classes and their relations. A vocabulary, also called ontology, is defined by using the RDFS (RDF Schema, RDF Vocabulary Description Language) or the OWL extension (Ontology Web Language). RDF triples can be stored in RDF stores (databases) and made accessible via SPARQL [17] entry point to pose semantic queries (SPARQL Protocol and RDF Query Language, recursive definition) on the RDF store. A network of SPARQL services and/or as LD/URI allows the creation of a network of LD, thus contributing to the construction of a global data model, which is the Linked Open Data, LOD [2].

In general, SPARQL queries are quite complex to be composed since their formalization strongly depend on the ontological structure of the RDF store model and the relationships among entities. This fact constraints the users to study the ontology in terms of entities and their relationships, also taking into account the external definition in terms of ontology segment, vocabulary, etc.

As an alternative, third parties LD search facilities based on keywords are also provided such as the semantic web crawler, such as Sindice.com [5]. Others solutions provide support to search on the semantic web via URI/LD. Other tools allow federating queries among multiple SPARQL entry points (RDF stores) have been proposed such as via Semantic Web Client Library [6]. This approach is typically applied for searching complementary aspects and composing the results in a unique semantic model.

Therefore, the complexity of accessing and using RDF stores and specifically LOD accessible via SPARQL entry point is limiting their usage. The understanding of LOD structure by using LD browsers is not an easy task and is also limited, since in most cases those browsers represent reticular relationships of LD with simplified tables and pages.

In the literature, to cope with the above mentioned problems, a large number of tools to edit and browse ontologies and knowledge bases have been proposed [7]. Most of them allow the editing of RDF stores and represent the entities by using hierarchical structures. A number of tools have been built on Protégé ontology editor [8]. Among the available tools, only

a few of them present a visual representation of the RDF store directly accessing to the SPARQL entry point. iSPARQL [9] is powerful tools that allow to access to an RDF repository via a SPAQRL query that can be visually represented and extended. On the other hand, the representation of results is still in tabular form and the navigation among relationships of the identified entities is very complex for who do not know the ontology structure. A number of tools for visual definition of SPARQL queries have been proposed, as Konduit [10], NITELIGHT [11].

Gruff application allows the visual composition of semantic queries grounded on Allegro Graph. Gruff generates the SPARQL query for accessing the entry point. The usage of Gruff should accelerate the learning of SPARQL language, while the complexity of usage is quite high. Gruff is a local application and includes capabilities for RDF storage browsing and analysis. A different approach has been taken by gFacet [12], which proposed a tool for posing interactive queries on a SPARQL entry point and obtaining interactive faceted results that can be used to refine the queries. Almost all the above mentioned tools are applications that need to be downloaded and installed. On the other hand, LodLive service is a web based RDF browser based on data graph representation (<http://lodlive.it>) [13]. It allows to access at LD and to single SPARQL endpoints. LodLive provides a user friendly user interface for browsing and navigation on the RDF entities starting from a specific URI. Once chosen the data sets and the URI, the representation of the accessed entity is based on circle surrounded by a number of small circles that can be accessed to expand the relationships.

On the other hand, none of the above mentioned LOD browsing data tools allow to fully exploit the nature of LD/LOD by expanding their rendering and navigation on multiple SPARQL entry points, and only LodLive is accessible via web, and present relevant limitations.

In this paper, Linked Open Graph, LOG, is presented. LOG.DISIT.org is a web based application for collaborative browsing and navigation into multiple SPARQL entry points (RDF stores). The LOG tool is web accessible and it is also in use to create the Social Graph in ECLAP social network as an embedded tool: <http://www.eclap.eu> [14].

The paper is structured as follows. In Section II, the main aspects of browsing into RDF stores are presented. Section III presents a comparative analysis of SPARQL visual browsers. It includes aspects to access and query, relationships among entities, general manipulation, URI details, and non-functional requirements. The comparison is used to put in evidence the main innovations of the proposed Linked Open Graph, LOG as: (i) management of multiple SPARQL entry points, (ii) saving and sharing of RDF graphs via web, (iii) learning and inspecting RDF graphs. Section IV presents some details about the computation of the LOG graphs and some larger and more complex example. Conclusions are drawn in Section V.

## II. RDF STORE AND EXTERNAL LINKS

The example reported in the introduction “*Carl knows Paolo*” consists of a subject, a predicate and an object or data value. These elements, in turn are represented by using URI.

The “*knows*” property may be defined to have as domain and range from class *foaf:Person* (from FOAF, [4]). Using this information, it can be inferred that both *Carl* and *Paolo* belong to the class *foaf:Person*. Moreover, the vocabulary states that class *foaf:Person* is a sub class of the more general class *foaf:Agent*, thus both *Carl* and *Paolo* belong to class *foaf:Agent*. The OWL version 2 language proposed by W3C allows defining disjunctive classes, union and intersection of classes, functional properties, symmetric, transitive properties, minimum and maximum cardinality of the associated elements of a property and other features. SPARQL language has been designed to query information on reticular structured information based on triples, and uses advanced algorithms to match portions of the RDF graph with a specified template. For example, the following query lists all the names of people that *Carl* knows indirectly through one or more other persons:

```
SELECT ?n WHERE {
  ?p1 foaf:mbox <mailto:carl@unifi.it>.
  ?p1 rdf:knows+ ?p2.
  ?p2 foaf:name ?n.
}
```

Different data sets may be defined by exploiting vocabularies (ontology segments) for defining properties and classes such as:

- *foaf:knows*, *foaf:Person* [4];
- OTN: [18] an ontology of traffic networks that is more or less a direct encoding of GDF (Geographic Data Files) in OWL;
- *dcterms*: [19] set of properties and classes maintained by the Dublin Core Metadata Initiative as *dc:title*;
- *vCard*: for a description of people and organizations [20];
- *wgs84\_pos*: vocabulary representing latitude and longitude, with the WGS84 Datum, of geo-objects [21].

Moreover, different RDF stores may be connected each other since they share common vocabulary or since one RDF store may refers with its links/URLs to other stores. Those links could be established after a process of data enrichment. For example, to connect the names of a well-known painter into a museum representation with the painter’s biography which is present on dbPedia [22]. On these bases, a number of SPARQL entry points to access at RDF stores are accessible such as: dbPedia [22], Europeana, LinkedGeoData, British Museum, Cultural Italia, Open Link LOD Cache, Linked Movie Data base, Getty vocabulary. A list of SPARQL end points can be found on <http://www.w3.org/wiki/SparqlEndpoints>. In addition, it is also possible to join two entities defined with different URI with a property *owl:sameAs*.

## III. ANALYSIS OF LOD GRAPH VISUAL BROWSERS

As described in the previous sections, the visual browsing of SPARQL entry points can be very useful for analysing the RDF store reticular structure, that is at the basis of the ontology and the related instances of predicates contained, the knowledge base. The users may use the LOD graphical viewers and browsers to (i) create RDF representations and models, (ii) save

them and share with other colleagues as a basis of discussion, (iii) learn about how SPARQL queries are created. On top of these SPARQL graph viewers, reasoners on different aspects can be provided, to make analysis about geographical and geometrical relationships, temporal relationships, etc. Moreover, different SPARQL versions provide limited capabilities in executing queries, such as problems counting elements, etc. Therefore, despite the first impression, the representation of an RDF reticular structure and thus its access are not a simple neither superficial task. The visual browsing of SPARQL entry points is not a simple task, especially if this work is performed by a Web Application. Thus, as described in the next pages, the LOG.disit.org service is not a simple browsing of related resources. In particular, specific algorithms are needed to cope with complexity of obtaining and processing complex reticular structures with web based applications, removing duplications, managing multiple entry points, generating complex SPARQL queries, etc.

In the following section, a number of demanded features and related problems are discussed with the aim of presenting LOG.DISIT features and comparing them with representative state of the art solutions: LodLive and Gruff. The identified features have been grouped in a few topics and discussed in different subsections: access and query, relationships versus entities, general manipulation, URI details, and non-functional features.

### III.A Access and Query LD and Stores

**Access and rendering of LD.** This means that the visual tool should be capable to represent a LD which is publically accessible as a URI, providing a set of triples. In Figure 1, the rendering of a URI<sup>1</sup> is depicted via LodLive [13]. The single URI is represented with a bubble, and the other small circles are links that can be clicked to expand the visualization to other LD/bubbles. Filled small circles are outbound links to LD, while unfilled small circles are inbound links coming from other LDs towards the former *URI(a)*.

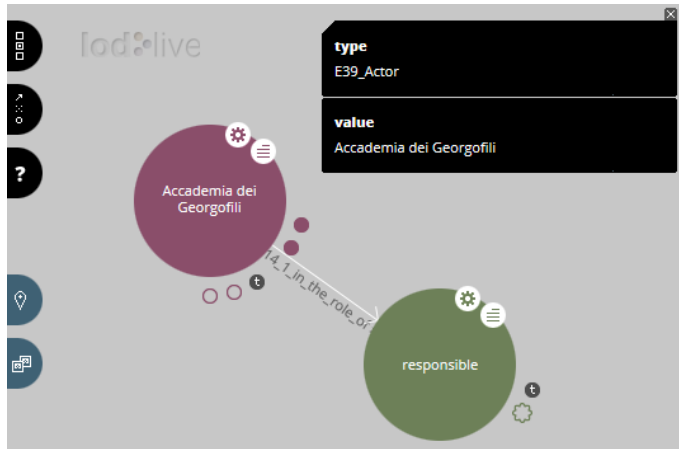


Figure 1: URI rendering with LodLive

These LD as RDF stores are accessible for the application or discovered by a semantic query to well-known SPARQL entry

<sup>1</sup> <http://dati.culturaitalia.it/resource/actor/accademia-dei-georgofili>

points. The small circles represent properties of the bubble, and some of them (presented in black) have coded letters as: “*t*” as types, “*s*” as *owl:sameAs*, “*b*” as blank nodes. The proposed rendering implies that the relationships of each single bubble/LD are not automatically explored. While their opening can be performed singularly and all together with a mouse click on the big bubble settings small icon.

#### Access and rendering URI from a SPARQL entry point.

A visual tool for browsing SPARQL entry points extract the results by using a couple  $\{ URL(i), Q \}$ , where  $Q$  is the semantic query or an URI. In this case, the tool needs to know both the SPARQL entry point of a given store  $a$  (an URL, that we can identify as  $URL(a)$ ), and at least a URI to be searched in the store (called here as  $URI(a)$ ) to get back the related description in terms of triples. In this case, the rendering of the triples can be similar to that of Figure 1 (representing the URI and the possible identified relationships recovered).

**Managing Entry Points with different URLs in URI.** In most cases, the  $URI(a)$  may have the same domain of their corresponding SPARQL entry point  $URL(a)$ , but it is not mandatory. And thus, the tools have to be capable to accept to start browsing from the couple URI, URL having different domains.

**Multiple SPARQL entry points.** The access and browse to a RDF store via the SPARQL entry point is a way to understand the knowledge base and the relationships among the included entities. In some cases, the entities/URIs ( $URI(a)$ ,  $URI(b)$ ) of different RDF stores (accessible via different SPARQL entry points:  $URL(a)$  and  $URL(b)$ ) may be connected each other. Typically, the connections can be via URI representing classes of common ontologies and definitions. The visualization of graphs associated with  $URL(a), URI(a)$  and  $URL(b), URI(b)$  on the same screen may allow to put in evidence the relationships among these two graphs. They may be the basis for (i) studying how to integrate different ontologies, for federating different RDF storages, (ii) understanding differences and relationships among different models, and/or (iii) for creating additional connections. For example, by creating an *owl:sameAs* relationship among two entities that represent the same concept in two models. In some cases, similar pattern have not been intentionally defined by using the same vocabulary since they are different for some aspect, while in other context they could be the same, otherwise deductions in the knowledge base would not take into account all needed facts.

**Making keyword based query.** In order to identify a starting URI for RDF graph rendering it could be possible to pose a keyword-based query on the RDF store. This feature is not always available on the RDF store (SPARQL entry points), and may be implemented in several different manners. Some implementation provide additional full text keyword based indexes on Lucene, other simpler solutions provide only substring search facility. The keyword based query is typically performed on all or specific ontology classes. Some of the tools allow selecting the specific class on which the keyword based query is performed. Both LodLive and LOG.DISIT provide this feature.

**Inspecting entry point for searching classes.** Once an entry point is identified, it is possible to pose queries to inspect

it to search for major classes. Thus, a textual search can be performed on the instances of one or more of those classes, in order to get back a list of entities/URI from which the graph visual browsing can start. This feature is quite difficult to use since the selection of the class(es) on which the search is performed may imply a certain knowledge about the ontology modelled in the RDF store of entry point.

### III.B Relationships among entities

**Showing relationships, turning them on/off, singularly and/or for category.** Once the first URI (the bubble in Figure 1) and related URIs are shown several relationships may be present in the graph, maybe hundreds or thousands, see Figure 2 from LOG.disit on dbPedia for URI related to Florence, Italy (it has been searched by a keyword based query on dbPedia SPARQL entry point, and thus it works when dbPedia entry is alive). Some of the relationships may be recursive (classes defined in term of their self); other are quite frequent and multiple, such as: *owl:sameAs*, *subject*, *type*. These should be marked or treated in different manner (as in the case of LodLive mentioned above). In any case, the users should be enabled to turn on/off some of the relationship categories to make the graph more readable and focussed on the entities and relationships under analysis. A LOD graph for an URI can present hundreds of different relationships kinds, and may be millions of triples to instances. Therefore, the usage of one line for each relationship kind is preferable to have a link for triple. The possibility of disabling relationship categories would shorten and simplify the analysis, as in LOG.DISIT.

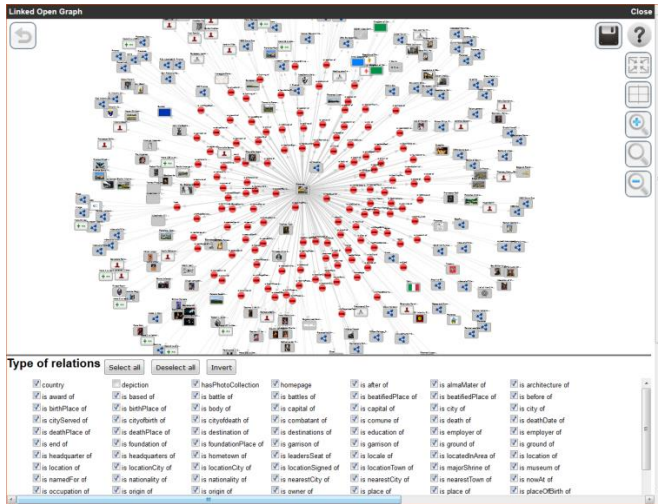


Figure 2: Florence URI on dbPedia, via LOG.DISIT, providing 364 elements: 237 entities and 127 multiple relationships (red circles).

Moreover, each category of relationship may bring to thousands or millions of entities (see Figure 3 for Gruff). For example, a library as Europeana has millions of elements, the civic number of a national street in Sii-Mobility may be thousands, see for example [http://SiiMobility.disit.org, http://servicemap.disit.org ]. This complexity has to be managed somehow, giving the possibility of accessing to a part of them for understanding the model, and to some specific relationships among the entities involved: for example by posing a specific query or faceting directly from each single

entity [12]. In some cases, the instances can be easily hidden from the graph disabling specific relationships of instance of.

**Representing relationships.** In the rendering of the RDF graph, a large number of entities (URI) and the relationships among them may be present. In most cases, the URI may have 1:N relationships that should be represented in different manner (some of them are very frequent such as *owl:sameAs*, *type*, or *those that bring to a blank node*). The high number of graphical elements can be reduced allowing closing/opening, expanding/compressing relations, filtering some relationships from the visualization (i.e., limiting the rendering to selected relationships) and may be also graphically representing entities and relationships by using coded styles.

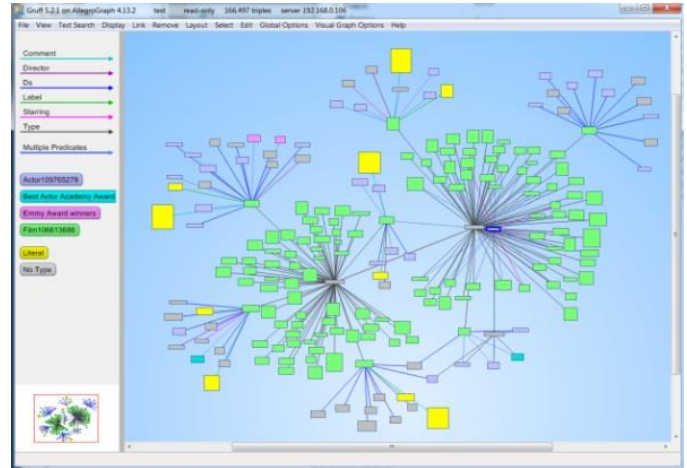


Figure 3: URI on a dbPedia segment, via Gruff.

On this regard, LodLive and Gruff assign a different colour to each URI according to their *type* (see Figure 3). When multiple types are present the colour can be determined by the first one, and thus the assignment may be misleading. In LodLive the color code is not constant so that at each graph reload the same graph may present totally different colors. A different approach could be to assign different icons according to their *type*, as in LOG.disit, and adopted in ECLAP social graph [14].

**Discovering inbound/outbound relationships, URI and queries.** At each URI a number of semantic queries can be associated with, for example, to recover the relationships:

(A) towards other entities (outbound, as *subject* in Gruff), it can be used:

```
SELECT ?object ?property WHERE {
  <http://dati.culturaitalia.it/resource/actor/accademia-dei-georgofili> ?property ?object.
  FILTER(isURI(?object))
}
```

(B) coming from other entities towards the former URI (inbound, also called as *object relationships* in Gruff), it can be used:

```
SELECT ?subject ?property WHERE { ?subject ?property
  <http://dati.culturaitalia.it/resource/role/isProducedBy>
  FILTER(isURI(?subject))
}
```



In some tool, the contextualized text of the query declined for a specific entity is accessible. It can be very useful for training the users in using the SPARQL and for shortening the data exploitation in external applications accessing to the SPARQL entry point API.

The inbound relationships can come from other SPARQL entry points, different from the one under inspection. This analysis implies to have a list of SPARQL entry points: as performed by LodLive and LOG.DISIT. In both cases, the list of accessible entry points for the tools is available for selection. Therefore, the set of SPARQL entry points allows for each URI to make this analysis, see Figure 4 for LOG.DISIT. The analysis allows counting the number of relationships in the different case, and for each of them to see sample the related query performed to get them. The query can be used to get all of them. In the case of Figure 4, 6 inbound and more than 5.6 million of outbound links have been found. In addition, also The British Museum is using that entity in about 40 million of triples, and dbPedia for more than 4.7 million etc. The discovered links can be opened to expand the browsing of the RDF graph to internal and/or external URIs, also belonging to other and multiple SPARQL entry points such as in LOG.DISIT.org, only, using multiple SPARQL rendering.

Thing	Relationship Type	Endpoint	Results	Action
inside	outbound	http://192.168.0.205:8080/openrdf-sesame/repositories/slimobilityultimate	6 results	View
inside	inbound	http://192.168.0.205:8080/openrdf-sesame/repositories/slimobilityultimate	5696829 results	View
endpoint	inbound	http://dbpedia-live.openlinksw.com/sparql/	4777083 results	View
endpoint	inbound	http://linkedgeodata.org/sparql	0 results	View
endpoint	inbound	http://europeana.ontotext.com/sparql	0 results	View
endpoint	inbound	http://collection.britishmuseum.org/sparql	40348775 results	View
endpoint	inbound	http://dati.culturaitalia.it/sparql/	41 results	View
endpoint	inbound	http://linkeddata.comune.fi.it:8080/sparql	0 results	View
endpoint	inbound	http://192.168.0.106:8080/openrdf-sesame/repositories/icaro7	0 results	View
endpoint	inbound	http://192.168.0.106:8080/openrdf-sesame/repositories/msptest2	6107 results	View
endpoint	inbound	http://openmind.disit.org:8080/openrdf-sesame/repositories/osim-rdf-store	832 results	View

Figure 4: Results of the relationships analysis for <http://www.w3.org/2002/07/owl#Thing> URI LOG.DISIT.org

**Discover paths between URI.** As a support to the analysis of the RDF graph, the identification of possible paths between two identified URI can be very useful. This analysis is a complex job to be performed in exhaustive manner for non-trivial cases, see for example the implementation of Gruff. Once identified the possible paths, the user would have to decide to see one or more of them according to some criteria.

**Creating triples/relationships.** An interesting feature that is moving towards the structural change of the RDF under analysis would be the insertion of new triples / new relationships, as in Gruff. This is possible in several RDF store editor and typically not available in browsers since the new triple should be stored somehow and would not be fine to store in a third party RDF store, even having the authorisation. Nevertheless, the creation of additional triples could be interesting to trial model integration among multiple SPARQL entry points.

A number of other features are also associated with the RDF relationships such as: the possibility of expanding and closing all the relationships, the possibility of counting the number of

relationships, and the special management of some of them *owl:sameAs*, links to blank nodes.

### III.C General Manipulation

**Undo of the actions performed, “back”.** The users may manipulate the LOD graph by means of several different actions such as opening/expanding URI and/or their relationships, turning on/off some relationship, etc. In the RDF visual graph manipulation, the possibly of undoing the actions performed with a back buttons may be very useful, together with the possibility of saving the reached status.

**Save and load LOD graphs.** The main aim of graph tools for visual browsing RDF stores is the construction of LOD graph rendering a situation for study and analysis entities and their relationships. The study of knowledge base as well as of ontology is frequently a long process in which several different navigations and openings are performed to explore relationships among the several entities/URIs. Therefore, a very valuable feature is the possibility of saving the status of the graph with all its linked URIs, and the relationships exploded (taking into account their on/off status). This graphical context should be the starting point for further analysis and not a simple image snapshot. Once saved the RDF graph analysis, it could be useful to be reloaded for further elaboration, and/or for sharing it with other colleagues in read or read/write modalities, thus enabling the collaborative work on the same RDF graph analysis. Only Gruff and LOG.DISIT allow saving and loading RDF graphs.

**Share and collaborate on LOD graphs.** The RDF graph sharing is based on saving the RDF graph on some cloud to provide the possibility to share it to other colleagues to make changes or simple access at the graph via web. Among the tool analysed, only LOG.DISIT provide this collaborative feature on LOD RDF graphs. LOG.DISIT allows to share the RDF graph as web data on the cloud, in read and read/write modalities.

**Export of RDF graph triples.** The export of the RDF entities involved in the visualized/pruned RDF graph can be a very useful feature for study the model in other tools.

A number of other features are also associated with the general manipulation of the visual graph such as: Re-laying out the graph the screen rearranging automatically the graphical elements, focusing on an URI (identifying an URI and restarting the navigation from that point), zooming and panning the graph, centering the graph (moving in the center of the graph the original URI).

### III.D URI Details

**URI attributes.** A number of attributes/values (literal) may be associated with the URI. These data should be accessible without involving graph representation. To this end, a simple table with a list of values can be provided as in LodLive and LOG.DISIT. Among the possible values, the GPS coordinates could be used for positioning the URI on geo-MAP (**Map allocation of URI**).

**URL to resources.** An URI in the LOD graph is the representation of an RDF entity in the store. On the other hand, the original data can be opened in the browser. Moreover, a URI may have among its attributes some URL to external digital resources. These URL should be accessible for opening



the digital resources into the browser or for download. They can be files, such as: images, video, documents, web pages, etc. In these cases, it can be useful to have the possibility to directly **Open play resources**.

**Representing entities.** In complex LOD graph the fast identification of URI type is very important. Not all the URIs have a relationships with an URI formalizing its type, and it is not rare to see an URI with multiple types. The URI can be represented by using specific icons on the basis of their: (i) type (problems in the case of multiple types), (ii) information and attributes (such as some connected image), (iii) specific icon associated with the URI (e.g., image of the person for *dc:author*), (iv) specific case, for example to represent the *Blank* nodes.

### III.E Summary of comparison

Table 1 reports the summary of the performed comparative analysis of Section III.

TABLE 1: SUMMARY OF COMPARATIVE ANALYSIS

	LOG	LodLive	Gruff
<b>Access and Query</b>			
Access and rendering of LD	Y	Y	N
Access and rendering URI from SPARQL entry point	Y	Y	Y
Managing Entry Points with different URL in URI.	Y	N	Y
Multiple SPARQL entry points	Y(10)	N	N
Making keyword based query	Y	Y	Y
Inspecting entry point for searching classes	Y	Y	Y
<b>Relationships vs entities</b>			
Showing relationships, turning on/off, singularly or globally	Y(3)	Y(2)	Y(2)
Representing relationships (managing complexity)	Y	Y(4)	Y(4)
Discovering inbound/outbound relationships, URI and queries	Y	Y	Y(7)
Discover paths between URI	N	N	Y
Creating triples/relationships	N	N	Y
Expand all relationships	Y	Y	N
Close all relationships	Y	N	N
Counting number of elements	Y	Y	Y
"sameAs" management	Y	Y	Y
Blank nodes rendering	Y	Y	Y
<b>General Manipulation</b>			
Undo actions performed, "back"	Y	N	Y
Save and Load LOD graphs	Y	N	(Y)
Share and collaborative LOD graphs	Y	N	N
Export of RDF graph triples	N	N	N
Re-laying out the graph	Y(6)	N	Y
Focusing on an URI	Y	Y	N
Zooming the graph	Y	N	Y(8)
Centering the graph	Y	N	N
Panning the graph with mouse/finger	Y	Y	Y
<b>URI Details</b>			
URI attributes (showing info or an URI)	Y	Y	Y(1)
Map allocation of URI	Y(9)	Y(9)	N
URL to resources	Y	Y	N
Open play resources	Y	Y	Y
Representing entities	Y	Y(5)	Y(5)
<b>Non Functional</b>			
Web based tool	Y	Y	N
Embed in web pages of third party service: ECLAP	Y	N	N
Graph Invoked by URL	Y(7)	N	N

1. Gruff presents literal attributes of URI as graph nodes, while LodLive uses a single aside panel, and LOG multiple frames, thus making simpler the comparison among nodes.
2. In Gruff: single and multiple links can be off at the same time, limited capability in tuning on all links of the same kind in the graph. In LodLive, links can be singularly turned on/off. The complexity is not managed.
3. In LOG, multiple links on/off of the same kind
4. LodLive and Gruff allow opening all or singularly, no middle way or precise control. LodLive presents a limited number of elements in some cases, and does not inform the user about the applied limitation.
5. LodLive and Gruff adopt different colours for representing different type of entities, and not icons.
6. In LOG.DISIT, the positioning of the entities and relationships is dynamically performed on the basis of a force model, in some case, this can be confusing.
7. Gruff provide support to discover inbound/outbound links (as object/as subject) only taking into account the current RDF store. LOG and LodLive perform the query on a range of SPARQL entry points (at their disposal in some database), while others can be added.
8. Gruff has a powerful zoom and large graph management; on the other hand, it is a standalone application in native code.
9. LodLive provide direct support for placing on a Map the URI if they present GPS coordinates. Integration with Map can be performed for LOG since the LOG graphs can be opened and recalled by an REST call / URL. See for example the [Http://servicemap.disit.org](http://servicemap.disit.org).
10. LOG allow the loading of multiple SPARQL entry points and the web sharing of LOG graph, by sending emails with the links to reload and manipulate them

### IV. LOG.DISIT.ORG COMPUTING

As described in Section III, the Linked Open Graph, LOG.Disit.org, allows opening multiple reticular RDF representations starting from different URIs (also called graph *root*) of different SPARQL entry points. All the starting URIs/URLs loaded are also listed on top of the LOG user interface. The listed URL/URI can be clicked to highlight the corresponding root URI.

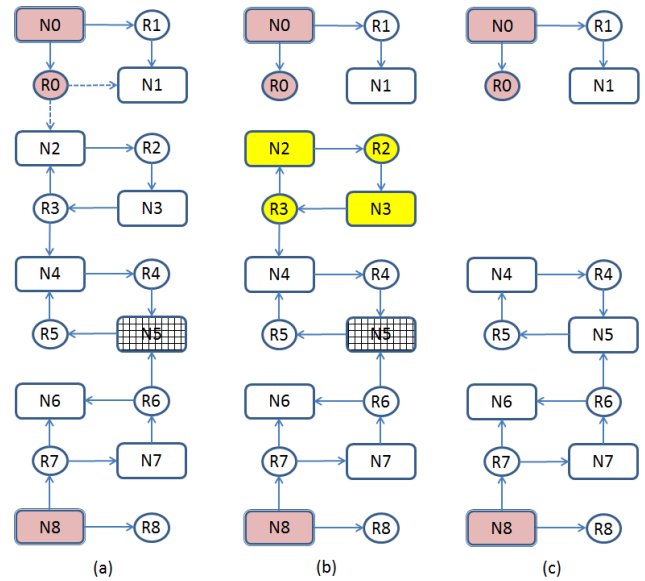


Figure 5: Graph reduction process in LOG

In LOG graph reported in Figure 5 an algorithmic aspects related to multiple entry points is discussed. In Figure 5a, the 1:N relationships (as R0, R1, ...) are represented with a unique arc exiting from the sourcing node, N0. Among the visual

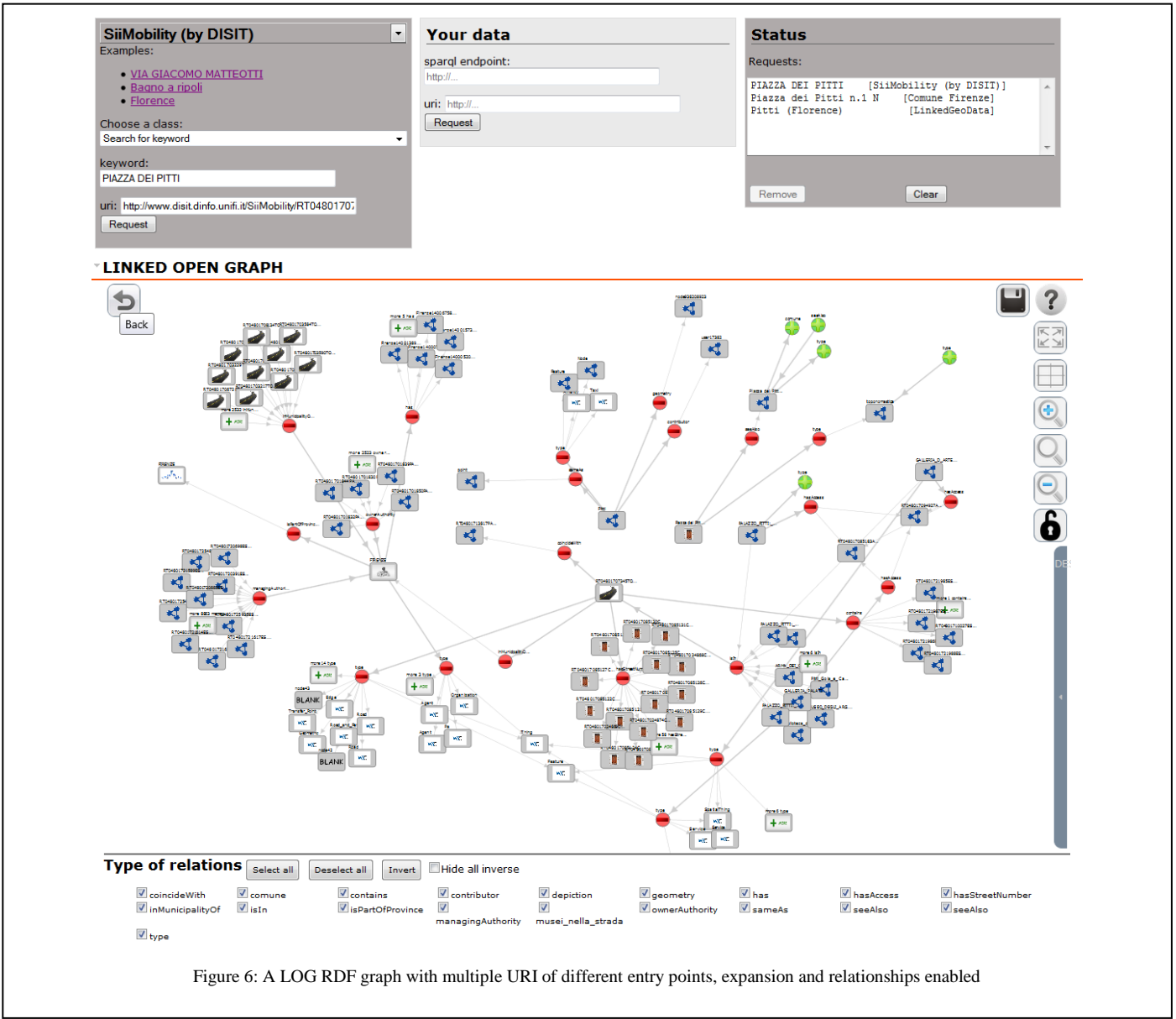
browsers analysed in the state of the art, LOG.disit is the only one managing multiple SPARQL entry points and allowing the web collaboration. Circles, as R0, represents the relationship and manages the multiplicity (for example towards N1 and N2). This approach (adopted in LOG to have only one line exiting from the entity per relationship kind), allows managing the complexity of large data sets. On the other hand, it computation adds an additional complexity in LOG drawing where multiple roots may be present.

In Figure 5a, a LOG case with two roots is presented: N0 and N8; the two roots share node N5 that holds a double multiplicity (belonging to two graphs). When the user closes R0, with a double click: the 2 relationships related arcs dotted in Figure 5a are deleted. According to that action, a graph analysis is needed. The analysis is started by performing a labelling process from both roots N0 and N8. This allows identifying all nodes that are connected from some root (all except N2, N3) in the graph. Thus, the elements which are not connected have to be removed (see Figure 5b), for example: N2, N3, R3 and R2. In addition, shared nodes, such as node N5 lose their

multiplicity. Figure 5c represents the final results after the application of the above described “closure” algorithm, where it is evident that some elements passed from one root to the other. A complementary operation is performed, when an inbound link of a node is opened (for example by using a query similar to that obtained in Figure 4), for example, N3 request the opening of R3, then a situation similar to Figure 5b can be reached.

In most cases, the removal of elements does not means to delete the elements from the internal graph model, but only its hidden from the graph. This approach allow to pass from less to more details in a very fast manner, but at the expenses of the loading time when data are collected form the remote RDF stores

With the **contextual menu** on the node/URI, the user may perform the analysis of the inbound and outbound relationships, or explore all the relationships (see Figure 4). Thus, in the browsing and construction of an LOG RDF graph, a number of progressive queries are performed. The graph is constructed on the basis of the resulting triples obtained from those queries: (i) some of the resulting relationships and URIs (nodes) could be



already present in the graph; (ii) a node may have multiple arcs entering and exiting to/from a node. They do not have to be drawn more than once; the duplications have to be avoided by using an efficient algorithm on the data model since real time rendering is needed. Thus, the algorithm verifies every new arc to check if it is already included or not; duplicated arcs are removed from the model. Then, nodes without arcs are also removed. The graph cleaning has to be performed every time nodes/URI and relationships/arcs are added or removed.

From the technical point of view, LOG.disit provides a server side application in PHP and exploit on client side: Javascript, JQuery, Ajax, and D3 graphic library [15].

When the graph related to a URI needs to be created the server side script needs to retrieve from the SPARQL endpoint the information to depict the node: the type, the label, when available the *foaf:depiction* image, the predicates that are associated with the node and for all the nodes that are to be represented the type, label and depiction image. To this end, the server-side script performs the following numbered steps:

1. it is requested the *rdf:type*, *rdfs:label* and *foaf:depiction* associated with the URI;
2. it is requested for the URI the number of occurrences of each predicate using a query such as:

```
SELECT ?p (count(*) as ?c) WHERE {
  <URI> ?p ?o.
  FILTER isURI(?o)
} GROUP BY ?p
```

This is performed to have an idea of the complexity of the relations with other nodes, it can happen that a node has thousands of associations with other nodes and in this case a query that gets all the triples where the URI is the subject can be unmanageable.

3. For the predicates that are not too numerous it is requested the information of the related nodes and predicates with a query as the following:

```
SELECT ?p ?o ?l ?t ?d WHERE {
  <URI> ?p ?o.
  OPTIONAL { ?o rdfs:label ?l }
  OPTIONAL { ?o rdf:type ?t }
  OPTIONAL { ?o foaf:depiction ?d }
  FILTER !(?p IN (<...>, <...>)) }
```

4. for each predicate <P-URI> that is too numerous a specific query is performed such as the following:

```
SELECT ?o ?l ?t ?d WHERE {
  <URI> <P-URI> ?o.
  OPTIONAL { ?o rdfs:label ?l }
  OPTIONAL { ?o rdf:type ?t }
  OPTIONAL { ?o foaf:depiction ?d } }
```

to retrieve information about related nodes.

5. the same operations of steps 2, 3 and 4 are performed using the <URI> as object and not as subject of the predicate.

A special case is the one related to *blank* nodes, generally identifiers used to refer to them are valid only for the specific document that contains them and thus these identifiers cannot be used in later queries to get information about the specific

blank node. Moreover, if a blank node is used in a SPARQL query it is treated as a variable matching nodes. Some RDF store solve this problem with specific extensions that are not standard and thus are difficult to be used in this context. To partially solve this problem we decided to retrieve for blank nodes also all the relations of the blank node with other nodes and send all this information to the client that needs to manage its access. This operation is limited since in case the blank node refers to another blank node this one cannot be explored. This problem may be solved in a future version using information from linked data that should contain all the blank nodes used to represent a resource.

For this reason the query used in the third step is changed to:

```
SELECT ?p ?bnode ?p2 ?o ?l ?t ?d WHERE {
  { <URI> ?p ?o. FILTER isURI(?o)
  } UNION {
    <URI> ?p ?bnode.
    ?bnode ?p2 ?o.
    FILTER isBlank(?bnode) && isURI(?o)
  }
  OPTIONAL { ?o rdfs:label ?l }
  OPTIONAL { ?o rdf:type ?t }
  OPTIONAL { ?o foaf:depiction ?d }
  FILTER !(?p IN (<...>, <...>, ...))
}
```

that makes a union of the results where the URI is associated with another URI and when the URI is associated through a blank node.

#### IV.1 LOG usage and example s

Technically, not all ontologies and RDF models and stores have been developed by using the same methods since they have been developed by different teams, using different styles, in different periods, and exploiting different vocabularies. This implies that different approaches to model the same entities and patterns may be possible, as well as different usage of “*sameAs*”, “*equivalent class*”, *blank* nodes, reuse of vocabulary and concepts, etc. The LOG can be very useful to understand these differences interactively studying the RDF store from remote, to learn and to explore the possibility of reusing and connecting them each other. The LOG.disit tool, with its additional features with respect to the state of the art browsing tools, can be a very useful tool for: analyzing RDF stores and models, comparing and discovering connections and relationships among RDF stores and models, discovering eventual problems in accessible knowledge base for their future reuse and connection.

In Figure 6, an example is presented. The upper part of the screen reports the controls and the list of roots URIs included and loaded in the graph. They have been obtained from: LOD of Florence, Sii-Mobility and LinkedGeoData The resulting LOG graph reported in the Figure 6 can be accessed (in read only mode) by using <http://log.disit.org/service?graph=3dfae71db76642b6ba23ce7dc7cb12bcf>, while the URL for modifying the LOG graph has been sent to the email of the LOG graph creator only, that could decide to share. On the bottom part of the screen, the list of active relationships is reported. They can be turned on/off and the whole section inverted. After to have loaded the first URI

(Pitti from Linked Geo data) the user discovered relationships (similarly to Figure 4) then decided to open the first URI related to Pitti, and worked a bit on some aspects to browse relationships. Then the decision of searching for Pitti in different SPARQL entry points (Sii-Mobility and Comune di Firenze) provoked the load of the corresponding nodes. Then a number of other nodes have been browsed with the aim of comparing the three different representations of the same entity discovering other similarities (sadly of unconnected entities) as Florence, and related streets. This process helped the user to conquer a global and integrated comparison of the aspects associated to the same topic in multiple RDF stores.

Concrete examples have to be contextualized with respect to the RDF store directly suggested in the LOG interface as follows. In the following other example of LOG.disit usage with connected and specialized graphs are reported.

**ECLAP** RDF store contains information about content and users of the ECLAP social network (<http://www.eclap.eu>). In this case, the LOG could be used to (i) compare the user profile graphs of different users, (ii) discover direct and indirect relationships among users by searching and calling their entry points, (iii) exploring relationships of a single user among its several connections with other social network actions and elements. The analysis can be focused on producing new metrics, new suggestions, and identifying new cause – effect relationships. The ECLAP model, via users and content are also connected to dbPedia and Geoname. In Figure 7, a study about the indirect relationships among two different users is reported. Some of the possible relationships have been disabled to focus on common favorite content and friendships.

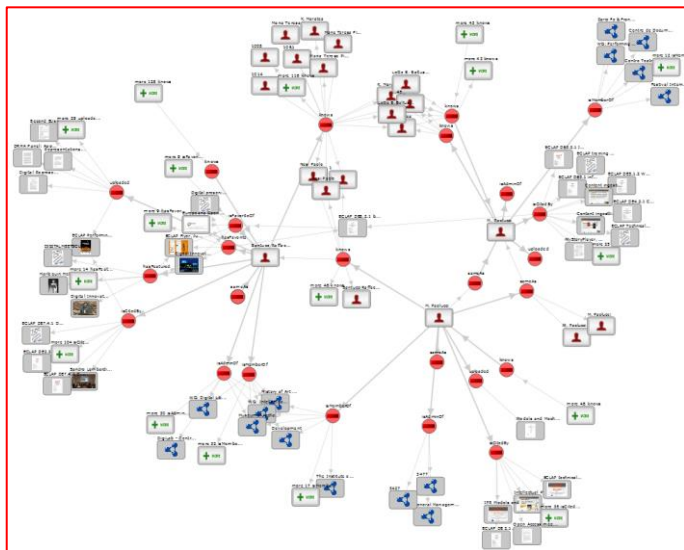


Figure 7: A LOG RDF, indirect relationships of two different users on ECLAP.

**OSIM** RDF Store contains a model and data related to the University of Florence knowledge, including all structures, research lab, researchers, their publications, relationships among them, related competences of people and structures and thus a taxonomy of concepts and competences. In this case, the LOG can be useful to browse and analyze the network of experts that are working on a given topic, their relationships, the places in which they have published, the projects in which

they worked, and who worked on what. The browsing of the store allows extracting more information than the simple semantic query on the user interface. There are some connections among users of ECLAP and the OSIM store since some of the users are also modeled in the OSIM store. The usage of multiple RDF stores allows to understand how these stores could be used to create new knowledge and services. For example, learning preferences on ECLAP and providing suggestions on OSIM or viceversa. In Figure 8, a LOG graph analysing connection and structures of the same user on ECLAP and OSIM RDF stores is presented.

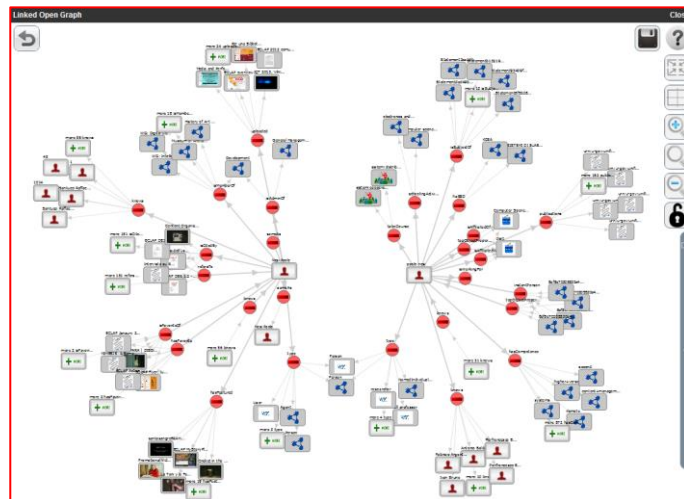


Figure 8: A LOG RDF graph analysing connection and structures of the same user on ECLAP and OSIM RDF stores.

**Senato** and **Camera** RDF stores contains the information related to laws and political decisions by the Italian govern, and thus also the involvements of the politicians. The two stores are not physically connected while relationships are evident in terms of laws, politicians, approvals of documents and their passages and demands from one camera to the other (the famous disputed Italian perfect bicameralism). In this case, the activation of the one URI in a store may really link to information in the other, and the complete view can be obtained only by a tool as LOG.disit.org, that allows you to join them together. Another interesting analysis can be performed to see the votes of politicians during their political life and the support they gave to different political groups and laws.

**Sii-Mobility** RDF Store models a large repository of geolocalized data regarding Smart City concepts and data connected to Tuscany: topographic information, administration, services, statistics, time line of busses, parking status, weather forecast. In this case, the LOG tools is very useful in the hands of potential SME interested in developing mobile applications during *Hackathon* for the definition of innovative Smart City services. For example, to (i) discover and understand the model and the information associated to a given service in the city, (ii) discover connections and similarities among different open data set of public administration, (iii) study the integration of open data with geographic information. In this particular case, Sii-Mobility provides an user interface to perform geographic queries and from the results the LOG graph can be open (<http://servicemap.disit.org>).

## V.CONCLUSIONS

The navigation on internet accessible RDF stores is becoming every day more relevant. They are frequently based on local and commonly accepted ontologies and vocabularies to set up large knowledge base to solve specific problems of modelling and reasoning. The growing needs of such structures increased the need of having flexible and accessible tools for RDF store browsing taking into account multiple SPARQL entry points to create and analyse reticular structure and scenarios of remote stores. The LOG tool presented in this article provides innovative features solving a number of problems related to graph computation to cope with high complexity of large LOD graphs with a web based tool. The complexity is mainly managed by providing tools for (i) progressive browsing of the graphs, (ii) allowing graph composition, (iii) providing support to pose specific queries, (iv) allowing the progressive discovering/selection of instances. A comparative analysis with reference solutions at the state of the art has been also provided, showing that LOG presents a number of innovative and very useful features for RDF store analysis and development. In general, RDF stores have been developed by using different methods, by different teams, using different styles, in different periods, and exploiting different vocabularies. The Linked Open Graph, LOG, is a web based tool for collaborative analysis, browsing and navigation on multiple SPARQL entry points. The LOG.disit tool, with its additional features with respect to the state of the art browsing tools, can be very useful to understand these differences interactively studying the RDF store from remote, to learn and to explore the possibility of reusing and connecting them each other.

The LOG tool is used in multiple projects as ECLAP for cultural heritage (<http://www.eclap.eu>), Sii-Mobility for smart city [16] and ICARO for smart cloud ontology analysis. It has been validated using multiple public accessible RDF stores such as: dbPedia, Europeana, Getty Vocabulary, Camera and Senato, GeoLocation, etc., putting in evidence the different cases and usage of LOG tools in the different scenarios, with a specific stress on the analysis of multiple RDF stores on the same graph.

## BIBLIOGRAPHY

- [1] T. Berners-Lee, "Linked Data", <http://www.w3.org/DesignIssues/LinkedData.html>, 2006.
- [2] C. Bizer, T. Heath and T. Berners-Lee (2009) Linked Data - the story so far. *Int. Journal on Semantic Web and Information Systems*, 5, (3), 1-22.
- [3] G. Klyne, J. Carroll, "Resource Description Framework (RDF): Concepts and Abstract Syntax - W3C Recommendation", 2004
- [4] FOAF, <http://www.foaf-project.org/>
- [5] G. Tummarello, R. Delbru, and E. Oren. 2007. Sindice.com: weaving the open linked data. In *Proc. of ISWC'07/ASWC'07*, Springer, Berlin, Heidelberg, pp.552-565.
- [6] O. Hartig, C. Bizer, J.-C. Freytag. 2009. Executing SPARQL Queries over the Web of Linked Data. In *Proc. of ISWC '09*, Springer, pp.293-309.
- [7] S. Ramakrishnan and A. Vijayan. 2014. A study on development of cognitive support features in recent ontology visualization tools. *Artif. Intell. Rev.* 41, 4 (April 2014), pp.595-623.
- [8] Protégé <http://protege.stanford.edu/>
- [9] iSPARQL, <http://oat.openlinksw.com/isparql/index.html>,

- [10] O. Ambrus, K. Moller, S. Handschuh, "Konduit VQB: a Visual Query Builder for SPARQL on the Social Semantic Desktop", *proc of VISSW2010, IUI2010, 2010, Hong Kong, China.*
- [11] A. Russell, P.R. Smart, D. Braines, Dave, N.R. Shadbolt, "NITELIGHT: A Graphical Tool for Semantic Query Construction", In, *SWUI 2008*, Florence, Italy,
- [12] Gfacet, <http://www.visualdataweb.org/gfacet.php>
- [13] D. V. Camarda, S. Mazzini, A. Antonuccio. 2012. LodLive, exploring the web of data. In *Proc. of the I-SEMANTICS '12*, ACM, pp.197-200. <http://lodlive.it>
- [14] P. Bellini, P. Nesi, "Modeling Performing Arts Metadata and Relationships in Content Service for Institutions", *Multimedia Systems Journal*, Springer, 2014. <http://www.eclap.eu>
- [15] D3, Data-Driven Documents, <http://d3js.org/>
- [16] P. Bellini, P. Nesi, N. Rauch, "Smart City data via LOD/LOG Service", *Workshop Linked Open Data: where are we?, LOD2014*, org. by W3C.
- [17] Prud'hommeaux, E., Seaborne, A., SPARQL Query Language for RDF, <http://www.w3.org/TR/2004/WD-rdf-sparql-query-20041012/>
- [18] OTN, Ontology of Transportation Networks, Deliverable A1-D4, Project REVERSE, 2005  
<http://reverse.net/deliverables/m18/a1-d4.pdf>
- [19] <http://dublincore.org>, <http://dublincore.org/documents/dcmi-terms/>
- [20] VCARD, <http://www.w3.org/TR/vcard-rdf/>
- [21] wgs84, [http://www.w3.org/2003/01/geo/wgs84\\_pos](http://www.w3.org/2003/01/geo/wgs84_pos)
- [22] dbPedia, <http://dbpedia.org/resource/>



# Creation and Use of Service-based Distributed Interactive Workspaces

Carmelo Ardito<sup>1</sup>, Paolo Bottoni<sup>2</sup>, Maria Francesca Costabile<sup>1</sup>, Giuseppe Desolda<sup>1</sup>, Maristella Matera<sup>3</sup>, Matteo Picozzi<sup>3</sup>

<sup>1</sup>Dipartimento di Informatica, Università degli Studi di Bari Aldo Moro

Via Orabona, 4 – 70125 – Bari, Italy

{carmelo.ardito, maria.costabile, giuseppe.desolda}@uniba.it

<sup>2</sup>Dipartimento di Informatica, Sapienza Università di Roma

Viale Regina Elena, 295 – 00161 – Roma

bottoni@di.uniroma1.it

<sup>3</sup>Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano

P.zza L. da Vinci, 32 – 201233 – Milano

{maristella.matera, matteo.picozzi}@polimi.it

**Abstract**—Distributed Interactive Workspaces (DIWs) are interactive environments, accessible through different devices, where end users create new content by exploring and aggregating data retrieved from distributed resources in the Web, tailor this content to their own personal needs, use it on different devices, and possibly share and co-create it with others. The need for collaborating with other people by means of DIWs is an important requirement that emerged in field studies conducted in different domains. This paper shows the extension of a platform for mashup composition to support collaboration through DIWs. In particular, it considers the possibility of producing annotated versions of DIWs, to add specific information and make it available to others without corrupting the original resources. It also investigates techniques for asynchronous collaboration that enable a distributed execution of the created interactive workspaces on different devices and by different users.

**Keywords** — *Distributed Interactive Workspaces, Collaboration, Human-Centric Service Composition.*

## I. INTRODUCTION

Web 2.0 has accelerated the evolution of the Web, becoming a driver for innovation. End users are now involved in the content creation process, a fact which has in turn amplified their will to become active creators of applications that simplify access to the huge quantity of data made available on the Web in heterogeneous formats and protocols.

The mashup phenomenon has been one of the results of this trend towards a “democratic” access to online resources. Mashups integrate heterogeneous services at different layers of the application stack, to provide unified views over integrated result sets fetched from different sources [1-3]. So far, mashups have been especially conceived as personal information spaces, i.e., vertical applications solving situational needs, that end users assemble by merging ready-to-use resources [4]. Mashups, however, have a great potential to accommodate the sharing and co-creation of knowledge [5]. As highlighted by the field studies discussed in this paper, in several domains the involved stakeholders need to share, co-create and execute mashup in a distributed manner. Nevertheless, while collaboration mechanisms have been extensively investigated

in different areas, the co-creation of interactive workspaces via mashup composition is still scarcely explored.

The work reported in this paper relates to the experience gained in the last few years in the analysis of different paradigms for mashup composition and in experiments with prototypes of a mashup platform [4, 6]. The platform exploits End-User Development (EUD) principles and offers a visual, live programming paradigm to let users, not necessarily technology experts, create service-based, interactive Web applications. The novel contribution of this paper is to enable the collaborative creation and use of *Distributed Interactive Workspaces* (DIWs). DIWs are component-based interactive applications, where content is produced by end users via the aggregation and manipulation of data fetched from distributed online resources, both local and third-party. DIWs can be deployed as personal applications through a client-side logic supporting the execution of the workspaces on multiple devices. DIWs can also exploit a centralized, server-side, execution logic to manage the sharing of workspaces among different users, the propagation of collaborative actions to active instances of a same workspace, and the distributed execution of a whole workspace, or of selected components, on different devices employed by different users. The peculiarity of the presented approach is that collaboration mechanisms can be applied to different elements of the mashup application: from the basic services the mashup raw data are fetched from, through the integrated content resulting from the adopted data integration policy, to the integrated visualizations that in our approach guide the composition process.

As a further contribution, while recent works proposing some form of collaboration in mashup composition only cover specific, limited aspects (e.g., awareness in synchronous editing) [7], this paper shows how collaboration can be supported in different modalities. In particular, it discusses how collaboration for DIW co-creation can be described along two dimensions: 1) the time at which it takes place, and 2) the resources, of a physical or computational nature, used by participants. Concerning the temporal aspect, this paper discusses both synchronous and asynchronous collaboration, among all participants or selected subsets thereof. As for



resources, the paper focuses on the different elements of a DIW, namely services, integrated data sets, and visualizations, and shows how they can be co-created during the collaboration process itself, or be produced or retrieved by individual participants working on their own. Although the two aspects are orthogonal, in the sense that systems and procedures can accommodate any combination of them, some problems may arise concerning the use of individual resources during collaborative work: How to smoothly integrate them in synchronous sessions? How to asynchronously communicate the availability of new resources to collaborators? How to protect resources intended only for personal usage, and how to change their status to public?

This paper shows how such questions have been addressed by integrating a general collaboration process in the creation and management of DIWs. The need for collaboration to co-create and share DIWs emerged from some formative studies in different application domains. Thus, the composition platform illustrated in [4] has been extended to enable annotation and co-creation of service-based interactive workspaces. The paper describes how it is possible to augment the available resources with collaboration-oriented information, and how to make this information available to others without corrupting the original resources. Synchronous collaboration, mainly based on live editing, is also addressed to allow multiple users to interactively modify (portions of) a shared DIW, being *aware* of the modifications operated by any participating collaborator.

The paper proceeds as follows. Section II summarizes the formative studies and the collaboration requirements that emerged. Section III introduces the collaborative mechanisms that can be applied on DIWs, while Section IV describes the platform architecture, with emphasis on the new modules supporting sharing, co-creation and distributed execution of service-based interactive workspaces. Section V reports on related work and Section VI concludes the paper.

## II. MOTIVATION FOR COLLABORATIVE AND DISTRIBUTED CREATION OF INTERACTIVE WORKSPACES

The mashup platform described in [4] allows end users to create interactive workspaces by composing heterogeneous data sources, be they public (e.g., remote resources available in the form of Web services and APIs) or private (e.g., local content for personal use). The platform supports the integration of data into *UI components* [6, 8], i.e., widgets that retrieve data from different services and display their integration into a unified user interface (UI). Given a set of distributed services registered into the platform, the user can define parametric, key-value queries through visual forms. The retrieved result sets can then be integrated within UI components. A number of *visual templates* supply possible UIs. Besides serving the rendering of the integrated data, visual templates also provide a unified schema for lightweight data mediation: a visual mapping process allows the user to select data items returned by the selected services and associate such items with visual elements playing the role of data collectors in a visual template. The association of data from multiple services with single UI data collectors determines the definition of *union* and *merge* operations on the involved data sets [6].

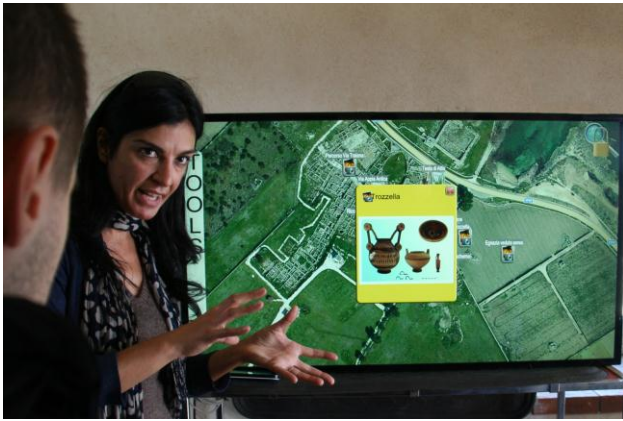
The created components can then be included within the workspace under definition, where they can be also synchronized according to an *event-driven, publish-subscribe* paradigm, which enables synchronization of *components'* behaviors. Each component exposes *events* and *operations* [1]. The coupling of components within a workspace is thus based on the subscription of *operations*, which become *listeners* for *events* exposed by other components. Invoking an operation changes the state of the interested components.

With respect to other proposals for mashup composition, the peculiarity of this platform lies in its visual composition mechanisms, which make it adequate for end-user development [8]. Due to the extensive use of visual representations guiding the composition process and to the separation between such mechanisms and the logic for mashup composition and execution, the platform is easily customizable with respect to specific user requirements and characteristics. As discussed in [4], customization mainly requires the adoption of *visual templates* offering adequate visual metaphors to the users.

In the next two subsections we report on two field studies we performed in different application domains in order to verify the usefulness for end users of content made available by distributed data sources, as well as the overall validity of our composition approach. The studies were also useful to identify improvements and extensions of the approach, in particular for the collaborative and distributed creation of interactive workspaces. The early prototypes of the platform used in the formative studies performed in the field with actual end users offered limited possibilities for collaborative and distributed creation of workspaces, which in those studies we called “Interactive Workspaces” (IW). For reasons of space, not all the details of the studies are reported in this paper.

### A. Field study in the Cultural Heritage domain

One of the studies was carried out in the context of visits to archaeological parks [4]. Before the visit, two professional guides composed their IWs relative to the archaeological park of Egnathia (in Southern Italy) using a desktop application, accessible through a PC placed in their office. A few days later, they experimented the use and update of IWs with a large multi-touch display (46-inch) and a tablet device (7-inch) during two guided visits of the archaeological park, involving 28 visitors. Before starting the visit, the professional guide interacted with the IW she created, in order to “enhance” her presentation of the history of the park. The IW was then deployed on a large multi-touch display available at the entrance of the park museum. Media contents, such as photos, videos, and wiki pages associated with park locations to be visited during the guided tour, were represented by an icon and a title placed on a Google map centered on the park. In this case, the map was the visual template adopted to guide the content selection and aggregation. By tapping on an icon, a pop-up window visualizes the corresponding media. For example, in Fig. 1a the guide has tapped on an icon on the map to show the 3D reconstruction of a church that was at that place at Roman times. During the park tour, the guide accessed her IW on the tablet, in order to show photos, videos and other information when appropriate (see Fig. 1b).



(a)



(b)

Figure 1. The guide is using her IW on: (a) a multi-touch display to illustrate the park history to the visitors; (b) a tablet during the park tour.

The study showed a general appreciation of the use of IW in the context of the visit and many interesting insights emerged. Guides would like to *communicate and share information* with other stakeholders and would appreciate collaborating with colleagues, both synchronously and asynchronously, during IW composition. During the interview at the end of the composition phase, the guides said they would like to be able to *compose collaboratively* the IW before a visit, even when working at home. They would also appreciate to *ask advice* to colleagues about new services that can provide material they are not able to find through the services they have access to. They would like to *share their IWs* with visitors to allow them to view and possibly add contents. The guides might also need to communicate with software engineers managing the platform, to *ask for modifications* of the user interface structure, or for the introduction of new templates for information visualization.

From interviews with the visitors, a general appreciation of the experience emerged, but also two main limitations. First, when the guides were explaining and introducing the visit showing in-depth contents through the large display, they were covering the screen with their body because they needed to be next to the multi-touch screen to interact with. Hence, visitors were not able to see the whole screen. Second, when the guides were outdoor and were showing contents through the tablet, most visitors were not able to see the screen because of its

limited dimensions with respect to the number of people in the visiting group. Multi-device collaboration mechanisms could be used to solve these issues by enabling remote control of the contents displayed on the large multi-touch screen or content delivery to visitors' mobile devices.

### B. Field study in the Technology-Enhanced Learning domain

Another field study, performed in a context of Technology-Enhanced Learning (TEL), allowed us to analyze the use of the platform in a situation in which students learn about a topic presented in class by their teacher, complementing the teacher's class by searching information on the Web [9]. The retrieved information can also be communicated and shared with the teacher and the other students using interactive whiteboards, desktop PCs and personal devices (e.g. laptop, tablet and smartphone).

The study was carried out at the technical high school "Antonietta Cezzi De Castro" in Maglie, a city in Southern Italy. It was organized over three days, involving a class of 16 students (9 females, 19 year-old on average) and a teacher. During the first day, the teacher composed an IW relative to "Communication Networks". Two days later, the teacher gave a lesson supported by the IW visualized on an interactive whiteboard. At the end of the lesson, he divided the students in groups of 2-3; each group was assigned the task of creating an IW about a specific Communication Networks sub-topic, e.g., protocols, packet switching, latency period. After a brief individual training session, all the groups accessed the laboratory to carry out their assignments. Fig. 2 shows a couple of students working with their IW, to which they are adding widgets visualized through the list visual template, to retrieve and integrate contents from Google, Slideshare and YouTube. At the end of this session, we simulated the sharing of their IWs with the teacher by manually integrating their components into a unique IW accessible by the teacher.



Figure 2. Two students working with their IW on a desktop PC.

After two days, teacher and students met again for a class on Communication Networks; this time the class was supported by the integrated IW running on the interactive whiteboard (see Fig. 3). The discussion on the retrieved information lasted for one hour and a half. At the end, teacher and students had 20 minutes to fill in a short questionnaire inquiring about platform pros and cons they perceived.



Figure 3. A student discussing about Communication networks by using the integrated IW on the interactive whiteboard.

A significant part of this field study was a design workshop that was conducted afterwards, in order to better understand the need for collaborating with other people by means of IWs. The design workshop aimed at engaging students and teacher in: 1) elicitation of positive and negative aspects of the overall interaction experience with the platform; 2) active participation in the design of new solutions, primarily stressing the envisioned possibilities of collaborative composition of an IW. The latter objective derived from the results of the field study in the Cultural Heritage domain indicating the willingness of professional guides to compose collaboratively the IW to be used during a visit. Four groups were formed, each involving four students, one interaction designer, one platform developer and one HCI researcher. One group also included the teacher. Stimulated by the researcher, participants elaborated their ideas about interaction possibilities. Then, they were asked to sketch such ideas (see Fig. 4). The design workshop lasted one hour and half. At the end, a plenary session of 30 minutes was held in which the more promising ideas were illustrated and discussed. They were instrumental for the design of both functionality and visual interface of the collaboration mechanisms that we next implemented in our platform, as it will be discussed in Section III.



Figure 4. A group sketching interaction ideas during the design workshop.

The analysis of videos and notes taken during the workshop revealed that all groups were very active. In general, it emerged that both students and teacher wish *more flexibility* in

organizing the interactive workspace, with *visual containers* in which retrieved content can be arranged and classified according to unforeseen needs. They stressed that the platform should be improved to *support collaborative activities* and contributed in the design of possible features and usage scenarios. The teacher proposed a “peer-learning” workspace, in which both teachers and students can *share their contents*, *offer comments* or *create a discussion thread*, and *express their appreciation* in a Facebook or YouTube style. The students envisaged the possibility of a *distributed collaborative creation* of a workspace, which could be asynchronous in case of a homework assignment or synchronous if carried out in class during a lesson.

### C. Summary of collaboration requirements

In both studies, the availability of live collaboration mechanisms and of annotations at different levels was identified as a key feature of interaction, composition and update of the DIW. Live editing was in particular singled out as a mechanism to show and share, in real time, personal contributions with other stakeholders that could enrich/improve a workspace. Annotations could then be used as personal memos, e.g.: remembering – by highlighting significant parts of a DIW; as expressions of thinking – by adding one’s own ideas, critical remarks, questions; and as clarifying elements – by reshaping the information in the DIW into one’s own verbal representations. The need for storing in a frozen form items from the dynamic content displayed in the workspace was also stated as a special kind of asynchronous collaboration. Therefore, annotations were in general deemed useful for sharing information and communication among DIW stakeholders, as they can support discussions among users having access to a same workspace.

## III. COLLABORATIVE INTERVENTIONS ON DIWS

Given the collaboration requirements illustrated in the previous section, we now discuss what covering such requirements entails, if a mashup composition paradigm is adopted for the creation of the interactive workspaces. An *interactive workspace* (IW) can be defined as an interactive document corresponding to any instance of a schema that is specified along three main dimensions:

- a composition model (CmM), describing the organization of the UI components in the IW and the way they synchronize by means of event-driven, publish-subscribe couplings;
- a content model (CnM), describing the actual content dynamically fetched by the different services and the way it is integrated into each UI components starting from the retrieved result sets. Given the dynamic nature of IWs, content is specified by means of queries on the involved services, which are in turn expressed according to some service-specific schema;
- a visual template model (VTM), describing the presentation aspect of the integrated data sets forming the IW through the association of queries to elements of the adopted visual template.

The organization of an IW along these dimensions is



specified in schemas expressed in an XML-based language; each workspace can be thus represented as a tree, where nodes define *composition elements* and leaves present the actual *content* users can interact with, and the visual template adopted for its visualization. In the definition reported above, distribution only refers to the component-based nature of the IWs, retrieving data from distributed resources. However, we also show how introducing collaborative mechanisms leads to distributed execution of IWs along different application instances, each instantiating the whole schema or only portions of it, depending on the users' access rights, the collaboration needs and the workspace sharing settings.

The collaboration dimension thus complements the previous IW definition leading to the notion of *Distributed Interactive Workspaces* (DIWs). Hence, a collaborative process for the creation of DIWs first of all requires the possibility to share a same application schema, according to specific users' access rights. Based on this, collaboration then consists in the production by the involved users of additional information, associated with some node or leaf in this tree by means of collaborative interventions.

A *collaborative intervention* on a document representing a DIW is seen as the creation of a collection of additional elements, together with a mapping describing the relation of each such element to the original document. The additional elements can refer to the original document as a whole, or to any subtree or leaf in its composition. The structure of each additional element is defined by a schema supporting at least the following data [10]:

- the *author*, as identified during the interaction;
- the *timestamp* when the new element is committed to the DIW repository;
- the *source*, as identified by the interactive selection of the part of the document to which the element refers;
- the actual *info* added by the author;
- its *visibility*, either private, public, or group-based.

An interactive process can exploit specific tools to identify the source, typically text selection or sketching. In general, arbitrary fragments of the original workspace, or sets of fragments across the tree structure, can be selected and associated with an additional element. For example, one can draw a shape to identify an area of a picture, or select several areas and collectively refer to them [11].

While interacting with the workspace, users can perform three types of interventions: *annotation*, *live editing*, *freezing&aggregation*, distinguished according to the nature of the info descriptor and to their usage within the platform.

With annotation, the *info* descriptor is of an arbitrary nature and can consist of any kind of digital data. The info added in an annotation can be used to generate an independent document, without corrupting the original workspace. The new document can be annotated in turn, thus supporting forms of *asynchronous collaboration*, for example by constructing annotation threads. Moreover, annotations can be used to request modifications of parts of the workspace itself,

delegating the realization of the request to authorized users, i.e., users with visibility on the annotation. In particular, every user can access his or her private annotations and all public annotations, while users belonging to some group can access annotations posted to that group. In general, the annotation process enriches the DOM of the loaded document with specific tags in correspondence of user selections. These tags are then saved with reference to that DOM. When a document on which some annotation was performed is loaded again, all the tags for the corresponding DOM are also loaded and used to create and render its enriched version.

With live editing, *info* describes a set of modifications to any structural element included in the workspace schema (e.g., a component or its underlying services in CmM, an integration query in CnM, the visual template), which are immediately activated on the instance in use by their author but also reflected to the aspects of workspace composition and behavior shared with other users, defining a form of *synchronous collaboration*. Examples of live editing interventions are: addition of a new component, or deletion of an existing one, or modification of the component query, resulting in an update of the content displayed in the component.

Finally, *freezing&aggregation* refers to a process by which, during a specific user's interaction, snapshots of (fractions of) actual contents are captured and added to a list (an *aggregator*) of contents. The content of the *info* descriptor is therefore the selected fraction of the content associated with the source at the time of the intervention. Users can build any number of aggregators and add any set of *frozen data* to each of them, by indicating the target aggregator at the time of freezing. The default visual template for aggregators iterates on the aggregator list to present frozen data, each according to its original format. All of these processes are enabled by special mechanisms through which the actions on any instance of the DIW are captured and propagated, if needed, to the other active instances of the DIW. As described in Section IV, this is made possible via an interaction between the client-side modules managing the composition and execution of each DIW instance and a server-side module managing persistency and evolution of the DIW schema.

In order to better understand how end users collaborate remotely in the creation of a DIW, let us consider the following example. Mario is a high school student; his teacher assigned him, and two of his classmates, a homework for which they have to collect documents and multimedia resources. Mario opens the platform and, by clicking on the *share* button (at the right side of 1 in Fig. 5), invites his friends Giuseppe and Alessandro to collaborate with him. After some minutes, Mario sees that his two friends are online. He types a message in the chat tool (4 in Fig. 5) to start collaborating. Both Giuseppe and Alessandro add a UI component to access a service. The live editing mechanisms allow Mario to understand what Giuseppe and Alessandro are doing by highlighting the border around the component: for example, an orange border highlights interventions from Giuseppe (he has added a UI component for the Slideshare service), while a pink border indicates those from Alessandro (he added the UI component for the Vimeo service). Furthermore, the annotation feature of the platform

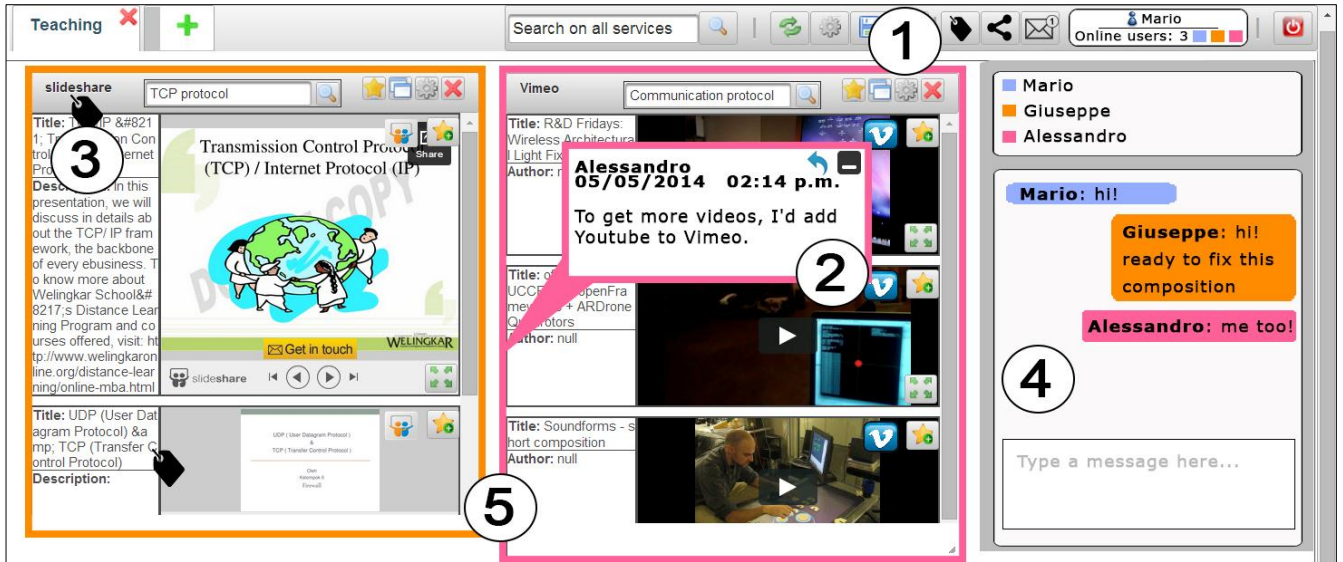


Figure 5. Example of DIW, highlighting features for asynchronous and synchronous collaboration.

allows each collaborator to attach a note to the DIW by clicking on the *note* button (at the right of 1 in Fig. 5). Alessandro has attached a note to the Vimeo component (the balloon indicated by 2 in Fig. 5) to indicate that he would like to perform the union operation of Vimeo and Youtube, in order to retrieve more videos. A notification (the envelop indicated by 1 in Fig. 5) makes Mario aware of this new element. Mario can minimize the note by clicking the minimize icon at the top-right corner of the balloon. He can also reply to it by clicking the reply icon near the previous one. Mario and his friends go on adding other components to the DIW, and retrieve content and multimedia resource by using such components. To save a specific result as “frozen content”, Mario has to click on the *star* icon available at the top right corner of every visualized content. All these contents will be shown in a specific aggregation container to be shared with the teacher.

It is worth noticing that the combination of live editing and annotation poses some specific issues, concerning the problem of orphan annotations, i.e., annotations referring to content items, which are no longer present in the document. While in principle some mechanism could be devised to retrieve content which has simply moved to a different position [12], we adopt here a conservative stance, removing any annotation referring to a node in a subtree which has been eliminated by a live editing process. This choice concerns annotations on service composition and visual template models, while those on content provided by a service can be retrieved if the same service offers the same content at some subsequent request, or if the content has been frozen to some aggregator.

#### IV. ARCHITECTURE OF THE DIW PLATFORM

Fig. 6 illustrates the organization of the platform supporting the composition paradigm and the collaboration interventions illustrated in the previous sections. The definition of a DIW is performed through the *Workspace Composition Environment*, an HTML/JavaScript Web application where end users can execute the composition actions and immediately see the result, i.e., a running application, thanks to the adoption of a live

programming paradigm. In particular, a *Workspace Manager* on the client-side intercepts the visual mapping and synchronization actions performed by an end user. Through its *Schema Manager* module, such actions are automatically translated into elements of a *Workspace schema*, expressed in an XML-based domain specific language [6], which describes the service queries, the association of the query results with specific visual templates, and possible synchronizations among different visual templates.

In order to support the live programming paradigm, as soon as new elements are added into the XML schema the workspace is immediately updated to show the changes, thus making it possible the interactive definition and execution of DIWs. Therefore, the workspace manager is in charge of:

- querying services dynamically (through the *Service Manager* module), according to the queries defined in the workspace schema. Pre-defined service URIs and query parameters are specified in service descriptors stored in proper repositories;
- supporting visual refinement of service queries, as end users can define new selection and projection queries over the result set from a UI component service, as well as new union and join queries to integrate data of additional services;
- displaying dynamically the retrieved result set in a visual format, according to the visual mapping by the users and expressed in the *Workspace schema*.

Execution of a DIW can occur in the very device where the composition is created, as well as in *Execution Environments* running on different devices. Executing a DIW indeed simply requires an Execution Engine (which can be coded according to any Web or device-native technology), able to interpret the Workspace schema (*Schema Interpreter*) and instantiate the adopted visual templates (*UI controller*), by rendering the corresponding UI and filling the visual elements with data requested to the involved services (*Service Querying module*).

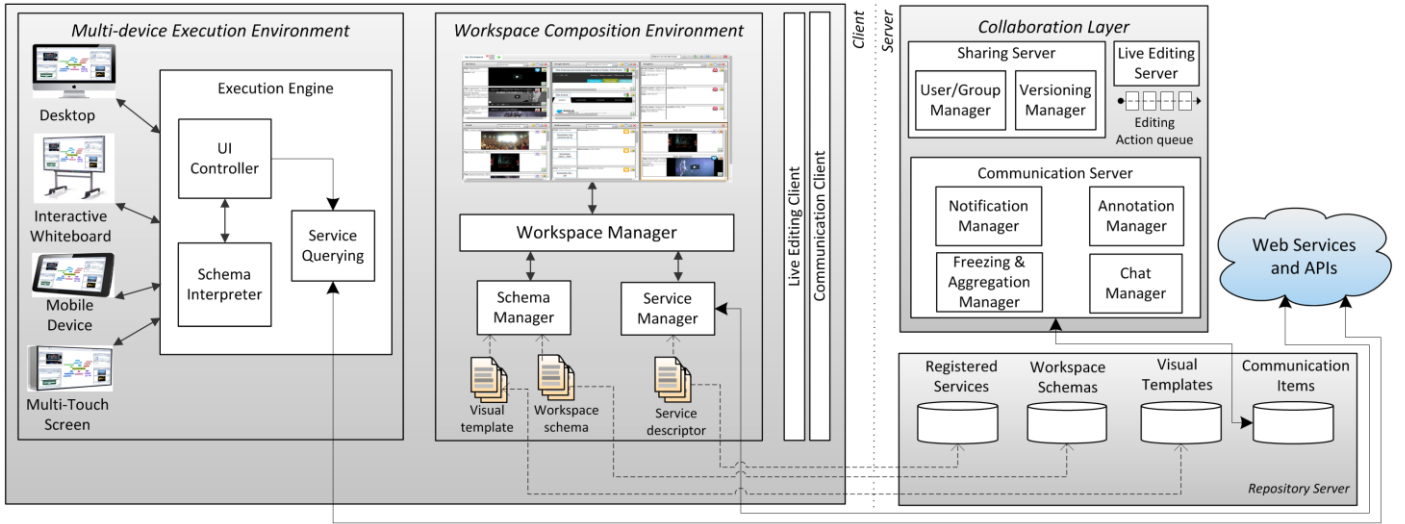


Figure 6. Architecture of the platform for the collaborative creation and use of DIWs.

The architecture of the original composition platform hosts all the modules for DIW composition and execution at the client side. In order to preserve such a “lightweight” approach, in the extended platform the pure composition logic still resides at the client side and the creation of workspace schemas is still operated on the clients. However, as highlighted in Fig. 6, a server-side *Collaboration Layer*, together with additional client-side modules, takes care of persistence management and of co-evolution of schemas. The server also hosts the required repositories (for *Registered Services*, *Workspaces* and *Visual Templates*) and the database to store the data items required to manage synchronous and asynchronous communication.

Specific modules take care of the collaboration aspects. The *Sharing Server* manages resource sharing, by handling users’ access rights and versioning of the released resources. This module also enables the distribution of portions of a same DIW on multiple devices. Based on the adopted sharing policies, the local execution engines instantiate and show only the part of the DIW a user or a device is authorized to use.

For live editing, every relevant modification on a DIW composition and execution is propagated to any other running instance of the same DIW. In particular, the *Live Editing Client* (LEC) captures the elements describing the modification to the DIW structure, and propagates them to the *Live Editing Server* (LES), which takes care of the DIW schema evolution by maintaining a representation of the distributed editing actions. Every editing session on a given DIW has an associated *Editing Action queue*. As illustrated in the sequence diagram in Fig. 7, any active instance of the DIW periodically queries the LES to know whether new actions, generated by other DIW executions, are available.

Any live editing action propagated to the clients is represented as a pair  $\langle \text{modifiedObject}, \text{notification} \rangle$ . The first element represents the argument of the action (e.g., a component, a service binding, an inter-component coupling, a query parameter) and all its properties. The second represents some metadata (e.g., the ID of the user who performed the

action), needed for notifying the change. The *modifiedObject* properties have effect on the composition schema; the *notification* element is used by the LEC to visually highlight the action (e.g., highlighting the widget in pink in Fig. 5). The LEC thus interprets the received actions and triggers events to let the *Workspace Manager* modify the composition schema accordingly and reload it. Changes are highlighted in the DIW based on the *notification* meta-data. To reflect DIW changes with minimal delay, the LEC periodically checks the composition status representation through the LES, to verify whether some actions must be loaded and rendered within the DIW instance. The sequence diagram in Fig. 7 illustrates communication between LEC and LES for publishing (by client *c11*) and retrieving (by client *c12*) edit actions.

This form of synchronization of all the active DIW instances implies a “distributed” representation of the DIW schema, maintained at each client. Server-side management of the action queue ensures synchronized evolution of all the active DIW instances. Differently from the composition model in the original platform, the DIW is now *stateful*. The schema is enriched with state meta-data, (e.g., parameter values to query single components, items selected in a data set) to synchronize each DIW instance not only on the composition structure (components and bindings), but also with respect to the displayed data set. Hence, composition is now *long-lasting*, maintaining structure and state across different sessions.

Interaction between client and server modules and schema persistence and evolution is also needed for enabling both synchronous and asynchronous communication among stakeholders in form of annotations, frozen data, and messages exchanged via chat sessions. With respect to live editing modules, the communication server and the communication client manage the addition of collaboration-oriented information to the original workspace, which we call *Communication Items*. Communication items are persisted by different modules (the *Annotation Manager*, the *Freezing & Aggregation Manager*, the *Chat Manager*), and are retrievable by authorized users when uploading the original document.



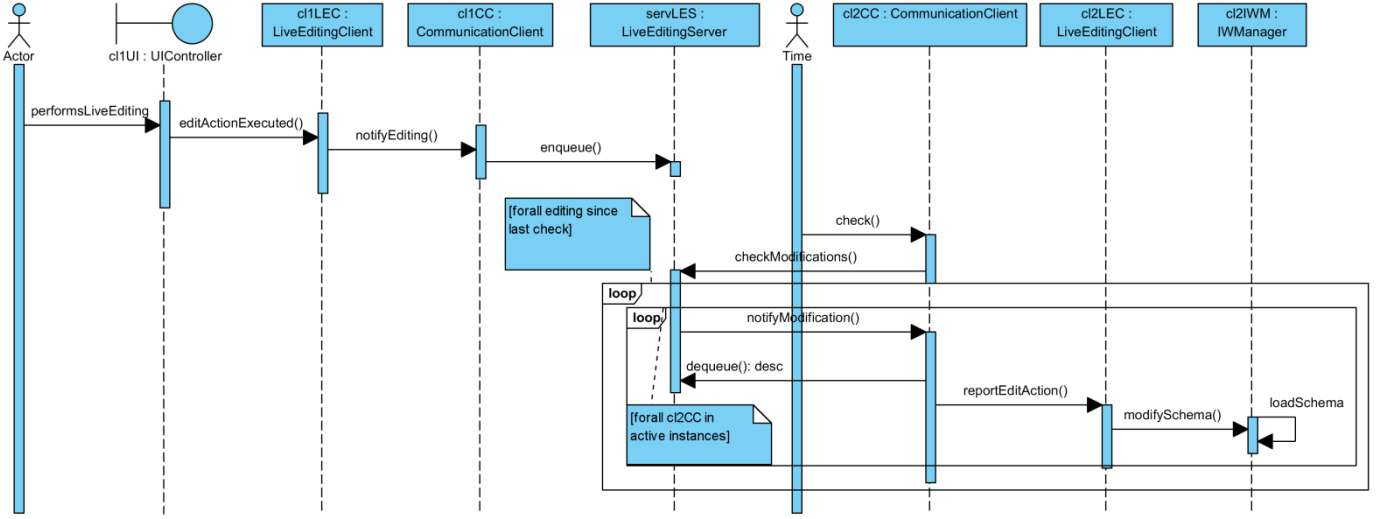


Figure 7. Communication between instances of Live Editing Client and Live Editing Server for publishing (client c1) and retrieving (client c2) edit actions.

Such communication items can also be presented to users in a live form. To this end, an instance of the DIW can inquire with the server if there are annotations for some of the nodes currently present in the composition instance whose timestamp is more recent than the last check for annotations. The way each communication item is displayed then depends on the nature of the item: annotations are displayed in popup windows attached to the workspace elements they refer to; frozen data are presented in an aggregation component in charge of grouping all such items collected by the different users sharing the workspace; chat messages are displayed in an ad-hoc viewer for chats.

## V. RELATED WORK AND DISCUSSION

While collaboration is a mature research field in some communities such as the one working on Computer Supported Cooperative Work, it is not still quite explored in the creation of Web artifacts, especially Web mashups. Tools have been proposed to ease mashup development for unskilled users unable to program the component integration logic [13, 14]. By offering intuitive visual notations to substitute programming, these tools reached the goal of enabling end user development of mashups. However, most of them offer paradigms for the creation of single-user applications, while they do not support the co-creation of shared information workspaces.

In the context of Web-based collaborative learning, Web Space Configuration is introduced in [15] as a basic container for instantiating W3C Widgets. Composition of widgets is independent of the runtime environment. Independence is exploited to support portability of the created applications, and sharing via broadcasting and co-editing. These are achieved by establishing a long-lasting connection by the owner of a Web space, which invites other users to join and see the Web space (broadcasting) and to apply changes (co-editing). Our approach also exploits independence from the runtime environment to enhance portability of the created interactive workspaces on different devices. The collaboration paradigm we propose also covers sharing and live co-editing. While Web space co-editing only focuses on the presence awareness aspect, we also support

action awareness, by propagating and notifying in real time any change applied by one of the collaborating users on a shared information space. In addition, we support synchronous and asynchronous communication through chat, annotation, and frozen data mechanisms. These dimensions are not covered by the Web Space Configuration approach. We also believe that aggregator components for freezing data items are original in the mashup world. In this field, indeed, applications are fully dynamic, meaning that they retrieve data via instant queries to the involved services. The field studies revealed that storing single specific data items is a recurrent need of real users.

In [16], the authors propose a crowdsourcing paradigm where user participation is adopted as a solution to responsive design in Web application development, trying to collectively solve problems related to the adaptation of Web applications to different device screens. System developers provide an interface where adaptive features can evolve at runtime with the help of users, who can refine the adaptations to better match peculiar usage context. This paradigm opens Web development towards the social dimension. However, its aim is limited to letting users customize the presentation of their Web applications to best fit their current device, while it neglects collaboration and coordination of different stakeholders. Moreover, it focuses on presentation adaptation, not covering modification of the application content at all. The approach presented in this paper is instead specifically targeted towards handling content. In [17] the same authors then discuss how to distribute the execution of mashups along different devices. However, they do not deal at all with synchronous and asynchronous collaboration.

In [7] a generic awareness infrastructure is proposed for providing basic awareness services reusable throughout different platforms. Awareness support is anchored at a standardized layer to provide an application agnostic solution. Different collaborating clients include a component, the *generic awareness adapter* that embeds awareness widgets and is devoted to propagating contextual information (from the client to the server and vice-versa). The approach is especially interesting because of its portability across different platforms

and its intrinsic extensibility, being based on the integration of widgets managing the different collaboration aspects. However, it only manages awareness in live-editing sessions, while our approach also covers asynchronous communication, shifting the focus from applications with duration limited to single sessions to long-lasting applications that support knowledge evolution and consolidation. Although the collaborative paradigm proposed in this paper is based on ad-hoc extensions implemented in the platform for interactive workspace composition, the architecture extensions have been designed as components, detached by the composition editor, with an adaptable event-driven logic. The collaboration components can be thus easily plugged (or unplugged), and adapted to other collaboration environments by customizing the events exchanged between client and server.

## VI. CONCLUSIONS

As users become increasingly familiar with Web 2.0 mechanisms to exchange ideas and instantly communicate with peers, collaboration becomes a fundamental feature of modern Web-based applications. However, service-based Web composition environments offering this feature are still lacking. This paper tries to fill this gap by showing how services, service-based resources, and composition models can be considered objects of collaboration. The presented collaborative features emerged from a series of studies where real users expressed their desiderata on possible collaborative aspects of a mashup platform. Besides operations such as Create, Read, Update, Delete applied to UI components and workspaces, already offered by the previous version of the platform, the collaboration extensions now support resource publishing and versioning on the common platform repository, and the definition of associated access rights for the other stakeholders. By proper setting of sharing policies, users can choose to distribute the entire DIW or part of it on multiple devices and among different peers. This feature is particularly useful in scenarios where each collaborating user contributes with single components to the creation of a shared workspace. Each single actor can manage an arbitrary complex workspace, but share with the others only some specific components or also some specific content items feeding shared aggregator components. We were also able to identify new forms of communication, for example through freezing and aggregation of single content items. This is a novel characteristic, peculiar for mashups but at the same time scarcely investigated in the mashup world, which could be easily extended to different classes of content-intensive collaborative systems.

The platform prototype described in Section III was created on the basis of the design workshop involving end users. Formative evaluation of this prototype, performed with a thinking aloud test in laboratory, indicated that the users appreciated it. Future work aims at further validating the extended composition approach, to assess in the field the effectiveness of the intermixing of service-based interactive composition and collaboration features. Therefore, new user-based studies have been planned in the same usage contexts where the field studies reported in this paper were conducted.

## VII. ACKNOWLEDGMENT

This work is partially supported by the Italian Ministry of University and Research (MIUR) under grant PON 02\_00563\_3470993 "VINCENTE" and by the Italian Ministry of Economic Development (MISE) under grant PON Industria 2015 MI01\_00294 "LOGIN". We are grateful to all people involved in the reported studies.

## REFERENCES

- [1] J. Yu, B. Benatallah, R. Saint-Paul, F. Casati, F. Daniel and M. Matera, "A framework for rapid integration of presentation components", *Proc. WWW '07*, 2007, ACM, pp. 923-932.
- [2] J. Yu, B. Benatallah, F. Casati and F. Daniel, "Understanding Mashup Development", *IEEE Internet Comp.*, vol. 12, no. 5, 2008, pp. 44-52.
- [3] N. Zang and M.B. Rosson, "What's in a mashup? And why? Studying the perceptions of web-active end users", *Proc. VLHCC '08*, 2008, IEEE Computer Society, pp. 31-38.
- [4] C. Ardito, M.F. Costabile, G. Desolda, R. Lanzilotti, M. Matera, A. Piccinno and M. Picozzi, "User-Driven Visual Composition of Service-Based Interactive Spaces", *Journal of Visual Languages & Computing*, vol. 25, no. 4, 2014, pp. 278-296.
- [5] C. Ardito, P. Bottoni, M.F. Costabile, G. Desolda, M. Matera, A. Piccinno and M. Picozzi, "Enabling End Users to Create, Annotate and Share Personal Information Spaces", *End-User Development - Is-EUD 2013*, LNCS 7897, 2013, Springer, pp. 40-55.
- [6] C. Cappiello, M. Matera and M. Picozzi, "End-User Development of Mobile Mashups", *Design, User Experience, and Usability. Web, Mobile, and Product Design - HCII '13*, LNCS 8015, 2013, Springer, pp. 641-650.
- [7] M. Heinrich, F. Grüneberger, T. Springer and M. Gaedke, "Reusable Awareness Widgets for Collaborative Web Applications – A Non-invasive Approach", *Web Engineering - ICWE '12*, LNCS 7387, 2012, Springer pp. 1-15.
- [8] C. Cappiello, F. Daniel, M. Matera, M. Picozzi and M. Weiss, "Enabling end user development through mashups: requirements, abstractions and innovation toolkits", *End-User Development - Is-EUD 2011*, LNCS 6654, 2011, Springer, pp. 9-24.
- [9] C. Ardito, M.F. Costabile, G. Desolda, R. Lanzilotti and M. Matera, 2014, *Creating Flexible Interactive Workspaces through Data Source Composition*, IVU Lab - technical report - n.02-2014. Available at: [http://ivu.di.uniba.it/papers/2014/IVU\\_TR\\_2-2014.pdf](http://ivu.di.uniba.it/papers/2014/IVU_TR_2-2014.pdf).
- [10] P. Bottoni, S. Levialdi, N. Pambuffetti, E. Panizzi and R. Trinchese, "Storing and retrieving multimedia web notes", *Int. J. Comput. Sci. Eng.*, vol. 2, no. 5/6, 2006, pp. 341-358.
- [11] M. Addisu, D. Avola, P. Bianchi, P. Bottoni, S. Levialdi and E. Panizzi, "Annotating Significant Relations on Multimedia Web Documents", *Multimedia Information Extraction*, 2012, John Wiley & Sons, Inc., pp. 401-417.
- [12] P. Bottoni, A. Cotroneo, M. Cuomo, S. Levialdi, E. Panizzi, M. Passavanti and R. Trinchese, "Facilitating interaction and retrieval for annotated documents", *Int. J. Comput. Sci. Eng.*, vol. 5, no. 3/4, 2010, pp. 197-206.
- [13] S. Aghaee, M. Nowak and C. Pautasso, "Reusable decision space for mashup tool design", *Proc. EICS '12*, 2012, ACM, pp. 211-220.
- [14] A. Namoun, T. Nestler and A. De Angeli, "Service Composition for Non-programmers: Prospects, Problems, and Design Recommendations", *Proc. ECOWS '10*, 2010, IEEE Computer Society, pp. 123-130.
- [15] S. Sire, E. Bogdanov, M. Palmér and D. Gillet, "Towards Collaborative Portable Web Spaces", *Proc. MUPPLE '09*, 2009.
- [16] M. Nebeling, S. Leone and M. Norrie, "Crowdsourced Web Engineering and Design", *Web Engineering - ICWE '12*, LNCS 7387, 2012, Springer, pp. 31-45.
- [17] M. Husmann, M. Nebeling and M. Norrie, "MultiMasher: A Visual Tool for Multi-device Mashups", *Current Trends in Web Engineering - ICWE '13 Workshops*, LNCS 8295, 2013, Springer, pp. 27-38.

# A Framework for Bimanual Inter-Device Interactions

Ali Roudaki

Jun Kong

Gursimran Walia

Department of Computer Science

North Dakota State University

Fargo, USA

{ali.roudaki, jun.kong, gursimran.walia}@ndsu.edu

**Abstract**— A shared interactive display (e.g., a tabletop) provides a large space for collaborative interactions. However, a public display lacks a private space for accessing sensitive information. On the other hand, a mobile device offers a private display and a variety of modalities for personal applications, but it is limited by a small screen. We have developed a framework that supports fluid and seamless interactions among a tabletop and multiple mobile devices. This framework can continuously track each user’s action (e.g., hand movements or gestures) on top of a tabletop and then automatically generate a unique personal interface on an associated mobile device. This type of inter-device interactions integrates a collaborative workspace (i.e., a tabletop) and a private area (i.e., a mobile device) with multimodal feedback. To support this interaction style, an event-driven architecture is applied to implement the framework on the Microsoft PixelSense tabletop. This framework hides the details of user tracking and inter-device communications. Thus, interface designers can focus on the development of domain-specific interactions by mapping user’s actions on a tabletop to a personal interface on his/her mobile device. A user study compared our interaction style with a standard tabletop interface and justified the usability of the proposed interaction.

**Keywords**—bimanual interaction; multimodal interface; tangible interface; human computer interaction

## I. INTRODUCTION

The benefits of tabletop-based applications have been investigated in different scenarios, such as collaboration [3] or pedestrian navigation [21]. However, it is hard to protect personal information on a tabletop. In addition, a public environment limits the usage of some modalities. For example, the usability of auditory feedback through a speaker is reduced in a noisy public environment. While a mobile device provides diverse multimodal feedback, it is limited by its small screen which makes it frustrating to browse a large amount of information.

Therefore, a synergistic interaction with mobiles and tabletops integrates their merits. While previous studies have explored the benefits of combining mobile devices and a large display together (e.g., a fluent switch between individual and group work [32]), one challenge in inter-device interactions is to minimize distractions when switching between devices. Another challenge is to develop a generic solution that is suitable for various scenarios and applications. Although different techniques have been developed to support interactions in a multi-device ecology [10, 19], few researchers

have focused on developing a generic platform that supports a variety of applications. Recently, based on PhoneTouch [29], Schmidt et al. [31] developed a generic platform supporting a novel interaction style that fits different applications. This interaction style is featured by pairing a phone touch event with the identity of the phone through an accelerometer.

In contrast to the PhoneTouch interaction style [31], our generic framework (hereinafter referred to as MobiSurf) supports a bimanual interaction style using a tangible object. Our approach uses a passive tangible object to perform coarse-grained selections on a tabletop while a mobile device is held by the dominant hand for fine-grained interactions. Based on the previous studies on the tangible interfaces [13, 34], we proposed various gestures (e.g., “Pointer Rotate”, “Pointer Move”, and “Pointer Share”) that enable a natural interaction with a tabletop device. The gesture of each user on the tabletop is detected and accordingly produces a unique personal interface with multimodal feedback on the associated mobile device. The MobiSurf framework is featured by a thin client on the mobile device, which is application-independent. In other words, during the development process, the client side, which encapsulates the functions of the interface generation and inter-device communications, is generic for different applications and does not need any revision. Such an implementation allows developers to focus only on application-specific developments by translating user gestures on a tabletop to appropriate messages on an associated mobile device (See Section III). To evaluate the usability of the proposed interaction style, we conducted a controlled empirical study. Participants were asked to complete two tasks using both a standard tabletop interface and the MobiSurf interface. The results from this study showed significant improvements in the usability when using the MobiSurf interaction style as compared to the standard interface.

## II. RELATED WORK

Researchers have explored the combination of mobile devices and tabletops to improve the usability for collaborative tasks, such as augmenting a computer with PDAs in the single display groupware [22] or exchanging information between a personal device and a public display [9]. This section reviews different techniques for user tracking and inter-device interactions.

### A. User Identification and Tracking

Because a tabletop represents a public space, it is necessary to identify the users to protect their personal data. Various

---

This work is in part supported by NSF under grant #CNS-1126570.

approaches have been proposed to pair a user's interactions on a tabletop with a unique ID, such as hand biometrics [30], utilizing the back of a user's hand as an identifier [26], infrared light pulses through a ring-like device [27], or using a tangible interface to authenticate users [34]. However, these approaches did not support direct data sharing between a public device and a personal one. Since most tabletop/mobile devices are equipped with a camera, the computer vision technique has been commonly used to associate a mobile device with a large display. For example, BlueTable [35] implemented a vision-based handshaking procedure by blinking an infrared light or flashing the display of a mobile device to establish a connection between a mobile device and a tabletop. Similarly, Schoning et al. [28] used the flashlight and Bluetooth unit of a mobile phone as a response channel to authenticate users. Ackad et al. [1] used the color detection to implement a handshaking protocol to identify a registered mobile device placed above a tabletop. This system also used a depth camera so that a user can continuously be tracked even if the device was removed from the tabletop. In addition, cameras were used in inter-device communications to replace radio-based techniques (e.g., WiFi or Bluetooth) for information exchange, such as FlashLight [11] or C-Blink [20].

Instead of using the computer vision technique, various approaches used gestures to associate a mobile device with a public display based on built-in sensors. Tilt correlation [12] compared the touch-derived tilt angle on a public display with the tilt sensor information from a mobile device to distinguish different mobile devices. Patel et al. [24] proposed a gesture-based authentication by shaking a device according to a required pattern. PhoneTouch [29] correlated the phone touch events detected by an interactive surface and by a mobile phone through an accelerometer to identify multiple mobile devices. The above sensor-based approaches used a mobile device for both the user identification and interaction. Consequently, a user cannot interact with the mobile device during the identification process. Instead, our approach introduces a tangible object for the user tracking so that the user tracking and interaction can be performed simultaneously.

Interacting with the computers using tangible objects has been widely studied. The pioneering work by Ishii and Ullmer [13] bridged the gap between a physical environment and a cyberspace through a tangible interface. Our approach uses a tangible object for tracing users' actions on a tabletop.

### B. Inter-Device Interaction

A mobile device provided gesture-based interactions and multimodal feedback; thereby making it suitable for being a remote controller for inter-device interactions [18, 19]. Several approaches leveraged a built-in camera to manipulate a remote object by directly touching or moving a mobile device, such as point & shoot for remote selection [2], camera-based pose estimation for remote operation [25], a privacy-respectful input method [16], snap and grab for sharing contextual multimedia contents [17], and touch projector for interacting with surrounding displays [4]. Alternatively, by using the accelerometer, movement-based gestures were developed for interacting with a distant display [6, 33].

Instead of remotely manipulating visual objects in a distant display through a mobile device, some approaches required a direct contact between a mobile device and a public display for inter-device interactions, such as placing a mobile device above a tabletop during the entire duration of interactions [7, 23], or freely moving a mobile device on top of a public display. Hardy and Rukzio [10] used an NFC mobile device as a stylus for interacting with an NFC-tagged display.

Although the above approaches supported mobile-tabletop interactions, they were not generic for supporting a variety of scenarios. Recently, Schmidt et al. [31] developed a generic platform for the synergistic usage of mobile devices and tabletops. Built on the PhoneTouch [29] technique, this platform used a mobile device as a stylus to select objects on a tabletop and analyzed touch events to recognize the identity of a mobile phone. Our framework is different with the above approach from the following perspectives. First, our approach implements a bimanual interaction, in which the non-dominant hand performed a coarse-grained selection through a tangible object while the dominant hand held the mobile device for a fine-grained interaction. According to Buxton et al. [5], the bimanual input outperformed the one-handed input for selection, positioning, and navigation tasks. Furthermore, the bimanual interaction is capable of tracking the path of hand's movements on the tabletop while accordingly producing continuous feedback to the mobile device. Secondly, our framework implements a thin client which is suitable for different applications without modification. The implementation of a thin client allowed developers to focus on mapping user's actions on the tabletop to interaction commands on the mobile device, while the framework itself can automatically produce a personal interface on the mobile device according to the mapping.

## III. SYSTEM DESIGN AND ARCHITECTURE

### A. System Design

The MobiSurf framework was built on three types of hardware components, i.e., a tabletop device, passive tangible objects (pointers), and mobile devices. Without losing generality, we implemented our framework on the Microsoft PixelSense tabletop which supported multi-touch interactions and was equipped with infrared sensors. Each passive tangible object served as a pointer to make a course-grained selection and was associated with a distinct mobile device for passing the interaction events from the tabletop to the associated mobile device, which was used as a personal area for accessing sensitive information with multimodal feedback. The pointer can be constructed with various shapes and different materials based on the users' needs.

### B. User Tracking

In a collaborative environment with multiple users, it is necessary to identify and track user's actions to provide the personalized information and protect privacy. Mobile devices have been used to identify and track a user, such as PhoneTouch [29] or Tilt correlation [12]. However, the above approaches constrained the usage of the mobile device for interactions during the process of user identification and tracking. In order to overcome the above limitation, MobiSurf

implemented bimanual interactions based on a tangible interface. More specifically, each tangible object (i.e., a pointer) has a unique ID defined by an infrared tag which consists of a geometric arraignment of infrared reflective and absorbing areas. In the beginning, a user needs to type in the ID of a pointer on a mobile device to pair the pointer with the mobile device. The pairing process is only performed once at the beginning of an interaction session. When a particular user terminates the connection with the tabletop, the pointer is automatically released and can be used by another user.

### C. Software Design

MobiSurf applies an event-driven architecture and serves as a middleware to set up two-way communications between a tabletop and a mobile device. On one side, the MobiSurf API accepts tabletop UI events, produces a set of commands, and forwards them to a mobile device. Those commands specify the actions performed on the mobile device (such as displaying a text box or generating a vibration). On the other hand, the MobiSurf API receives responses from the mobile device and accordingly notifies the tabletop application. Based on the above communication mechanism, the mobile application has two functionalities. First, based on the received messages, it either performs proper actions or renders proper UI elements on the mobile device. Second, it generates responding messages based on the user's actions on the mobile device and sends them back to the MobiSurf API. Such an event-driven design results in a thin client which makes the mobile side application-independent. Because MobiSurf is completely compatible with standard UI elements, developers can extend a standard tabletop interface with the feature of inter-device bimanual interactions by defining the semantics of user actions on a tabletop. In summary, MobiSurf hides the details of user identification and inter-device communications and minimizes the coupling between the mobile device and tabletop through an event-driven architecture.

### D. Event Handling

Different events and messages are supported in MobiSurf.

**User Connection and Disconnection.** Once a user requests to connect to the MobiSurf API, the API raises a "NewClient" event and sets up a virtual connection with the mobile device through TCP/IP. Pairing a pointer with the mobile device is also completed as a part of the user connection. When a user terminates the connection, the "ClientRemoved" event is raised, which releases the pairing between the pointer and the mobile client.

**Actions on the Tabletop.** After the connection and pairing steps, the MobiSurf API continuously tracks user's actions on the tabletop through the paired pointer and raises corresponding internal events, i.e., "PointerOver", "PointerRotated", "PointerShare", and "PointerMove". For instance, PointerShare is triggered when two or more pointers are placed in proximity. This event is designed for content sharing among mobile devices. Table I summarizes all internal events handled by the MobiSurf API.

TABLE I. MOBISURF API EVENTS

Event	Description
NewClient	New mobile client connected
ClientRemoved	Mobile client disconnected
MessageReceived	Acknowledgement from the mobile device
PointerOver	Pointer is over a UI Element
PointerRotate	Pointer twisted over a UI Element
PointerShare	Multiple pointers are in proximity
PointerMove	Continuous movements

**Mobile Interaction.** Based on an internal event, MobiSurf allows interface developers to specify proper action messages (Table II) and sends them to the paired mobile client. For example, when a user moves his/her pointer to a text box that invites a password, MobiSurf triggers the "PointerOver" event. Based on the developer's specification, an action message is sent to the paired mobile client, which produces a text field on the mobile device for inputting the password. Therefore, the user can apply the mobile device as a private channel to input a password with improved privacy. In summary, the actions on the mobile device are classified into two groups. The first-group actions can dynamically generate UI elements on the mobile device, and the second-group ones can produce various multimodal feedback. Any combination of actions in Table II can be defined and sent to the mobile device.

TABLE II. MOBISURF ACTIONS

Action	Description
Lighting	Flash the LED light
Vibrate	Vibrate the mobile device
Beep	Generate a beep sound
TextMode	Display a textbox on the mobile device
ListMode	Display a listbox on the mobile device
Button	Display a button on the mobile device
WebLink	Open a web page on the mobile device
Text	Display textual contents on the mobile device
Speech	Speak a text on mobile device
Image	Show an image on the mobile device
Media	Play a voice or video file
AlertDialogue	Show a text alert message on screen
DataRequest	Request data stored on a mobile device

**Acknowledgement.** After the mobile client completes the required interaction, it sends an acknowledgement to the MobiSurf API. Accordingly, MobiSurf raises the "MessageReceived" event which includes the actions completed on the mobile device (such as free-style typing or selection from a list) and the information being input along with the user ID. The user's actions include "ButtonClick" (i.e., the user tapped a button on the mobile device), "ItemSelect" (i.e., the user selected an item from a list box on the mobile device) and "TextEntered" (i.e., the user inputted texts to a text box on the mobile device).

Users' gestures on a tabletop in general have a metaphoric basis. However, due to different application domains and users' backgrounds, the same gesture may intend to different commands under different interaction scenarios. Therefore, MobiSurf supports an open framework, which provides the flexibility for interface developers to determine the action of a

gesture on a tabletop by mapping the tabletop-based gesture (See Table I) to commands on a mobile device (See Table II).

#### IV. MOBISURF INTERACTION STYLE

Schmidt et al. [31] summarized six challenging issues in the use of multi-touch tabletops. We introduce the MobiSurf interaction style to address those issues.

##### A. Data Transfer

Users can transfer data from a tabletop application to a mobile device and vice versa. For example, a user can store his/her bookmark on the mobile device and then copy it to a tabletop application. In addition, multiple users (i.e., two or more users) in the same group can share information.

**Tabletop to Mobile.** A user first moves his/her pointer to a tabletop UI element that includes the information of interest; then, he/she twists the pointer which triggers the “PointerRotate” event. By mapping the “PointerRotate” event on the above UI element (e.g., a text field) to a specific action (i.e., the Text action) on the mobile device, the requested information is transferred to the mobile device.

**Mobile to Tabletop.** Automatically transferring personal information from a mobile device to a tabletop can facilitate data entry and avoid redundant inputs. A user places the pointer over a UI element that invites inputs from the user, which triggers the “PointerOver” event. A message including the “DataRequest” action is sent to the mobile device. The user can choose the corresponding information on his/her mobile device for sharing with the tabletop.

**Mobile to Mobile sharing.** A user (i.e., the source) selects the information being shared (i.e., image or text) on his/her mobile phone and moves his/her pointer in proximity to the pointer of the target user. The above action triggers the “PointerShare” event on a tabletop. Based on the type of information being shared, the tabletop application sends proper action messages (i.e. “Text”, “Image”, or “Media”) to the target mobile device. In order to avoid an accidental sharing, the “PointerShare” event can also produce an “AlertDialogue” action on the source device to confirm the sharing. The information sharing can be easily extended to three or more users by placing their pointers in proximity.

##### B. Personalization

A mobile device is ideal to supplement tabletop interactions with personal information, such as copying personal information from a mobile device to a tabletop to automatically fill in a form or defining a personal area on a public display.

**Auto Fill.** Auto fill avoids redundant data entry by automatically copying personal data from a paired mobile device to the tabletop, which is essentially implemented as data transfer from a mobile device to a tabletop.

**Region Selection.** The PhoneTouch based approach [31] only tracks separate phone touch events, while MobiSurf supports tracking the continuous movement of a pointer that corresponds to the “PointerMove” event. A tabletop application can record a moving path that includes a sequence of coordinate data from the “PointerMove” events. The information of a moving path is useful in various scenarios. For

example, we can define a personal area according to the moving path (i.e., a close area where the first point in the moving path is identical to the last one). The personalized information or adaptation (based on the mobile-to-tabletop sharing) can then be applied to this personal area.

##### C. User Interface Composition

In a collaborative environment, moving interaction commands (such as a menu) from a tabletop to a mobile device can utilize the screen more efficiently and allow multiple users to simultaneously operate commands displayed in proximity. Based on the location of a pointer on the tabletop, MobiSurf can use the mobile device as a tangible controller by displaying contextual menus.

**Expanded Screen.** During the interaction on a tabletop, a user’s mobile phone displays contextual menu items in response to the user’s action. For example, a user moves his/her pointer to an image on the tabletop which triggers the “PointerOver” event. Then, appropriate contextual menu items (e.g., save, edit, email) are sent to the mobile device through corresponding action messages (e.g., Button or ListMode).

**UI elements to Phone.** A user can move his/her pointer over a UI element and rotate the pointer (i.e., the PointerRotate) to transfer the UI element along with its content to his/her mobile phone (i.e., TextMode or ListMode). A user can manipulate the UI element on his/her mobile device (i.e., changing the content) and send it back to the tabletop.

##### D. Authentication

Authentication on shared interfaces has always been challenging [15]. MobiSurf uses the mobile device as a private channel to authenticate the identity of the user.

**Password entry.** Traditional username-password based authentication is still popular in many existing systems. However, it is not secure to input a password or other sensitive information on a shared display. By mapping a “PointerOver” event on a UI element (e.g., a text field for a password) to a “TextMode” action on the mobile device, the MobiSurf interaction style uses a mobile device as a personal device to input sensitive information. Once a user moves his/her pointer over the username/password entry panel, the login interface will be generated on his/her mobile device to protect privacy.

**Advanced Authentication.** Some applications that require advanced security need both the username/password and hardware authentication. Since a mobile device is a personal device, its International Mobile Equipment Identity (i.e., IMEI) can be used for the hardware authentication. Based on MobiSurf, an advanced authentication can be implemented by sending both the TextMode (i.e., inputting the password) and DataRequest (i.e., sending the IMEI information) actions to a mobile device.

##### E. Localized and Private Feedback

Since a mobile device supports different output modalities (e.g., vibration, flash or sound), it is ideal to produce personalized and private feedback.

**Multimodal Feedback.** MobiSurf supports various types of feedback on the mobile device, such as a flashing light,



beep, vibration, or speech. For example, when a user twists his/her pointer over a text block on the tabletop, the text-to-speech service can be utilized (based on the “Speech” action) through a speaker or earphone on the mobile device. With an open event-driven framework, any combination of output modalities can be defined through action messages.

**Accessibility.** Researchers have investigated on improving the accessibility of a public interface for blind users [14]. MobiSurf enables visual-impaired users to interact with a tabletop by providing personalized interactions and feedback through a mobile device. More specifically, MobiSurf tracks user’s actions on a tabletop (e.g., “PointerOver” or “PointerMove” events) and accordingly generates vibration and voice feedback (through the “Vibrate” and “Speech” actions) to guide the user to access various parts in an interface.

#### F. Input Expressiveness

In addition to the traditional interaction methods, MobiSurf exploits tangible objects as an additional input method.

**Movement Tracking.** The “PointerMove” event supports the movement-based gesture to control a tabletop application, such as a region selection or accessibility.

**Multiple Pointers.** MobiSurf supports pointers with different shapes as long as a proper infrared tag is attached to each pointer. Furthermore, multiple pointers can be registered to a single user. The shape, the number of tags associated with a user, and gestures performed on a pointer define a design space for tangible interactions. For example, in a chess game application, each player has 16 pointers with different shapes and the movement of a pointer produces the corresponding feedback on both the tabletop and the mobile device.

In summary, MobiSurf seamlessly integrates multiple mobile devices with a tabletop and is potentially useful in different applications. For example, MobiSurf can provide a user-friendly interface for both visually impaired users and normal users. More specifically, vibration and speech feedback on a mobile device allows disabled users to access information while normal users read visual information on a tabletop.

## V. EVALUATION

A controlled empirical study was conducted to investigate the user experience.

#### A. Research Hypotheses

The following hypotheses were formulated for this study:

H1: MobiSurf interface is better than a standard tabletop interface in terms of effectiveness, ease of use, user satisfaction, privacy, and comfort.

H2: MobiSurf interface is at least as easy to learn as the standard tabletop interface.

H3: There are no distractions when switching between the mobile and tabletop devices for the MobiSurf interface.

#### B. Participating Subjects and Apparatus

Forty-four undergraduate students participated in the study. None of the participants had any prior experience with tabletop devices. The study design required each participant to perform

two tasks (i.e., user authentication and content sharing) using both the standard tabletop interface and the MobiSurf interface. A complete counterbalancing of the order of the tasks and the interfaces was performed to avoid the pitfalls of the “carryover” effect in a standard repeated measures design.

MobiSurf can be implemented on any kind of mobile devices. Without losing generality, a prototype was implemented on a Microsoft PixelSense Samsung SUR40 tabletop and Android mobile devices.

#### C. Experiment Design

The study began with a pre-study questionnaire followed by a training sessions on the standard tabletop and MobiSurf interfaces. Next, the participants were asked to perform two tasks, using each interface respectively. The experimenter recorded the completion time of each subject for each task on each method. After performing each task, the participants filled out a post-study questionnaire to provide feedback on their reaction to each interface.

**Step 1 – Prestudy Questionnaire:** During this step, we collected information from the participants about their reading skills, the prior knowledge of touch screen interfaces, the experience (i.e., whether they own a smartphone or not) and the comfort level with smartphones.

**Step 2 – Training Session:** The experimenter trained the participants on using a standard tabletop interface as well as a MobiSurf interface. The training session included the general description of both interfaces. Then, participants practiced a text entry task and a content sharing task, designed only for the training purpose, using both interfaces.

**Step 3 – Performing Tasks:** Researchers have designed and evaluated the authentication and content sharing tasks on tabletop devices [15, 17, 24, 25, 28]. These tasks indicated the common interaction applications on tabletops; thus, they were selected in our study to compare the usability of the MobiSurf interface against the standard tabletop interface. The participants were not allowed to ask for any help during the study. They performed the tasks based on written instructions to accomplish each task.

**Task #1- For this task,** we provided a standard login interface and assigned each subject with a random username-password combination that was of the same length for all the subjects.

**Task #2- Content Sharing.** In this task, each user downloaded a company’s mission statement, revised it, and finally sent it back to a coordinator. In a standard tabletop interface, email was used to transfer information between different devices. Each participant was asked to email a text block from the tabletop to a designated email address. Then, the participant opened the email on his/her mobile device to download the mission statement, edited it, and emailed the revised version to the experimenter’s email. In order to make a fair comparison, in the tabletop-only interaction, a user clicks a button to open an email interface, which was embedded and displayed in the same tabletop application, for transferring data. Alternatively, the MobiSurf interface used gestures to support the email function. First, each participant was asked to use the

PointerRotate gesture, which mimics the disk rotation for data exchange, to transfer the mission statement from a tabletop to his/her mobile device. After revising the statement, the updated version could be shared with the coordinator through the PointerShare gesture.

Step 4 – Survey Questionnaire: At the end of each task using each interface, participants were asked to fill an online survey questionnaire to rate his/her experience on using the interface.

#### D. Analysis and Results

An alpha value of 0.05 is selected to judge the significance of the results.

H1 – Usability of MobiSurf and Tabletop: To compare the usability of the treatments, each interface for each task was rated using a 5-point scale on five relevant characteristics, i.e., effectiveness, ease of use, user satisfaction, privacy, and comfort. Using the individual score, we calculated the median score for each treatment method on each characteristic separately for the user authentication task (See Table III) and the content sharing task (See Table IV).

TABLE III. RATINGS ON FIVE CHARACTERISTICS FOR TASK 1

Task 1 – User Authentication					
	Effectiveness	Ease of use	Satisfaction	Privacy	Comfort
MobiSurf	5	5	5	5	5
Tabletop	3	4	4	2	2

TABLE IV. RATINGS ON FIVE CHARACTERISTICS FOR TASK 2

Task 2 – Content Sharing					
	Effectiveness	Ease of use	Satisfaction	Privacy	Comfort
MobiSurf	5	5	5	5	4
Tabletop	3	3	4	3	3

In addition to the noticeable positive ratings for the MobiSurf interface, we also noticed the differences between the ratings on the two tasks. Specifically, the difference on the privacy characteristic between the two interfaces on task 1 is larger than that on task 2, which can be attributed to the nature of the tasks. That is, task 1 (user authentication) is more privacy concentrated compared with task 2 (content sharing) which implies that the MobiSurf interface is preferred by users with the privacy needs. In addition, we noticed that for all five characteristics, the median ratings for the MobiSurf interface were significantly higher than that for the tabletop interface. A Paired-sample Wilcoxon Signed-Rank test on each pair of rating values revealed significant differences between two methods for each characteristic (i.e.,  $p < 0.001$ ).

H2 – Ease of learning on MobiSurf and tabletop interfaces: MobiSurf introduces a new interaction style, which should not significantly increase the learning time. A Paired-sample Wilcoxon Signed-Rank test compared the median ratings for each method on each task. As presented in Table V, in task 1, the results showed that the participants rated the ease of learning on both interfaces equally positive, and the difference was not statistically significant (i.e.,  $p = 0.323$ ). In task 2, the participants rated the MobiSurf interface significantly easier to learn than the tabletop interface (i.e.,  $p = 0.012$ ). These results

verify our hypothesis that the MobiSurf interface is at least as easy to learn as the standard tabletop interface.

TABLE V. RATINGS ON EASE OF LEARNING

Ease of learning		
	Task 1 - User Authentication	Task 2 – Content Sharing
MobiSurf	5	5
Tabletop	5	4

H3 – Distraction: The MobiSurf interaction style requires a user to switch his/her focus between a mobile device and a tabletop. In the study, we specifically asked participants to rate the level of distraction they felt due to the device switch during an interaction session. In Table VI, which presents the median score, score 5 indicates “the user strongly agrees that the combination of a mobile device and a tabletop does NOT distract his/her attention while performing a task”. Based on the participants’ ratings, 39 of the 44 participants for tasks 1 and 36 participants for task 2 rated “agree” (with a score of 4) or “strongly agree” (with a score of 5) that they did not feel distraction using the MobiSurf interface. This result verified our hypothesis that the switching between devices in the MobiSurf interaction style does not distract users.

TABLE VI. RATINGS ON USER DISTRACTION

User Distraction	
Task 1 – User Authentication	5
Task 2 – Content Sharing	4

Fig. 1 provides an overview of participants’ efficiency results. It shows the mean and standard deviation (SD) of the task completion time for each task and interface combination. For both tasks, the SD values for both interfaces are relatively small and indicate consistency across the data. Regarding task 1, the subjects spent an average of 21 seconds using the tabletop interface vs. an average of 24 seconds using MobiSurf. Based on the researcher’s observations during the user study, the device switch in MobiSurf may result in longer time to complete task 1. Regarding task 2, there is a visible difference and significant improvement in the task completion time when using the MobiSurf interface (an average of 33 seconds) vs. tabletop interface (an average of 115 seconds).

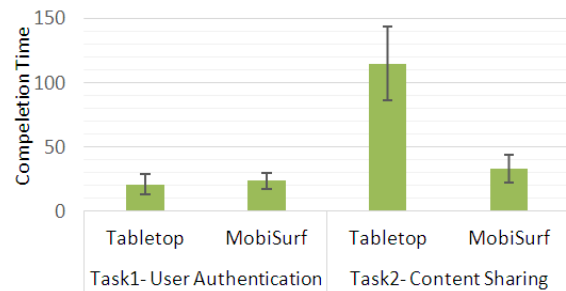


Figure 1. The mean values and SD of task completion time

We then used multiple regression [8] to find any statistically significant correlations between their background data and their efficiency when using the MobiSurf and tabletop interfaces. The results showed that the subjects’ experience of using smartphones (i.e., owning a smartphone) had a significant correlation to the efficiency using the MobiSurf

interface for both tasks 1 ( $p=0.0098$ ) and 2 ( $p=0.0462$ ). That is, subjects familiar with smartphones spent significantly less time to complete tasks 1 and 2 using the MobiSurf interface. In the MobiSurf interface, a smartphone was used as an external controller to supplement a tabletop interface. Therefore, the previous user experience for smartphones affects the efficiency for the MobiSurf interface.

In summary, MobiSurf provides a balanced interaction style that considers both usability and privacy for collaborative applications. Our results showed that the MobiSurf interface provided a seamless interaction between a personal device and a shared tabletop with minimal distraction. We identified two main limitations during the user study. First, the proposed approach utilizes tangible objects to track users over the display, but it is only applicable to horizontal displays where a user can comfortably place the tangible object on top of a tabletop. Modifications are necessary to extend our approach to vertical or wall mounted devices. Second, the use of infrared sensors to identify pointers' infrared tags limits our system to indoor environments.

## VI. APPLICATIONS

Based on the MobiSurf framework, interface designers can design inter-device bimanual interactions through the following steps.

1) **Design a standard tabletop interface.** Interface developers elicit requirements, and design a tabletop interface according to existing guidelines.

2) **Identify inter-device user interface (UI) elements.** After designing the standard tabletop interface, developers select UI elements that require inter-device interaction. For example, a textbox for a password within a standard tabletop interface is extended with an inter-device interaction by using a mobile device for a private input while some text blocks or images can be given with the sharing capacity across different devices.

3) **Define the intended action of a tabletop gesture on a UI element.** Developers must define the action of a gesture (e.g., "PointerOver" or "PointerRotate") on each inter-device UI element. According to the intended action, corresponding action message(s) are transferred to and performed on an associated mobile device. The mapping between a tabletop gesture on a UI element and its action provides the flexibility for interface developers to define the application-dependent semantics for user's actions.

Based on the above design process, we have designed three applications, discussed as follows.

Although tabletops are getting more and more popular as public interfaces, they lack accessibility features for blind users. Though commercial tools (e.g., Apple's Voice Over, and Google's Eyes-Free) have been proposed to support accessibility features on tabletop devices, they in general combine the gesture-based input with the speech-based output for blind users, which may be potentially limited in a public environment due to ambient noises. In addition, the voice input on a tabletop is only useful for single-user interactions because different sources of sound can interfere with each other.

Furthermore, most approaches require significant modifications on the hardware or software, which makes the new system inaccessible for non-blind users. Without compromising the usability for non-blind users, we developed a prototype (See Figure 2) to facilitate blind users to access information in a public environment. The prototype gives vibration and voice feedback to a visually impaired user through his/her personal mobile device. Therefore, one user's interaction does not interfere with others. More specifically, non-blind users interact with the tabletop application in a normal way. On the other hand, the prototype provides accessibility features for visually impaired users. The boundary in the following screenshot (Figure 2) includes a serial of red blocks that guide blind users to browse different parts of the interface. In the beginning, a blind user places his/her pointer on the top-left corner of the screen. Then, speech feedback, which directs the user to the destination, is delivered to the user's mobile device when the user moves the pointer along red blocks. The user is alerted through vibration if he/she accidentally moved the pointer off the track. When the user's pointer arrives at the destination, the application generates an action message "Speech" on the user's mobile to request the corresponding information. The user can input the required information through speech on his/her mobile, which does not interfere with other users.

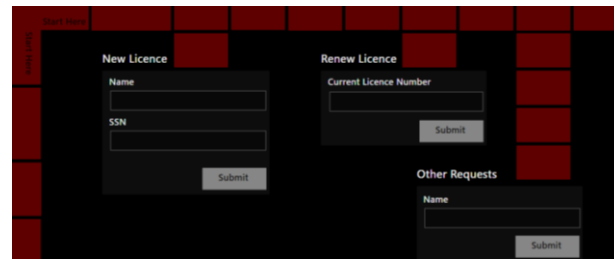


Figure 2. A screenshot for the accessible public interface

In addition to applications in a public environment for both disabled and normal users, MobiSurf is useful for collaboration, such as brain storming or class discussion. Taking advantage of the large screen of a tabletop, people can gather around a digital surface and interact with the system simultaneously. The MobiSurf interaction technique is especially useful to protect privacy and support data sharing in collaborative tasks. For example, a meeting coordinator can easily issue a private poll on a particular idea and others can privately send their responses to the coordinator through the "PointerShare" gesture. Furthermore, during the discussion, a participant can access or input sensitive information through his/her mobile device while he/she uses the tabletop as a public space for discussion and information sharing. Another potential application is to support class discussion in educational applications. For example, an instructor can use a public display to share contents with students (i.e., the mobile-to-tabletop sharing) while students can use private sharing to send back their answers to the instructor (i.e., the mobile-to-mobile sharing). Voice, vibration and other multimodal feedback can add more interactivity.

## VII. CONCLUSION AND FUTURE WORK

This paper presents the MobiSurf framework that supports inter-device bimanual interactions in different scenarios through tangible objects and mobile devices. MobiSurf is featured with an event-driven architecture, which maps user's actions on a tabletop to corresponding interaction commands on a paired mobile device. We investigated different scenarios, to which the MobiSurf interaction style is applied. A user study evaluated the usability of the MobiSurf interaction style. The results indicated that the proposed interaction style was as usable as the standard multi-touch interaction while offering improved privacy and efficient data sharing. In the user study, some advanced features, such as "Region selection" or "Multiple pointers", were not evaluated. We will evaluate these features in future experiments. The future work also includes developing complex real-world applications, especially focusing on improving the accessibility for vision-impaired users on tabletop interfaces, and evaluating their usability. In addition, we plan to evaluate the ease of instantiating the framework in applications.

## REFERENCES

- [1] C. J. Ackad, A. Clayphan, R. M. Maldonado and J. Kay, "Seamless and Continuous User Identification for Interactive Tabletops Using Personal Device Handshaking and Body Tracking", Proc. CHI EA, 2012.
- [2] R. Ballagas, M. Rohs, and J. Borchers, "Sweep and Point & Shoot: Phocam-based interactions for large public displays", CHI Ext. Abstracts, 2005.
- [3] M. I. Berkman and A. Karahoca, "A direct touch table-top display as a multi-user information kiosk: Comparing the usability of a single display groupware either by a single user or people cooperating as a group", *Interacting with Computers*, vol. 24, no 6, pp. 461-471, 2012.
- [4] S. Boring, D. Baur, A. Butz, S. Gustafson, and P. Baudisch, "Touch projector: Mobile interaction through video", Proc. CHI, 2010.
- [5] Buxton, W., Hill, R., and Rowley, P. "Issues and techniques in touch-sensitive tablet input", Proc. SIGGRAPH, 1985.
- [6] R. Dachsel and R. Buchholz, "Natural throw and tilt interaction between mobile phones and distant displays", CHI Ext. Abstracts, 2009.
- [7] F. Echter, S. Nestler, A. Dippon and G. Klinker, "Supporting casual interactions between board games on public tabletop displays and mobile devices", *Personal and Ubiquitous Computing*, vol. 13, no. 8, pp. 609 - 617, November 2009.
- [8] A. Field, "Discovering Statistics using IBM SPSS Statistics", 2nd edition, 2007.
- [9] S. Greenberg, M. Boyle and J. Laberge, "PDAs and Shared Public Displays: Making Personal Information Public, and Public Information Personal", *Pers. and Ubiqu. Comp.*, vol. 3, p. 54-64, 1999.
- [10] R. Hardy and E. Rukzio, "Touch & Interact: Touch-based interaction of mobile phones with displays", Proc. MobileHCI, 2008.
- [11] T. Hesselmann, N. Henze and S. Boll, "FlashLight: Optical Communication between Mobile Phones and Interactive Tabletops", Proc. ITS, 2010.
- [12] W. Hutama, P. Song, C. Fu and W. Goh, "Distinguishing Multiple Smart-Phone Interactions on a Multi-touch Wall Display using Tilt Correlation", Proc. CHI, 2011.
- [13] H. Ishii and B. Ullmer, "Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms", Proc. CHI, 1997.
- [14] S. Kane, M. Morris, A. Perkins, D. Wigdo, R. E. Ladne and J. Wobbrock, "Access Overlays: Improving Non-Visual Access to Large Touch Screens for Blind Users", Proc. UIST, 2011.
- [15] D. Kim, P. Dunphy, P. Priggs, J. Hook, J. Nicholson, J. Nikolson and P. Olivier, "Multi-Touch Authentication on Tabletops", Proc. CHI, 2010.
- [16] A. De Luca and B. Frauendienst, "A privacy-respectful input method for public terminals", Proc. NordiCHI, 2008.
- [17] A.J. Maunder, G. Marsden and R. Harper, "SnapAndGrab – Accessing and sharing contextual multi-media content using Bluetooth enabled cameraphones and large situated displays", Proc. CHI, 2008.
- [18] C. McAdam and S. Brewster, "Distal Tactile Feedback for Text Entry on Tabletop Computers", Proc. BCS-HCI, 2009.
- [19] C. McAdam and S. Brewster, "Using Mobile Phones to Interact with Tabletop Computers", Proc. ITS, 2011.
- [20] K. Miyaoku, S. Higashino, and Y. Tonomura, "C-Blink: A hue-difference-based light signal marker for large screen interaction via any mobile terminal", Proc. UIST, 2004.
- [21] J. Muller, M. Jentsch, C. Kray, and A. Kruger, "Exploring factors that influence the combined use of mobile devices and public displays for pedestrian navigation", Proc. NordiCHI, 2008.
- [22] B. A. Myers, "Using handhelds and PCs together", *Comm. ACM*, 44:34-41, 2001.
- [23] A. Olwal and S. Feiner, "Spatially aware handhelds for high-precision tangible interaction with large displays", Proc. TEI, pp. 181-188, 2009.
- [24] S. N. Patel, J. S. Pierce and G. D. Abowd, "A gesture-based authentication scheme for untrusted public terminals", Proc. UIST, 2004.
- [25] N. Pears, D.G. Jackson, and P. Oliver, "Smart phone interactions with registered displays", *IEEE Perv. Comp.*, 8:14-21, 2009.
- [26] R. Ramakers, D. Vanacken, K. Luyten, K. Coninx, J. Schoning, "Carpus: A Non-Intrusive User Identification Technique for Interactive Surfaces", Proc. UIST, 2012.
- [27] V. Roth, P. Schmidt, B. Guldenring, "The IR Ring: Authenticating Users' Touches on a Multi-Touch Display", Proc. UIST, 2010.
- [28] J. Schöning, M. Rohs and A. Krüger, "Using Mobile Phones to Spontaneously Authenticate and Interact with Multi-Touch Surfaces", Workshop on designing multi-touch interaction techniques for coupled private and public displays, 2008.
- [29] D. Schmidt, F. Chehimi, E. Rukzio and H. Gellersen, "PhoneTouch: A Technique for Direct Phone Interaction on Surfaces", Proc. UIST, 2010.
- [30] D. Schmidt, M. Chong, and H. Gellersen, "HandsDown: Hand-contour-based user identification for interactive surfaces", Proc. NordiCHI, 2010.
- [31] D. Schmidt, J. Seifert, E. Rukzio and H. Gellersen, "A Cross-Device Interaction Style for Mobiles and Surfaces", Proc. DIS, 2012.
- [32] J. Seifert, A. L. Simeone, D. Schmidt, C. Reinartz, P. Holleis, M. Wagner, H. Gellersen and E. Rukzio, "MobiSurf: Improving Co-located Collaboration through Integrating Mobile Devices and Interactive Surfaces", Proc. ITS, 2012.
- [33] A. S. Shirazi, T. Döring, P. Parvahan, B. Ahrens and A. Schmidt, "Poker Surface: Combining a Multi-Touch Table and Mobile Phones in Interactive Card Games", Proc. Mobile HCI, 2009.
- [34] A. Wiethoff, R. Kowalski and A. Butz, "inTUIt – Simple Identification on Tangible User Interfaces", Proc. TEI, 2011.
- [35] A. D. Wilson and R. Sarin, "BlueTable: Connecting wireless mobile devices on interactive surfaces using vision-based handshaking", Proc. GI, 2007.

# SWOWS and Dynamic Queries to build Browsing Applications on Linked Data

Paolo Bottoni

Sapienza, University of Rome  
Department of Computer Science  
bottoni@di.uniroma1.it

Miguel Ceriani

Sapienza, University of Rome  
Department of Computer Science  
ceriani@di.uniroma1.it

## Abstract

*The Linked Data Initiative is pushing dataset maintainers to publish data online in a highly reusable way through a set of open standards, such as RDF and SPARQL. The amount and variety of available structured data in the Web is increasing but its consumption is still quite limited. In particular applications used to explore linked data are mostly either generic linked data browsers or applications with hard-coded logic tailored for specific needs. SWOWS is a platform for declarative specification of applications consuming linked data. In this paper we describe the use of the platform for creating browsing applications tailored to specific contexts, and show how the declarative paradigm preserves flexibility of the application. To this end, the platform has been extended to allow the dynamic generation of SPARQL queries. An example of a linked data browser created with the platform is given.*

## 1. Introduction

In recent years the Web is evolving from an interlinked set of documents to an interlinked web of data and services. The structured data available online is increasing both in quantity and diversity [7]. The comprehensive data model proposed by the World Wide Consortium (W3C) is the Resource Description Framework (RDF) [14]. Other de facto standards, e.g. the Freebase model [8], or HTML embedded formats, such as Microformats [25], Microdata [22] or RDFa [1], can be mapped to the Resource Description Framework (RDF) model [14] for interoperability (for example using Any23<sup>1</sup>).

One of the key advantages of the linked data model is that it allows the serendipitous exploration and reuse of existing data. Potentially, any expert of a specific domain can build a fully customized visualization from a set of possibly different linked data sources.

<sup>1</sup><http://any23.apache.org/>

In practice, building such a visualization currently requires advanced programming skills. Moreover it involves a number of system choices that constrain the effective reuse of a visualization. To mitigate this problem, we have developed the Semantic Web Open dataflow System (SWOWS) [11, 10], a platform that allows the declarative construction of interactive linked data applications. Applications are built from a basic set of operators (based on SPARQL, the standard query language over the RDF model [21]) adopting the pipeline metaphor.

In this paper we discuss the use of SWOWS to build applications to browse specific sets of linked data. We show how such applications can be built based purely on Web and Semantic Web standard technologies and how the declarative approach can help in organizing flexible visualizations.

The SWOWS platform has been extended to allow manipulation of SPARQL queries as data. A developer using SWOWS can thus write SPARQL queries which take as input other SPARQL queries and/or produce new queries in output. The queries written with SWOWS have the role *higher functions* have in other programming languages (Scala [30] being possibly the most popular nowadays). The use of this functionality (together with other extensions to the original SWOWS platform) will be showcased through an example application.

In the rest of the paper, Sect. 2 introduces the technology background and Sect. 3 discusses related work, while Sect. 4 specifically describes SWOWS. New features are shown in Sections. 5 and 6, where dynamic query generation is illustrated. An example application is described in Sect. 7 and Sect. 8 discusses conclusions and future work.

## 2. Technologic Background

The relational model is widely used to represent virtually any kind of structured information. The **Resource Description Framework (RDF)** [14] generalises it to the universe of structured data in the World Wide Web, better known as the Semantic Web [5]. In the RDF data model, knowledge is represented via *RDF statements* about *resources*, where a

resource is an abstraction of any piece of information about some domain. A RDF statement is represented by a *RDF triple*, composed of *subject* (a resource), *predicate* (specified by a resource as well) and *object* (a resource or a literal, i.e. a value from a basic type). A *RDF graph* is therefore a set of RDF triples. Resources are uniquely identified by a Uniform Resource Identifier (URI) [4], or by a local (to the RDF graph) identifier if they have not meaning outside of the local context (in which case they are called *blank nodes*). The resources used to specify predicates are called *properties*. A resource may have one or more *types*, specified by the predefined property `rdf:type`. A *RDF dataset* is a set of graphs, each associated with a different name (a URI), plus a default graph without a name. We use RDF through the framework to represent any kind of information and its transformations. In RDF, *prefixes* can be used in place of the initial part of a URI, representing specific namespaces for vocabularies or set of resources.

We extensively use **SPARQL**<sup>2</sup> [21], the standard query language for RDF datasets. SPARQL has a relational algebra semantics, analogous to those of traditional relational languages, such as Structured Query Language (SQL). The SPARQL `CONSTRUCT`, one of the SPARQL query forms, takes as input a RDF dataset and produces a RDF graph. While the SPARQL Query Language is “read-only”, the SPARQL Update Language [33] defines a way to perform updates on a *Graph Store*, the “modifiable” version of a RDF Dataset. A SPARQL Update *request* is composed of a number of *operations*. The current version of the standard is SPARQL 1.1, but much of the existing work refers to the previous version, SPARQL 1.0 [32]. SPARQL 1.1 algebra offers an expanded set of operators, effectively allowing the expression of queries that were not expressible before.

The ubiquity of Web browsers and Web document formats across a range of platforms and devices drives developers to build applications on the Web and its standards. Requirements for browsers have dramatically changed from the first days of the Web. Now a browser is an interface to an ever-growing set of client capabilities, exemplified by **Rich Web Client** Activity at W3C. All modern browsers natively support the Scalable Vector Graphics (SVG) standard [15], a language representing mixed vector and raster content, based on Extensible Markup Language (XML) [12]. Together with the long established Document Object Model (DOM) Events [31, 23] and ECMAScript<sup>3</sup> [16] support, it allows the realisation of complete interactive visualisation applications. Indeed, ECMAScript libraries for interactive data visualization are proliferating, from standard visualisations [19, 3, 9] to specialized visualisations for specific domains [36], especially leveraging the SVG technology.

<sup>2</sup>Originally a recursive acronym SPARQL Protocol and RDF Query Language, the extended form has then been dropped from W3C documents

<sup>3</sup>Commonly called JavaScript, the dialect from Mozilla Foundation.

### 3. Related Work

Several languages were proposed to define **SPARQL views** in a way analogous to SQL views. A SPARQL view is a graph intensionally defined by a SPARQL `CONSTRUCT` query; the input dataset can be composed by both “real” (extensionally defined) graphs and views. *RVL* [28] is an early effort, using an imperative language for defining views based on an independently defined query language (RQL [24]). *vSPARQL* [35] is an extension of SPARQL 1.0 grammar allowing named views defined with `CONSTRUCT` queries and reusable in other queries. Schenk and Staab, working on *Networked Graphs* [34], propose an RDF-based syntax to define views, which are graphs defined in terms of SPARQL 1.0 `CONSTRUCT` queries on explicitly defined graphs and other views. Although powerful enough to define read-only applications (possibly together with visualization tools described below), network of views do not easily model interactive applications. In particular, they face the problem of how to represent events and time-dependent information, including the application state.

Two **pipeline languages** have been proposed to define RDF transformations, namely DERI Pipes [27] and SPARQLMotion [26]. They offer a set of basic operators on RDF graphs to build the pipelines and they are both endowed with a graphical environment to create the pipelines using the available operators (free in the case of DERI Pipes, in a commercial software for SPARQLMotion [13]). In SPARQLMotion, both pipelines and queries are represented in RDF (for queries using the SPIN-SPARQL [18] syntax).

Visualbox [20] and Callimachus [2] have been proposed for **linked data visualisation**. In their two-step model/view approach, SPARQL queries select data and a template language generates the (XML-based) visualisation. SPARQL Web Pages (SWP) is a RDF-based framework (to be used with SPARQL Motion or on its own) to describe and render HTML+SVG visualisations of linked data. HTML and SVG are mapped to two corresponding vocabularies and together with the UISPIN Core Vocabulary allow the association of a RDF resource with the description of its visualisation. The description may be also statically associated with a class of resources, with each specific resource mapping defined through a SPARQL query. In all these proposals the execution model corresponds to managing a single HTTP request, as with typical application server technologies like Java Servlet or PHP. Persistence and logical relationships between requests and client state must be managed explicitly (e.g. saving/loading data related to a session and encoding parameters in requests)<sup>4</sup>.

Generation/manipulation of queries at runtime is widely

<sup>4</sup>Both Callimachus and SWP offer some aid for building interactive applications via special functions and syntaxes, but the execution model remains request oriented.



used both in SQL and in SPARQL. Several systems (for a SPARQL example, see Jena [29]) provide basic support through *parameterized queries*, in which some *parameters* are bound (through some external mechanism) to actual (scalar) values at execution time. This mechanism is not sufficient when the structure of the query must be changed dynamically (e.g. for multiple or complex search criteria and/or ordering rules). Generic **dynamic query generation** is usually achieved through string-based, semantically-unaware, manipulation or through a programmatic interface offered by the host language. There have been efforts to represent queries using semantically rich structures not tied to a specific host language [37], specifically for SPARQL the already mentioned SPIN-SPARQL [18] vocabulary. While this vocabulary potentially allows dynamic query generation, its use for this purpose is documented only for a specific case of query rewriting [17]<sup>5</sup>.

#### 4. SWOWS

We set the following **requirements for SWOWS**: to be based on a *dataflow language* in which data transformations are represented as pipelines; to use pipelines through cascading *declarative views* on the input or other views; to represent data as *RDF*; to be able to connect to existing *RDF sources*; to exhibit *interactivity* through Web interface input/output; to represent pipelines *as RDF* to share and reuse; to support interoperability with *XML*; to use *existing standards* whenever possible.

The **platform** we propose allows users to define linked data applications through a pipeline language based on RDF. Users create pipelines using a visual representation, through a Web-based *editor*, in turn interacting with a *pipeline repository*. Other software, called the *dataflow engine*, executes the pipeline (after downloading the corresponding RDF Graph from the repository) on a possibly separate server, with a Web-based interface as well. Figure 1 shows a simplified lifecycle of the pipeline, from editing and saving it in a directly controlled repository, to eventually sharing it for use “as it is” or reuse in other pipelines.

A **SWOWS pipeline** is a side-effect-free dataflow programming module, taking as input an RDF Dataset and returning another RDF Dataset. The available components (shown in Figure 2) are: the *default input graph* and the *(named) input graphs*; the *default output graph* and the *(named) output graphs*; the *transform processors*, that execute a SPARQL 1.1 [21] query against a RDF dataset; the *single graph stores*, whose content is incrementally modified during an execution of the pipeline by executing a SPARQL 1.1 Update [33] on it each time one of its input graphs changes (the update takes as input a RDF

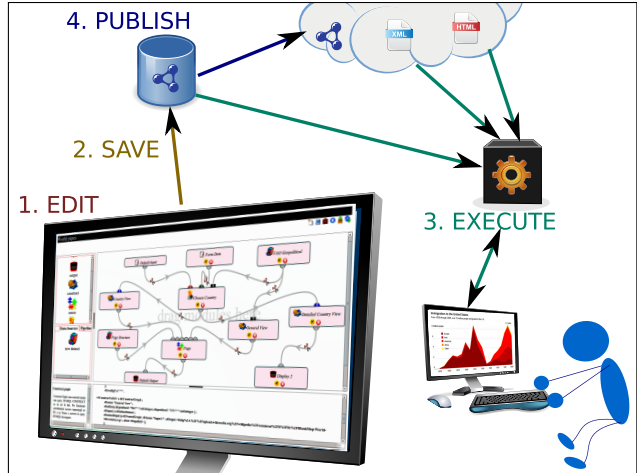


Figure 1. A schematic view of the platform

dataset composed by the previous snapshot of the store as default graph and a set of input named graphs); existing *pipelines* which can be used as components in the current pipeline; *file data sources*, i.e. RDF graphs generated by loading local or remote files (serialized in one of the standard RDF formats); another way to access data from outside the pipeline is through *SPARQL Federated Queries* used in transform processors or simple graph stores.

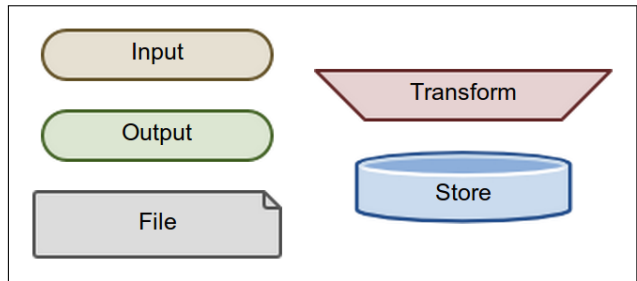


Figure 2. The pipeline base component types

A pipeline can be designed just for reuse by other pipelines. If a pipeline has to be executed (i.e. it is a *top level* pipeline), its default output graph must comply with an *XML DOM Ontology*<sup>6</sup> in RDF. It will represent a HTML or SVG document, to be rendered by the user interface. Its default input graph will receive the DOM Events generated in the interface, described with a *DOM Events Ontology*<sup>7</sup>.

The main blocks of the **implemented application** are the *editor*, the *pipeline repository* and the *dataflow engine*. The editor is a rich Web application with its client side logic coded in HTML+CSS+JavaScript and embedded

<sup>5</sup>SPIN-SPARQL vocabulary is widely used as a way to attach SPARQL rules to RDF resources.

<sup>6</sup><http://www.swows.org/2013/07/xml-dom>

<sup>7</sup><http://www.swows.org/2013/07/xml-dom-events>

in the Callimachus Web application [2], used as pipeline repository. The dataflow engine is a Java-based (using Apache Jena [29]) Web application maintaining the state of each running pipeline instance; when a new instance is launched (e.g., from the editor) the engine initialises the pipeline and returns its output to the client, along with a piece of JavaScript logic to report handled events back to the server; each time an event is fired on the client, the dataflow engine is notified and answers with the changes to be executed on the client content. On the client side, any browser supporting JavaScript can use both the editor and the generated application. The software is freely available<sup>8</sup>.

## 5. Transform Processors

In the previous version of the platform the behaviour of a transform processor was always specified through a `CONSTRUCT` query. It has been changed to allow also for a `SELECT` query or a `UPDATE` request, maintaining the fact that the transform processor executes a stateless operation on a RDF dataset, resulting in a single RDF graph.

The `SELECT` queries are run according to the corresponding semantics, but the result set is represented as a RDF graph, containing the bindings, their order and each value associated with a binding. This allows the derivation of some order (depending on the `ORDER BY` clause) over a set of items. Ordering could also be achieved through `CONSTRUCT` queries but in a contrived and potentially inefficient way.

The `UPDATE` requests are run considering the input dataset as a Graph Store; the output is given by the content of the default graph after the `UPDATE` (the changes to the other graphs are discarded). The `UPDATE` is useful if most of the input default graph must be copied to the output, as using `UPDATE` only the changes need to be expressed.

## 6. Dynamic Queries

One of the advantages of a declarative platform based on a flexible model such as RDF is to easily generalise functions. For example if one wants to generalise an operation of filtering or aggregation on some property, one can take the property from a configuration graph or as result of another query: properties in RDF are resources and can be treated as such. In practice, due to some limitations in SPARQL expressivity (e.g., there is no direct way to express a `FORALL` operator), it can be tricky to build generalised queries; moreover, due to current limits in SPARQL engines and protocol infrastructure, they can be highly unefficient. For these reasons we propose **dynamic query generation**, using a RDF vocabulary to represent any SPARQL query. The vocabulary

<sup>8</sup><http://www.swows.org/>

reuses SPIN-SPARQL [18] classes and properties (namespace <http://spinrdf.org/sp#>, prefix `sp:`). Some properties (namespace <http://www.swows.org/spinx#>, prefix `spx:`) are added to avoid the use of lists when the order in the query structure is not important. As an example, whenever SPIN-SPARQL uses the `sp:elements` property to associate a list of elements with a SPARQL operator in a query (e.g., `ex:op1 sp:elements (ex:el1 ex:el2)`), we use the `spx:element` property to associate each single element with the given operator (e.g., `ex:op1 spx:element ex:el1, ex:el2`). This choice simplifies writing queries that manipulate queries. The namespace <http://www.swows.org/spinx/var/> (prefix `var:`) is further defined to contain all the used variables (e.g., variable `?foo` represented as `var:foo`) in order to simplify definition and comparison of query variables.

## 7. An Example: Exploring Artworks

We will demonstrate in practice how a simple Linked Data browsing application can be built with the proposed platform and methodology. Our aim is to build an interface to visually navigate through a set of world artworks. There are a number of contexts in which it is important to offer a visual navigation interface, reducing the need to use text to interact with the application. One such context is a device available in a public space, that may be designed for casual use and international audience. The data used comes from Freebase, a community-curated (and derived from other online sources like Wikipedia) database of well-known people, places, and things. The data is rich enough to allow for faceted navigation using different dimensions: author, art form (painting, sculpture, etc.), art genre (portrait, allegory, etc.), creation date (usually at least the year).

The usual approach for faceted navigation is to use the different categories to build multi taxonomy filters for searching items. This approach is indeed established and effective; we want to experiment instead with the use of different dimensions to guide the user in free exploration of dataset item by item. This kind of navigation can be useful when the user has no previous complete knowledge of the used taxonomies and/or the user interface will be used for casual exploration, wandering through a data set. Usually this is provided through related items links, that provide a single dimension for this kind of exploration.<sup>9</sup>

The user interface for the application is composed of: a *main area* in which an image is shown together with information about the currently selected item; a variable number of *related artworks areas*, in which links to related artworks

<sup>9</sup>In the present work we present the idea of multifaceted visual exploration just to show the feasibility of the proposed approach for design of browsing applications. We think that the idea on its own deserves to be studied and tested in future research.

(visually represented as thumbnails) are shown according to different axes (in Figure 3 one of the possible configurations of the application is shown<sup>10</sup>). The pipeline design has been based on the separation between presentation elements (position and sizes of the different areas) and content elements (different filters used for related artworks). A specific graph defines the association between the two kinds of elements.



**Figure 3. A screenshot of the example application, here with the main area in the center and related artwork areas at the four sides.**

For this example we consider the Freebase dataset to be available from a SPARQL Endpoint, considering the standard mapping defined by Freebase maintainers. It is possible to use the public FactForce endpoint<sup>11</sup> (aggregating also several other datasources [6]), or to install a cloud-based distribution as the Sindice Freebase Distribution<sup>12</sup> for a more reliable access.

The pipeline, shown in Fig. 4, is composed by the following components: some Data Sources corresponding to queries: one for each available filter (here four, but could be any number) and one, *filters-common.rq*, to be used as template to dynamically build the query; *Framing* that represents the wireframe of the user interface, both as SVG structure and as specification of the areas for the other components; *FiltersQuery* that generates the dynamic query for the areas based on *AreaFilter*, the association between areas and filters; *SelArtwork*, that stores the selected artwork, setting the initial one and changing it when the thumbnail for another one is clicked (events coming from *Default Input*); *RelatedArtwork*, that runs the dynamic query to obtain the sets of related artworks; *View*, that creates the dynamic part of the *Default Output* (the SVG) to be merged with the static part of the visualization from *Framing*.

The “heart” of the pipeline is the generation and execution of the query that retrieves the data of related artworks for each area based on the defined areas, the defined filters

(the queries) and the association between them. The areas are defined like the following (Turtle syntax):

```
1 <#bottomBarArea>
2   a <#ImageArea>;
3   <#svgElement> <#bottomBar>;
4   <#x> 20;
5   <#y> 550;
6   <#width> 800;
7   <#height> 70;
8   <#cols> 7;
9   <#rows> 1.
```

The queries representing the available filters are defined in the following style (this is same-genre.rq):

```
1 PREFIX aw: <http://rdf.freebase.com/ns/visual-art.artwork.>
2
3 SELECT DISTINCT ?relatedArtwork
4 WHERE {
5   ?selectedArtwork
6     aw:artist ?artist;
7     aw:art_genre ?genre;
8     aw:art_form ?form.
9   ?relatedArtwork
10    aw:artist ?diffArtist;
11    aw:art_genre ?genre;
12    aw:art_form ?diffForm.
13   FILTER(?diffArtist != ?artist).
14   FILTER(?diffForm != ?form).
15 }
```

Any kind of *SELECT* query will be accepted as long as it projects the variable *?relatedArtwork* and use the variable *?selectedArtwork* corresponding to the selected artwork.

Finally the association between areas and filters is defined through a graph like the following one:

```
1 <#topBarArea> <#filter> <same-artist.rq>.
2 <#bottomBarArea> <#filter> <same-genre-form.rq>.
3 <#leftBarArea> <#filter> <same-genre.rq>.
4 <#rightBarArea> <#filter> <contemp.rq>.
```

The *FiltersQuery* component generates the dynamic query and has as inputs various named graphs: *<#filters>*, from the file data sources corresponding to the queries defining the filters (the RDF graphs, being connected to the same input, are merged, but each query is still distinguishable, as each query root element corresponds to the original query URI); *<#common>*, from a file data source corresponding to *<filter-common.rq>*, a query used to hold constraints that should be satisfied by any filter (e.g. having at least an image and a minimum set of information); *<#areas>*, from *AreaFilters* component, giving the association between areas and filters described above; *<#framing>*, from *Framing* component, holding the definition of the areas, also described above. The default graph is given by the union of *<#common>* and *<#filters>*. *FiltersQuery* is defined an Update request; it is useful to recall that the output graph is obtained from the default input graph, applying on it the Update request. The Update request is the following (prefix declaration omitted):

```
1 INSERT {
2   <#FiltersQuery>
3   a sp:Select ;
4   sp:resultVariable var:area , var:relatedArtwork;
5   sp:where <#filterUnion>;
```

<sup>10</sup>Some images are not shown due to broken URLs in the dataset.

<sup>11</sup><http://factforge.net/sparql>

<sup>12</sup><http://sindicetech.com/freebase>

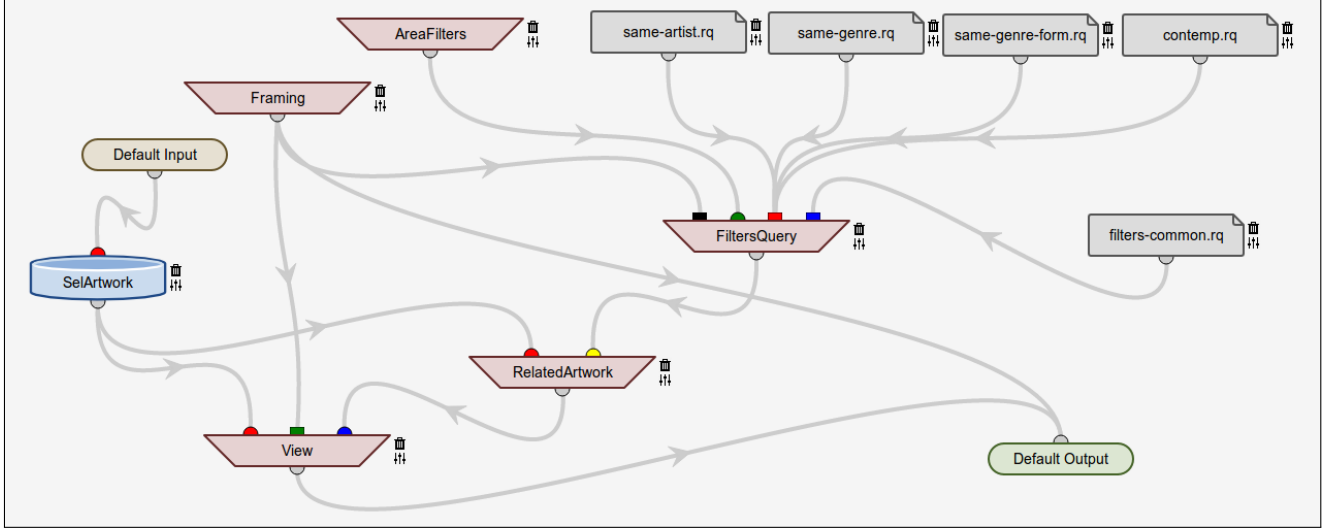


Figure 4. The pipeline of the example application.

```

6      sp:orderBy var:area.
7      <#filterUnion> a sp:Union.
8      <#selected>
9      a sp:TriplePattern;
10     sp:subject <#global>;
11     sp:predicate <#selectedArtwork>;
12     sp:object var:selectedArtwork.
13 }
14 WHERE {};
15
16 INSERT {
17   <#filterUnion> spx:element
18   [ a spx:ElementGroup;
19     spx:element
20     <#selected>,
21     [ a sp:Service;
22       sp:serviceURI <http://factforge.net/sparql>;
23       spx:element
24       [ a sp:SubQuery; sp:query ?filter ] ] ].
25   ?filterWhere spx:element
26   [ a sp:SubQuery;
27     sp:query <filters-common.rq> ].
28   ?filter
29     spx:resultVariable
30     [ sp:as var:area; sp:expression ?area ],
31     var:selectedArtwork;
32     sp:limit ?filterLimit.
33 }
34 WHERE {
35   GRAPH <#framing> {
36     ?area
37     a <#ImageArea>;
38     <#cols> ?areaCols; <#rows> ?areaRows. }.
39   GRAPH <#areas> { ?area <#filter> ?filter. }.
40   BIND(?areaCols*?areaRows AS ?filterLimit).
41   GRAPH <#filters> { ?filter sp:where ?filterWhere. }.
42 };

```

The Update request is composed by two separate Update operations. The first one (lines 1–13) is constant (in fact the `WHERE` clause is empty) and builds the fixed part of the query to be generated. The second Update request depends from the association between areas and filters (line 38) to build the part of the query corresponding to each filter and consisting of a common part (lines 16–21) and having as subquery the query corresponding to the filter (lines

22–24). That filter query is also enriched in some ways: the `filter-common.rq` query is added as a subquery (lines 25–26); a `LIMIT` clause (`sp:limit`) is added to retrieve exactly the number of items that will be shown in each area (lines 34–37, 39, 31); the corresponding area is added to the variables projected by each filter query (lines 28, 19); the queries are joined to the selected artwork through projecting the variable for each query (line 31) and joining each query with a constant part that reads the selected artwork from the input of the generated query (lines 8–11, 19).

As the output of this component is connected to the input named `#query` of *RelatedArtwork*, this query is executed in that component, taking as input the selected artwork. The purpose of using a `SELECT` query is having the items ordered. Even in the case the order of the related artworks in an area is undefined (the tuples are ordered only with respect to `?area`), the `SELECT` result gives an arbitrary order that can be used to draw the corresponding thumbnails. Finally, the `CONSTRUCT` query in the component *View* (not shown), builds the SVG page based on an image and information on the selected artwork and the thumbnails of related artworks organized by areas.

## 8. Conclusions and Future Work

We discussed the use of the platform SWOWS and dynamic query generation to build interactive linked data applications, especially in the case of flexible browsing applications. Dynamic query generation and other extensions to the platform as originally conceived were presented, justified by application to concrete cases.

We want to keep experimenting in this direction, and possibly also to leverage this experimentation to build

higher level interfaces, designed also for usage by non-expert users, as a way to flexibly interact with linked data.

## References

- [1] B. Adida, M. Birbeck, S. McCarron, and I. Herman. RDFa Core 1.1 - 2nd Edition. W3C Recommendation 22 August 2013.
- [2] S. Battle, D. Wood, J. Leigh, and L. Ruth. The Callimachus Project: RDFa as a Web Template Language. In *Proc. COLD 2012*, volume 905 of *CEUR*, 2012.
- [3] N. G. Belmonte. JavaScript InfoVis Toolkit. <http://philobg.github.io/jit/>, 2011.
- [4] T. Berners-Lee, R. Fielding, and L. Masinter. Uniform Resource Identifier (URI): Generic Syntax. RFC 3986 (INTERNET STANDARD), Jan. 2005. Updated by RFC 6874.
- [5] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.
- [6] B. Bishop, A. Kiryakov, D. Ognyanov, I. Peikov, Z. Tashev, and R. Velkov. Factforge: A fast track to the web of data. *Semantic Web*, 2(2):157–166, 2011.
- [7] C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *Int. J. on Semantic Web and Information Systems*, 5(3):1–22, 2009.
- [8] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proc. ACM SIGMOD'08*, pages 1247–1250. ACM, 2008.
- [9] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics*, 17:2301–9, 2011.
- [10] P. Bottoni and M. Ceriani. A Dataflow Platform for In-silico Experiments Based on Linked Data. In *Proc. DNIS 2014*, pages 112–131, 2014.
- [11] P. Bottoni, M. Ceriani, and S. Valentini. A user interface to build interactive visualizations for the semantic web. In *ISWC 2013 (Posters & Demos)*, pages 165–168, 2013.
- [12] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, F. Yergeau, and J. Cowan. Extensible Markup Language (XML) 1.1 (Second Edition). W3C Recommendation 16 August 2006, edited in place 29 September 2006.
- [13] T. COMPOSER. TOPBRAID COMPOSER 2007 Features and getting Started Guide Version 1.0, created by TopQuadrant, US, 2007.
- [14] R. Cyganiak, D. Wood, and M. Lanthaler. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014.
- [15] E. Dahlström, P. Dengler, A. Grasso, C. Lilley, C. McCormack, D. Schepers, J. Watt, J. Ferraiolo, J. Fujisawa, and D. Jackson. Scalable Vector Graphics (SVG) 1.1 (Second Edition). W3C Recommendation 16 August 2011.
- [16] ECMA. EcmaScript language specification, standard ecma-262, 5.1 edition. <http://www.ecma-international.org/ecma-262/5.1/>, 2011.
- [17] C. Follenfant, O. Corby, F. Gandon, D. Trastour, et al. Rdf modelling and sparql processing of sql abstract syntax trees. In *PSW-Ist Workshop on Programming the Semantic Web*, 2012.
- [18] C. Fürber and M. Hepp. Using SPARQL and SPIN for data quality management on the semantic web. In *Business Information Systems*, pages 35–46. Springer, 2010.
- [19] Google. Google charts, 2010.
- [20] A. Graves. Creation of visualizations based on linked data. In *Proc. WIMS'13*, page 41. ACM, 2013.
- [21] S. Harris et al. SPARQL 1.1 Query Language. W3C Recommendation 21 March 2013.
- [22] I. Hickson. HTML Microdata. W3C Working Group Note 29 October 2013.
- [23] G. Kacmarcik, T. Leithead, J. Rossi, D. Schepers, B. Hhrmann, P. Le Hgaret, and T. Pixley. Document Object Model (DOM) Level 3 Events Specification. W3C Recommendation 13 November 2000.
- [24] G. Karvounarakis, A. Magkanaraki, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl, and K. Tolle. RQL: A functional query language for RDF. In *The Functional Approach to Data Management*, pages 435–465. Springer Berlin Heidelberg, 2004.
- [25] R. Khare and T. Çelik. Microformats: a pragmatic path to the semantic web. In *Proc. WWW'06*, pages 865–866. ACM, 2006.
- [26] H. Knublauch et al. SPARQLMotion Specifications, 2010. [sparqlmotion.org](http://sparqlmotion.org).
- [27] D. Le-Phuoc, A. Polleres, M. Hauswirth, G. Tummarello, and C. Morbidoni. Rapid prototyping of semantic mash-ups through semantic web pipes. In *Proc. WWW '09*, pages 581–590. ACM, 2009.
- [28] A. Magkanaraki, V. Tannen, V. Christophides, and D. Plexousakis. Viewing the Semantic Web through RVL Lenses. In *Proc. ISWC 2003*, volume 2870 of *LNCS*, pages 96–112. Springer, 2003.
- [29] B. McBride. Jena: a semantic Web toolkit. *Internet Computing, IEEE*, 6(6):55–59, Nov/Dec 2002.
- [30] M. Odersky, P. Altherr, V. Cremet, B. Emir, S. Micheloud, N. Mihaylov, M. Schinz, E. Stenman, and M. Zenger. The Scala language specification, 2004.
- [31] T. Pixley. Document Object Model (DOM) Level 2 Events Specification. W3C Recommendation 13 November 2000.
- [32] E. Prud'hommeaux and A. Seaborne. SPARQL Query Language for RDF. W3C Recommendation 15 January 2008.
- [33] S. Schenk, P. Gearon, et al. SPARQL 1.1 Update. W3C Recommendation 21 March 2013.
- [34] S. Schenk and S. Staab. Networked graphs: a declarative mechanism for SPARQL rules, SPARQL views and RDF data integration on the web. In *Proc. WWW '08*, pages 585–594. ACM, 2008.
- [35] M. Shaw, L. T. Detwiler, N. Noy, J. Brinkley, and D. Suci. vSPARQL: A view definition language for the semantic web. *Journal of Biomedical Informatics*, 44(1):102 – 117, 2011. Ontologies for Clinical and Translational Research.
- [36] S. A. Smits and C. C. Ouverney. jsPhyloSVG: A Javascript Library for Visualizing Interactive and Vector-Based Phylogenetic Trees on the Web. *PloS one*, 5(8):e12267, 2010.
- [37] J. Van den Bussche, S. Vansummeren, and G. Vossen. Towards practical meta-querying. *Information Systems*, 30(4):317–332, 2005.

# Hybrid Classification Engine for Cardiac Arrhythmia Cloud Service in Elderly Healthcare Management

Huan Chen<sup>1</sup>, Guo-Tan Liao<sup>3\*</sup>, Min-Sheng Chien<sup>4</sup>,  
Ting-Chun Kuo<sup>5</sup>

Department of Computer Science and Engineering  
National Chung Hsing University  
Taichung, Taiwan

Bo-Chao Cheng<sup>2</sup>

Department of Communications Engineering  
National Chung Cheng University  
Chiayi, Taiwan

**Abstract**—The self-regulation ability of the elderly is largely degenerated with the age increases, and the elderly often expose to great potential hazards of heart disorders. In practice, the Electrocardiography (ECG) is one of the well-known non-invasive procedures used as records of heart rhythms and diagnosis of unusual heart diseases. In this paper, we propose a healthcare management system, named CardiaGuard, which is specialized in monitoring and analysis the heart disorder events for the elderly. The CardiaGuard cloud service is an expert system designed based on the hybrid classifier implemented using Support Vector Machine (SVM) and Random Tree (RT) classification algorithm. We conduct a comprehensive performance evaluation which shows the proposed hybrid classification engine are able to detect six types of cardiac disorders with higher accuracy rate than the SVM-based classifier alone. CardiaGuard poses a great solution to enhance the quality of good clinical practice on the healthcare management for the elderly in cardiology.

**Keywords**—*Electrocardiograph (ECG); Healthcare; RR interval (RRI); Android-based; Arrhythmia classification; Support Vector Machine (SVM)*

## I. INTRODUCTION

Heart disease is the major cause of death for countries. In today's complex life, it is easy to find physical and emotional disorders on every citizen. Sleep disorders is a common problem in the elderly, which causes even more severe physical conditions such as heart disorders. For more comfortable and complete healthcare service, a lot of healthcare system models have been proposed to help diagnose, monitor and provide services to heart disease patients in recent years [1][2]. For clinical observation, there exists an inextricable connection between the disorders and the autonomic nervous system.

Although the advance in medical technology extends the lifespan of human beings, it is still a challenging task to meet the requirements of quality clinical practices under limited available medical resources and healthcare resource. Moreover, the patient's family wants the necessary and appropriate action taken at any time when a patient is in need of care or when an unusual emergency situation is occurring. With the rapid

development of information and communication technology, more ehealth services are introduced in order to improve access, efficiency and quality of health care services. There is an urgent need for the home care system to send real-time monitoring of physical information to a central health management system, which would serve as a platform where the doctor could communicate with his/her patients who need helps.

To mitigate the potential impact of the limited medical resource, home telecare and remote healthcare services emerge to offer monitoring, alerting services or providing medical information or healthcare tutorial remotely in recent years. And to provide quality telecare services, a good healthcare management system (HMS) is critical since all services are carried on the platform. An effective HMS would benefit not only patients, but also benefit to the patient's family with enough information and suggestion, as a result taking correct actions when taking care of patients.

The degradation of the ability to adapt and respond to changing health situations and environments causes a lot of potential hazards on the health of the elderly. Among all potential threats, heart disorders are more critical that may cause sudden death. Electrocardiography (ECG), one of well-known non-invasive procedures, is commonly used to detect abnormal heart rhythms and a powerful tool in diagnosing heart disorders, which is evolved from a long history references to and correlation with known cardiac disorders. Further, each individual has his own unique ECG signal which may also be influenced by his mood, illness and also his environment. Conventional analysis tools (e.g., rule-based analysis) do not have a personalized learning heart disorder capability. Therefore, the clinical experts still need to get involved and carefully identify the symptom. There are many studies on ECG signal processing such as baseline correction, noise removal, R-wave detection, QRS algorithm and disease diagnosis [3][4][5][6]. With medical knowledge and techniques, hence, Electrocardiography (ECG) can be used as an important cue for diagnosis of heart disease and trace of treatment.

On the other hands, real-time heart rate variability (HRV) analysis, measured based on QRS detection and beat-to-beat

\*Corresponding author: G.-T. Liao. This research was supported in part by the National Science Council (NSC) in Taiwan under the grant numbers NSC 102-2220-E-005-008 and NSC 102-2221-E-194 -006.



intervals, is essential to emotional recognition. One of the most important parameters is RRI that is the interval between two successive R points, the peak of the QRS complex in electrocardiography (ECG) wave. Recent studies also show that RRI is very useful to identify Premature Ventricular Contraction (PVC), a common type of heart arrhythmias [7]. What we are interested in is how to build a reliable cloud ehealth service to recognize the disease detection and arrhythmia classification through HRV and RRI analysis. In this paper, we propose a healthcare management system, named CardiaGuard, built on the top of cloud computing platform. Taking advantages of the cloud computing platform (such as flexibility, cost reduction, reliability, security gains), CardiaGuard is able to provide reliable and safe real-time surveillance of the heart disorder events for the elderly.

## II. BACKGROUND

Heart Rate Variability (HRV) [11][12] is the variation of time interval between heartbeats that represents a physiological phenomenon. Generally speaking, HRV analysis can be divided into two categories: time domain and frequency domain measurements. For time domain analysis of HRV, it uses the simplest parameters to be calculated, such as the RR intervals (RRI), mean NN intervals (MNN,  $RR_m$ ), the standard deviations of all NN intervals (SDNN), the square root of the mean of the sum of the squares of differences between adjacent NN intervals (RMSSD) and the proportion derived by dividing the number of interval differences of NN intervals greater than 50 ms by the total number of NN intervals (PNN50). For instance,  $RR_m = \frac{1}{N} \sum RR(i)$ .

FFT is simple in calculation to get Power Spectral Density (PSD).

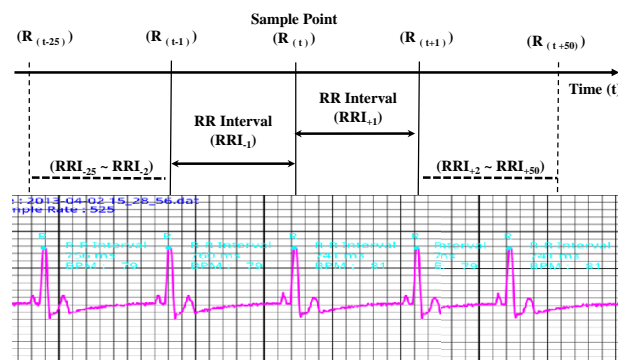


Fig. 1. Example of illustrating RR intervals

There are many methods for classification proposed in literature, including K-Nearest Neighbors (KNN) [13], Linear Discriminant Analysis (LDA) [14], multinomial logistic regression (also called Softmax Regression) [15], Genetic Programming (GP) [16], fuzzy weighted [17], neural network (NN) [18], Bayesian classifiers [19], support vector machine (SVM) [20], Hidden Markov Models (HMM) [21], rule-based algorithms [22] and Random Tree [23]. Several popular algorithms are used for disease identification and classification of cardiac arrhythmias. For reducing system complexity and improving accuracy, one kind of methods has a step for feature reduction. Genetic programming can be used to select effective features to distinguish between different types of arrhythmias. Another kind of methods is data pre-processing like normalization and weighted values. Involving both feature selection and data pre-processing, a novel layered hidden Markov model (LHMM) was proposed to detect cardiac arrhythmia that is two layers of HMMs, the Layer 1 HMM extracts the features of ECG waveform and human activities, and the Layer 2 HMM classifies the type of cardiac arrhythmias. On the other hand, the rule-based algorithms and the deterministic automaton rely on medical knowledge.

knowledge, we want to know how accurate the time intervals RRI can provide assistance for classifying of multiple cardiac arrhythmias. Based on statistical principle, we focus on the comparison of SVM, Softmax Regression and NN algorithms. Cascade neural network is a type of artificial neural networks (ANN) has two properties, Cascade Correlation and supervised learning algorithm. In many cases, SVM provides better classification results than the NN methods. About Softmax Regression, it is usually used to solve nonlinear classification problem for optimization by gradient descent or by Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (LBFGS). Moreover, SVM training models mainly divide into three kinds of kernel functions, linear, Radial Based Function (RBF) and quadratic functions. Here, we especially analyze RBF SVM (by choosing optimal parameters,  $\sigma$  and  $c$ ). In this paper, we take MIT-BIH arrhythmia database as training data to show the performance of classifying Premature Ventricular Complex (PVC) and other arrhythmias.

### III. CARDIAGUARD OF HEALTHCARE CLOUD SERVICE

The proposed personal healthcare system [24] includes three subsystems as shown in Fig. 2: (1) the ECG sensor node [25], (2) the mobile device responsible for ECG signal process and data communication, and (3) the surveillance system and healthcare service. The detailed model of our methodology of arrhythmia classification is illustrated by Fig. 3.

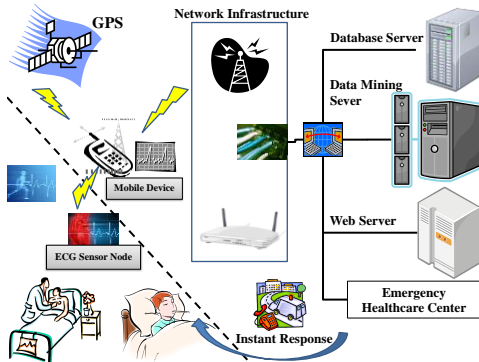


Fig. 2. Health monitoring system architecture

- Preprocessing: Although it may cause the loss of real useful ECG signals, it is a critical step for improving classification accuracy before measuring the amplitudes to perform baseline drift correction for ECG data. One way is to use the integer coefficient digital filter and the ten-point moving average filter [26]. Another way is to remove the baseline wander by median filter [4], and remove the high frequency noise by symN wavelet, such as sym10 [27].
- R-wave detection: There are many QRS detection methods including So-and-Chan QRS detection algorithm [6] and Hilbert transform with automatic threshold [28].
- Feature extraction: We only adopt the continuous RR intervals ( $RRI_{-25} \sim RRI_{+50}$ ) as attributes of each sample.

Generally, there are parameters of time domain and frequency domain analysis.

- Training model and arrhythmia classification: In our experiments, we conduct the SVM-based algorithms.

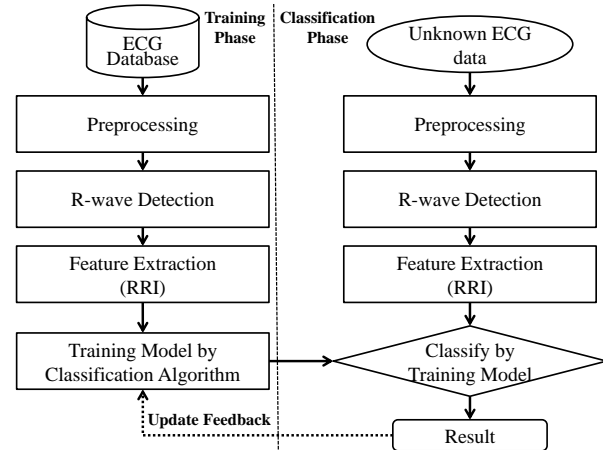


Fig. 3. Methodology of arrhythmia classification

A key point is to set the alarm service that can trigger the healthcare worker to query the user's situation and make the instant response such as immediately calling ambulance if necessary. In common, the range of heart rate is between 60-100 times per minute. For example, when the heart beats per minute is larger than 120 or lower than 40, the alarm service with the threshold will be triggered. The healthcare system application in smart home is developed as real-time ECG monitoring platform by HRV analysis and cardiac arrhythmia classification. Furthermore, when the service cost of the related hard devices and system applications is low enough to be popular for everyone in the future, this kind of healthcare system will benefit the whole world. Especially note we can take Pulse to Pulse Interval (PPI) of the patient from the so-called heart rate watch by Photoplethysmography (PPG) as similar RRI even when the real RRI information (HRV related) is not monitored [29][30][31]. Due to flexibility, it is more likely to convince the elderly to wear smart devices such as a watch with biosensor. In other words, in our proposed healthcare management system, it needs to be able to extract the typical RRI or PPI information.

#### A. Hybrid classification engine

In our CardiaGuard cloud service, the major purpose is to classify cardiac arrhythmias no matter whether the arrhythmias had been detected in personal ECG records (PER). Hence, we propose the three-layer classification flow model as shown in Fig. 4. Layer 1 is by general ECG database including the well-known ECG database, PER, and etc. Layer 2 and Layer 3 are for classification by PER. Layer 1 is designed to mainly detect other arrhythmias, not existed in PER. In other words, through other person's ECG records, we may earlier find the presentation of the new arrhythmia for self-health awareness.

On the other hands, the design principle of the Layer 2 is to acquire higher classification of arrhythmias existed in PER as the presentation of everybody's ECG could be different. Using PER, it is easier to catch the property of features for personal healthcare management. Moreover, Layer 3 Random Tree (RT) filter is designed to improve classification accuracy through the observation of doctors. If we find some type is falsely identified as another type many times, we can do further process like using Random Tree method (RT) for classification of suspicious arrhythmias between limited type scopes. For example in our experiments about RT filter, we use RT classification method to determine the type 'L', classified in Layer 2, is more like 'L' or 'E'. That is because we know the performance of classification for the two types ('L' and 'E') is very good by RT method. Finally, combining the classification results of Layer 1-3, the healthcare service can offer the precise personalized arrhythmia classification and detection of new arrhythmias for specific person.

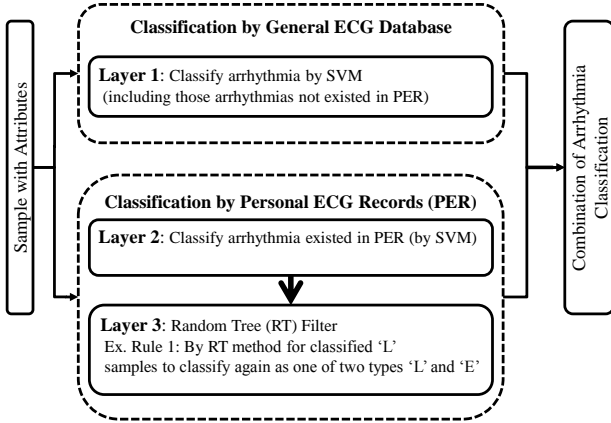


Fig. 4. CardiaGuard classification flow model

#### B. Use-case of cloud service

In the elderly healthcare management system, there is a SVM-based arrhythmia classification cloud service. Among all potential threats, heart disorders are more critical that may cause sudden death. Because ECG signals of different patients of the same disease are not all the same that each patient may have specific patterns in self ECG signals, there is a special design of the cloud service for the elderly healthcare management system that the training model is made based on the arrhythmia analysis of the patient's self ECG signals with detailed examinations carried out by doctors. The cloud service for the elderly healthcare needs the cooperation of doctors and hospitals, and the use-case flowchart is illustrated in Fig. 5.

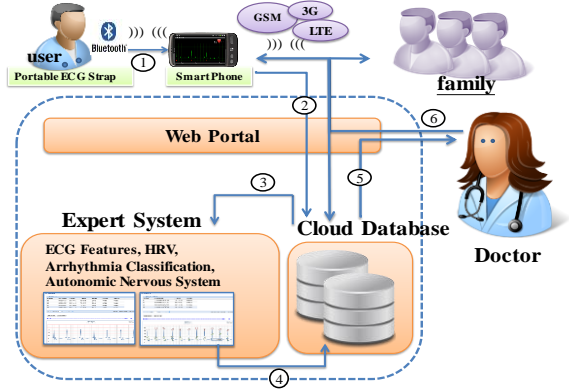


Fig. 5. Use-case flowchart of cloud service for healthcare

- Step 1. ECG data is gathered and then transmitted to smart device by Bluetooth communication. ECG records can be shown and uploaded through the smart device with network connection capability.
- Step 2. ECG records are transmitted to the cloud database by WiFi, 3G or Long Term Evolution (LTE).
- Step 3. The unanalyzed ECG records will be read by the expert system, and perform QRS detection process. Next, the expert system automatically classified out arrhythmias by SVM classification method.
- Step 4. The data information including QRS and classification results are sent to the cloud database for storage.
- Step 5. The doctors can observe the original data and the classified record by web service. At the meantime, the doctors can manually fix the arrhythmia results for on-line real-time correction. The ECG records with arrhythmia information processed by doctors will be given as ECG training sample for personalized arrhythmia classification service.
- Step 6. The patients and their relatives can view the results of arrhythmia analysis via web service.

#### IV. EXPERIMENT

For testing the efficiency of arrhythmia classification under different RRI attributes, we take the ECG records (R207, R124 and others) from MIT-BIH database as training data, and adopt RBF SVM as the default classification method. Here, in our experiments, RBF SVM is with the optimal parameters (sigma and C) [32]. In practical, the system can be set with the optimal RBF parameters for personalized service. TABLE I summaries the arrhythmia notations and counts in R207 and R124. And shown in TABLE II, we denote  $RRI_1$  and  $RRI_2$  by 2-RRI, denote from  $RRI_3$  to  $RRI_{+6}$  by 9-RRI, and denote from  $RRI_{-24}$  to  $RRI_{+48}$  by 72-RRI.

TABLE I ARRHYTHMIA TYPE ON R207 AND R124

Notation	Meaning of Arrhythmia Type	Count	
		R207	R124
L	Left bundle branch block beat	1457	0
R	Right bundle branch block beat	85	1531
A	Atrial premature beat	107	2
V	Premature ventricular contraction	105	52
!	Ventricular flutter wave	472	0
E	Ventricular escape beat	105	0
J	Nodal (junctional) premature beat	0	29
j	Nodal (junctional) escape beat	0	5

TABLE II EXAMPLE OF MULTIPLE RRI ATTRIBUTES

Number of RRIs	Arrange(t) of $RRI_0$
2	-2, -1
9	-3 ~ +6
18	-6 ~ +12
36	-12 ~ +24
72	-24 ~ +48

#### A. Undetected arrhythmia by general records (Layer 1)

In this section, we discuss the classification performance of joint ECG records from different patients. Generally speaking, taking others' ECG records as training samples, the performance of self ECG arrhythmia classification is not very good. Even if taking joint ECG records as training samples, including self ECG records, the classification performance still does not achieve high accuracy. However, through the results of TABLE III, it shows multiple RRIs can help solve the problem. Here, 8 different ECG records are taken in the general ECG database, including R111, R118, R124, R200, R207, R213, R228 and R232. Among the 5 kinds of RRI attribute schemes (with 9 arrhythmia types), the mode with 72-RRI can have the highest accuracy in the K-fold cross-validation experiments (K=10). That is, arrhythmia classification by multiple RRIs can perform well in the 1-layer classification.

TABLE III AVERAGE ACCURACY BY CROSS-VALIDATION WITH DIFFERENT RRI ATTRIBUTES IN 1-LAYER CLASSIFICATION (8 RECORDS)

Method	2-RRI	9-RRI	18-RRI	36-RRI	72-RRI
SVM	74.96%	91.85%	92.52%	<b>93.09%</b>	<b>93.91%</b>
NN	70.98%	80.14%	81.83%	<b>83.62%</b>	<b>89.53%</b>

#### B. Analysis on personal record (Layer 2 classification)

For personalized healthcare management, personal ECG records can help our system more precisely classify those arrhythmias which were diagnosed and detected before. The following results can prove that the SVM-based algorithm is effective and efficient because it has good accuracy performance and takes execution time less than NN method. TABLE IV shows classification accuracy of multiple RRIs, used as attributes of each sample. The results present the classification accuracy of from 9-RRI to 72-RRI is better than 2-RRI. The detailed classification results of R207 by SVM based on 2-RRI, 9-RRI, 18-RRI, 36-RRI and 72-RRI are listed in from TABLE V to TABLE IX.

TABLE IV CLASSIFICATION ACCURACY OF R207 &amp; R124 BY SELF-TRAINING BASED ON RBF SVM WITH DIFFERENT RRI ATTRIBUTES

Method	Record	2-RRI	9-RRI	18-RRI	36-RRI	72-RRI
		Record	Record	Record	Record	Record
SVM	R207	96.39%	99.26%	<b>100%</b>	97.77%	<b>99.33%</b>
	R124	95.90%	96.01%	96.24%	<b>96.45%</b>	<b>97.86%</b>
NN	R207	95.71%	96.98%	<b>99.35%</b>	98.95%	<b>99.82%</b>
	R124	95.97%	97.13%	97.30%	<b>97.84%</b>	<b>97.53%</b>

TABLE V CORRECTLY CLASSIFICATION OF R207 FOR 2-RRI

Classified Arrhythmia	L	R	A	V	!	E	Accuracy (%)
<b>L (LBBBB)</b>	1450	7	0	0	0	0	99.51
<b>R (RBBBB)</b>	1	84	0	0	0	0	98.82
<b>A (APB)</b>	0	0	107	0	0	0	100
<b>V (PVC)</b>	0	0	0	102	3	0	97.14
<b>! (VFW)</b>	0	0	0	5	467	0	98.94
<b>E (VEB)</b>	68	0	0	0	0	37	35.23

TABLE VI CORRECTLY CLASSIFICATION OF R207 FOR 9-RRI

Classified Arrhythmia	L	R	A	V	!	E	Accuracy (%)
<b>L (LBBBB)</b>	1456	0	0	0	0	1	99.78
<b>R (RBBBB)</b>	0	85	0	0	0	0	100
<b>A (APB)</b>	0	0	101	0	0	0	100
<b>V (PVC)</b>	0	0	0	104	0	0	100
<b>! (VFW)</b>	0	0	0	0	472	0	100
<b>E (VEB)</b>	16	0	0	0	0	89	84.76

TABLE VII CORRECTLY CLASSIFICATION OF R207 FOR 18-RRI

Classified Arrhythmia	L	R	A	V	!	E	Accuracy (%)
<b>L (LBBBB)</b>	1457	0	0	0	0	0	100
<b>R (RBBBB)</b>	0	83	0	0	0	0	100
<b>A (APB)</b>	0	0	95	0	0	0	100
<b>V (PVC)</b>	0	0	0	103	0	0	100
<b>! (VFW)</b>	0	0	0	0	472	0	100
<b>E (VEB)</b>	0	0	0	0	0	105	100

TABLE VIII CORRECTLY CLASSIFICATION OF R207 FOR 36-RRI

Classified Arrhythmia	L	R	A	V	!	E	Accuracy (%)
<b>L (LBBBB)</b>	1455	0	1	0	0	1	99.86
<b>R (RBBBB)</b>	1	79	0	0	0	0	98.75
<b>A (APB)</b>	1	0	82	0	0	0	98.79
<b>V (PVC)</b>	0	0	0	100	0	0	100
<b>! (VFW)</b>	0	0	0	0	472	0	100
<b>E (VEB)</b>	44	2	0	1	0	58	55.23

TABLE IX CORRECTLY CLASSIFICATION OF R207 FOR 72-RRI

Classified Arrhythmia	L	R	A	V	!	E	Accuracy (%)
<b>L (LBBBB)</b>	1456	0	1	0	0	0	99.93
<b>R (RBBBB)</b>	0	74	0	0	0	0	100
<b>A (APB)</b>	0	0	59	0	0	0	100
<b>V (PVC)</b>	0	0	0	94	0	0	100
<b>! (VFW)</b>	0	0	0	0	472	0	100
<b>E (VEB)</b>	13	1	0	0	0	91	86.66

The results of R207 clearly present the classification accuracy based on 18-RRI and 72-RRI has good performance (18-RRI > 72-RRI > 9-RRI > 36-RRI > 2-RRI). Furthermore, we summarize the experiments of R124 as shown in TABLE X. The accuracy metric is denoted as correctly classification. And the results of R124 obviously present the effectiveness of 72-RRI that is 72-RRI > 36-RRI > 18-RRI > 9-RRI > 2-RRI. Then, through the analysis of classification results, we consider 72-RRI is the most reliable manner among these five cases of RRI-based sample attributes.

Moreover the overall performance evaluation is carried out by the way of cross-validation (K=10). The results are listed in TABLE XI. But when the system administrators think 72-degree samples are too heavy, the 18-RRI sample attributes is the good choice as its overall cross-validation performance is also good. Roughly speaking, multiple RRIs do improve the classification accuracy than 2-RRI mode.

TABLE X CORRECTLY CLASSIFICATION ACCURACY OF R124

RRIs Arrhythmia	2- RRI	9- RRI	18- RRI	36- RRI	72- RRI
<b>R</b>	99.93%	100%	100%	100%	100%
<b>A</b>	0%	0%	0%	0%	100%
<b>V</b>	31.91%	36.17%	40.42%	48.93%	55.31%
<b>J</b>	3.44%	0%	0%	0%	58.62%
<b>j</b>	40%	40%	80%	80%	100%

TABLE XI AVERAGE ACCURACY BY CROSS-VALIDATION WITH DIFFERENT RRI ATTRIBUTES

Method	RRIs Record	2- RRI	9- RRI	18- RRI	36- RRI	72- RRI
SVM	R207	94.89%	96.08%	<b>96.97%</b>	95.55%	<b>96.77%</b>
	R124	95.66%	95.89%	<b>95.92%</b>	95.75%	<b>96.11%</b>
NN	R207	94.85%	95.22%	<b>96.37%</b>	95.77%	<b>96.59%</b>
	R124	<b>95.84%</b>	95.20%	94.99%	95.44%	<b>95.78%</b>

### C. Analysis on RT filter (Layer 3 classification)

About the Layer 3 classification, we take R207 as an example. Firstly, we observe that some type 'E' is often falsely identified as another type 'L'. Secondly, we know the performance of classification between the two types 'L' and 'E' can be very effective by RT classification method. So, this

rule can be added in Layer 3 RT filter that corrects the false positive of the type 'E' as shown in TABLE XII. In fact, the so-called correctly accuracy reflects another sensitivity measurement for whole classification performance.

TABLE XII PERFORMANCE OF FROM LAYER 2 TO LAYER 3 CLASSIFICATION WITH DIFFERENT RRI ATTRIBUTES FOR R207

True 'E' falsely classified as 'L'	Error count (Correctly Accuracy)	2- RRI	9- RRI	18- RRI	36- RRI	72- RRI
	Layer 2	68 (35.23%)	16 (84.76%)	0 (100%)	44 (55.23%)	13 (86.66%)
	Layer 3	15 (85.71%)	0 (100%)	0 (100%)	0 (100%)	0 (100%)

### D. Performance Analysis in CardiaGuard

Based on the description of Layer 1 above, we set a database of 8 records (including 9 arrhythmia types), and let R207 be the testing data (known including 6 arrhythmia types). By the experiment, we show the performance efficiency and difference between single layer and multi-layer co-decision. As shown in TABLE XIII, we summarize observations as follows. And in the C<sub>4</sub> case, the integration of Layer 1 and Layer, where Layer 3 is mainly personal-relative and Layer 1 is an assistant of other new arrhythmias, has the mutual decisive property.

- The performance of C<sub>3</sub> (by Layer 3) is promoted from C<sub>2</sub> (by Layer 2).
- Using 72-RRI as attributes, the C<sub>1</sub> classification could achieve the accuracy of up to 100% in this experiment.
- Although the performance of the C<sub>4</sub> case is obviously affected by C<sub>3</sub>'s high accuracy for 18-RRI, 36-RRI and 72-RRI, the classification of by 2-RRI or 9-RRI has the difference between the C<sub>3</sub> and C<sub>4</sub> cases. The difference means the detection of the 3 arrhythmias ('J', 'j', 'N'), which cannot be detected by C<sub>3</sub>, but can be detected by C<sub>1</sub> or C<sub>4</sub>. Overall, C<sub>4</sub> could main high classification accuracy and possess the ability to find other arrhythmias which are new for personal health record.

TABLE XIII ANALYSIS OF CARDIAGUARD EXPERIMENTAL ACCURACY

Case	RRIs	2- RRI	9- RRI	18- RRI	36- RRI	72- RRI
C <sub>1</sub> : Layer 1 (L1)		79.40%	96.25%	96.93%	99.30%	100%
C <sub>2</sub> : Layer 2 (L2)		96.39%	99.26%	100%	97.77%	99.33%
C <sub>3</sub> : Layer 3 (L3)		99.35%	100%	100%	100%	100%
C <sub>4</sub> : Layer (1&3)		96.18%	99.44%	100%	100%	100%
Difference of C <sub>3</sub> & C <sub>4</sub> Cases	Count('J', 'j')	1	0	0	0	0
	Count('N')	73	13	0	0	0



## V. CONCLUSION

This paper introduces a healthcare management system which can help the elderly to be aware of cardiac arrhythmias and urgent needs of healthcare assistance. The system prototype and hybrid classification methodology are illustrated above. The experimental results show the proposed hybrid classification engine (Layer 2: SVM-based classifier plus Layer 3: RT-based filter) improve the accuracy rate for E type arrhythmia from 35.23% to 85.71% under 2-RRI. Moreover, our proposed hybrid classification engine is able to achieve 100% accuracy rate under the conditions of multiple RRIs (from 9-RRI to 72-RRI). In future work, we will study more types of arrhythmias which cause sudden cardiac death (SCD) and conduct a clinical trial with the medical organization after the approval from the Institutional Review Board (IRB).

## REFERENCES

- [1] P. DeToledo, S. Jimenez, F. del Pozo and J. Roca, "Telemedicine Experience for Chronic Care in COPD," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, issue 3, July 2006, 567-573.
- [2] T. Amouh, M. Gemo, B. Macq, J. Vanderdonckt, A. W. El Gariani, M. S. Reynaert, L. Stamatakis and F. Thys, "Versatile clinical information system design for emergency departments," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 2, June 2005, 174-183.
- [3] H. H. So and K. L. Chan, "Development of QRS detection method for real-time ambulatory cardiac monitor," *The 19-th International Conference, IEEE/EMBS*, vol. 2, pp. 289-292, 1997.
- [4] R. C. Gonzalez and R. E. Woods, "Digital Image Processing," 2nd Edition, Prentice Hall, 2002. ISBN: 0-201-18075-8
- [5] P. de Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. O'Malley, "Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea," *IEEE Transactions on Bio-medical Engineering*, vol. 50, no. 6, pp. 686-696, 2003.
- [6] S. Li, Y. Ji, and G. Liu, "Optimal Wavelet Basis Selection of Wavelet Shrinkage for ECG De-Noiseing," *International Conference on Management and Service Science*, pp. 1-4, September 2009.
- [7] Y. Wang, K. N. Plataniotis and D. Hatzinakos, "Integrating Analytic and Appearance Attributes for Human Identification from ECG Signals," *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, Baltimore, MD, 2006.
- [8] W.-T. Sung, J.-H. Chen and K.-W. Chang, "Mobile Physiological Measurement Platform with Cloud and Analysis Functions Implemented via IPSO," *IEEE Sensors Journal*, vol. 14, issue 1, January 2014, 111-123.
- [9] E. J. Vergara, J. Sanjuan and S. Nadjm-Tehrani, "Kernel level energy-efficient 3G background traffic shaper for android smartphones," *The 9-th International Wireless Communications and Mobile Computing Conference (IWCMC 2013)*, pp. 443-449, July 2013.
- [10] MIT-BIH, MIT-BIH Arrhythmia Database. <http://www.physionet.org/physiobank/database/mitdb/>
- [11] G. D. Clifford, "Signal processing methods for heart rate variability," Ph.D. dissertation, University of Oxford, 2002.
- [12] LabVIEW & Biosignal Analysis, ISBN-13: 9789866301001, 2009.
- [13] I. Christov, I. Jekova, and G. Bortolan, "Premature ventricular contraction classification by the Kth nearest-neighbours rule," *Physiological measurement*, vol. 26, no. 1, January 2005, 123-130.
- [14] A. Sengur, "An expert system based on linear discriminant analysis and adaptive neuro-fuzzy inference system to diagnosis heart valve diseases," *Expert Systems with Applications*, vol. 35, 2008, 214-222.
- [15] H. R. Sanabila, M. I. Fanany, W. Jatmiko and A. M. Arimurthy, "Bootstrapped Multinomial Logistic Regression on Apnea Detection Using ECG Data," *International of Advanced Computer Science and Information Systems (ICACSIS 2010)*, pp 181-186, ISSN: 2086-1796, Bali, 2010.
- [16] M. Tavassoli, M. M. Ebadzadeh and H. Malek, "Classification of cardiac arrhythmia with respect to ECG and HRV signal by genetic programming," *Canadian Journal on Artificial Intelligence, Machine Learning and Pattern Recognition*, vol. 3, no. 1, January 2012, 1-8.
- [17] K. Polat, S. Sahan and S. Güneş, "A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia," *Expert Systems with Applications*, vol. 31, issue 2, August 2006, 264-269.
- [18] R. Ceylan, Y. Özbay and B. Karlik, "A novel approach for classification of ECG arrhythmias: Type-2 fuzzy clustering neural network," *Expert Systems with Applications*, vol. 36, issue 3, part 2, April 2009, 6721-6726.
- [19] R. J. Muirhead and R. D. Puff, "A Bayesian classification of heart rate variability data," *Physica A: Statistical Mechanics and its Applications*, vol. 336, issue 3-4, May 2004, 503-513.
- [20] B. M. Asl, S. K. Setarehdan and M. Mohebbi, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal," *Artificial Intelligence in Medicine*, vol. 44, issue 1, September 2008, 51-64.
- [21] S. Hu, Z. Shao and J. Tan, "A Real-Time Cardiac Arrhythmia Classification System with Wearable Electrocardiogram," *2011 International Conference on Body Sensor Networks (BSN 2011)*, pp. 119-124, Dallas, Texas, May 23-25, 2011.
- [22] M. G. Tsipouras, D. I. Fotiadis and D. Sideris, "An arrhythmia classification system based on the RR-interval signal," *Artificial Intelligence in Medicine*, vol. 33, issue 3, March 2005, 237-250.
- [23] S. Shandilya, K. R. Ward and K. van Najarian, "A time-series approach for shock outcome prediction using machine learning," *IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW 2010)*, pp. 440-446, Hong Kong, December 2010.
- [24] L.-H. Wang, T.-Y. Chen, S.-Y. Lee and H. Chen, "Implementation of a Personal Health Monitoring System in Cardiology Application," *2012 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS 2012)*, pp. 236-239, December 2012.
- [25] D.-S. Lee, S. Bhardwaj, E. Alasaarela and W.-Y. Chung, "An ECG Analysis on Sensor Node for Reducing Traffic Overload in u-Healthcare with Wireless Sensor Network," *2007 IEEE Sensors Conference*, pp. 256-259, Atlanta, GA, October 28-31, 2007.
- [26] D. C. Ready, *Biomedical signal processing principles and techniques*, Tata McGraw-Hill Education, ISBN: 0070583889, 2005.
- [27] G. Garg, S. Gupta, V. Singh, J. R. P. Gupta and A. P. Mittal, "Identification of optimal wavelet-based algorithm for removal of power line interferences in ECG signals," *2010 India International Conference on Power Electronics (IICPE 2010)*, New Delhi, January 28-30, 2011.
- [28] N. M. Arzeno, Z.-D. Deng and C.-S. Poon, "Analysis of First-Derivative Based QRS Detection Algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 2, February 2008, 478-484.
- [29] F.-C. Chang, C.-K. Chang, C.-C. Chiu, S.-F. Hsu and Y.-D. Lin, "Variations of HRV analysis in different approaches," *Computers in Cardiology*, vol. 34, pp. 17-20, Durham, North Carolina, 2007.
- [30] K. Davoudi, M. Shayegannia and B. Kaminska, "Vital signs monitoring using a new flexible polymer integrated PPG sensor," *Computing in Cardiology Conference (CinC 2013)*, pp. 265-268, Zaragoza, September 22-25, 2013.
- [31] A. Axel Schäfer and J. Vagedes, "How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram," *International Journal of Cardiology*, vol. 166, issue 1, June 2013, 15-29.
- [32] C.-C. Chang and C.-J. Lin, "LIBSVM : a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, issue 3, April 2011. (Last updated: March 2013) <http://www.csie.ntu.edu.tw/~cjlin/libsvm>



# Gait Recognition Based on Joint Distribution of Motion Angles

Wei Lu\*, Wei Zong, Weiwei Xing

(School of Software Engineering, Beijing Jiaotong University, Haidian District, Beijing 100044, China)

**Abstract**—Gait as a biometric trait has the ability to be recognized in remote monitoring. In this paper, a method based on joint distribution of motion angles is proposed for gait recognition. The new feature of the motion angles of lower limbs are defined and extracted from either 2D video database or 3D motion capture database, and the corresponding angles of right leg and left leg are joined together to work out the joint distribution spectrums. Based on the joint distribution of these angles, we build the feature histogram individually. In the stage of distance measurement, three types of distance vector are defined and utilized to measure the similarity between the histograms, and then a classifier is built to implement the classification. Experiments has been carried out both on CASIA Gait Database and CMU motion capture database, which show that our method can achieve a good recognition performance.

**Index Terms** —biometrics, gait recognition, joint distribution, feature histogram

## I. INTRODUCTION

BIOMETRICS refer to the identification of humans by their characteristics or traits[1]. The characteristics include but not limited to face, fingerprint, iris, gait and DNA. However the current recognition methods, such as face, fingerprint or iris based, require a cooperative subject or physical contact. So it is nearly impossible to identify people at a distance by using these methods. However, gait as the way people walk does not have these constraints. In the past decades, many studies have proven that gait has the potential to become a powerful biometric for surveillance and access control, since it has advantages such as noninvasive, hard to conceal and capable of being acquired at a distance [2]. In fact, besides being well-suited to identify people at a distance, gait also have the potential to be applied in the medical field. For example, recognizing changes in walking patterns early can help to identify conditions such as Parkinson's disease and multiple sclerosis in their earliest stages [3]. Although gait has some limitations, e.g., it may not be as unique as fingerprint or iris, and may be affected by one's clothing, shoes or physical condition, the inherent gait characteristic of an individual still makes it irreplaceable and useful in visual surveillance.

Nowadays, video is not the only way to collect the gait any more. According to the ways of data collection, gait recognition methods can be divided into three categories: Video Sensor (VS)-based, Floor Sensor (FS)-based and

Wearable Sensor (WS)-based [4]. VS-based method collects gait data by video cameras. Without physical contact, VS-based method is the most noninvasive way and can get the most natural way of one's walking. Moreover, video cameras are widely used in our daily life, so it is quite easy to get the gait data in a variety of occasions. However, the image processing is required, and the images captured from cameras should be preprocessed in order to get the gait information that can be used directly. The most widely used field of VS-based method is remote monitoring. FS-based method is also called footprint method, which puts the sensors on the floor and record the information of one's footprint such as length and location to be studied. WS-based method needs the subject wear sensors and collects the motion data recorded by them. Different from VS-based method, WS-based method can get the motion data directly and the data is more suitable for gait analysis. WS-based method is popularly used in gait analysis mainly for medical purpose. However, it is not a good choice for remote monitoring as the non-invasive is the key feature. The gait data can be represented in 2D or 3D. 2D data is presented by the sequence of images, which is widely used in early days. While in recent years, gait recognition based on 3D data became a new trend. The 3D data is mainly acquired and calculated by motion capture technology.

In this paper, a method based on joint distribution of motion angles for gait recognition is proposed, which can work on 2D video database or 3D motion capture database. The motion angles of lower limbs from the original data are extracted to propose the new feature, thus the joint distribution spectrums can be work out. Then the feature histogram is built and the distance between the histograms is calculated. Finally, the classification is implemented on the distance vector to recognize the gaits.

As for the experiments data in this paper, we use the CASIA Gait Database [5] from the Institute of Automation, Chinese Academy of Sciences as the 2D data source, and the CMU motion capture database in ASF/AMC format [6] as 3D data. In the stage of motion angle extraction, we only use the motion angles of lower limbs which can work well even when the subject is with half-occlusion in the images. After the motion angles are extracted, we take the result as a time-free model instead of considering them as a time sequence model. The spatial distribution of them, however, is what we concerned. In other words, we only care about the postures of the subject when he is walking. Experimental results are compared with other similar work which demonstrates our method can reach a higher accuracy.

The reminder of the paper is organized as follows: Section

---

\* Corresponding author: luwei@bjtu.edu.cn

2 is the related work. Section 3 shows the feature definition and extraction. The classification is described in Section 4. Section 5 presents the experiments and the results. Section 6 concludes the paper.

## II. RELATED WORK

Gait recognition method has been well studied and many methods have been proposed which can be classified into model-free and model-based approaches [7]. In the model-free approaches, moment of shape is one of the most common used features. In addition silhouette and statistical features are widely used in a lot of work. In this approach, the correspondence between successive frames is established based upon the prediction or estimation of features related to position, velocity, shape, texture and color. Alternatively, they assume some implicit notion of what is being observed [8]. In the model-based approaches, the prior information or a known model is needed for fitting human gait. Though the model-based method is more complex, it has some advantages such as immunity to noise [9]. The model-based approaches assume a priori model to match the data extracted from video, and features correspondence is automatically achieved. These two kinds of methods both follow the general framework of features extraction, features correspondence and high-level processing. The essential difference between these two approaches is whether a model is used to fit the bounding contours of the walker.

The model-free approaches gained a rapid development in the early days. L. Wang, et al. [10] put forward a recognition method based on shape analysis, and presented the static pose changes of these silhouettes over time. Then Procrustes shape analysis is implemented to obtain gait signature. R. T. Collins, et al. [11] presented a key frame identification technology which is view point dependent on the basis of the outline template of the human body. N. Cuntoor, et al. [12] studied the dynamic characteristics of the front view and side view, took the motion features such as the motion of arms, legs and body shaking features into consideration for gait information identification, to a certain extent, improved the recognition rate. In [13, 14], based on the appearance and view point, A. Kale, et al. presented the binarization contour as the feature using the Dynamic Time Warping (DTW) to deal with the speed changes in the process of walking, and strengthen the fault tolerance of original data. In addition, A. Kale, et al. [9] use the width of the outer contour of the binarized silhouette as the image feature and built a HMM model to distinguish the dynamic features of the gait. M. Hu, et al. [15] also built a HMM model for gait recognition. And they take the local binary pattern descriptor as the motion feature. J. W. Davis and A. F. Bobick [16] proposed the temporal template first which is for appearance-based action representation. They used the motion energy images (MEIs) and motion history images (MHIs) to represent motion sequences. In recent years, the depth information of the silhouette is also used as the motion feature for gait recognition. P. Chattopadhyay, et al. [17] put forward a novel feature known as Pose Depth Volume which is just based on the depth frame.

As a model-based approach, D. Cunado, et al. [18, 19] mapped the leg movement of human body to a pendulum

model. It contained a pendulum motion model and a structural model. The lower limb was modeled as two interconnected pendulums. C. Y. Yam, et al. [20] calibrated the rotation of thigh and curves manually, and extracted the transformation of the angles as the feature. A. Yilmaz and M. Shah [21] proposed a model-based method to extract the joints of lower limbs from lateral walking sequences. I. Bouchrika and M. S. Nixon [22] studied the effects of covariates, such as footwear, load carriage, clothing and walking speed for gait recognition. The accuracy of the model-based method is not as high as the one based on the contour, however, the fault tolerance is strengthened. The method proposed in this paper is also model based, and the model we used refers to [23].

In recent years, with the development of 3D technology, more and more research focus on the motion features analysis on 3D data. The 3D data are more accurate than the traditional video data and include more information of human gait. G. Ariyanto and M. S. Nixon [8] used the Soton Multi-biometrics Tunnel to acquire 3D data, and an articulated cylinder model was built to analyze the features such as height, stride and footprint poses. J. Gu and X. Ding [24] addressed a common viewpoint-free framework for action recognition and human identification from gaits. They used a vision-based markerless pose recovery system to extract 3D human joints from 3D reconstructed volume data. D. Gafurov [4] extracted 3D joint information by wearable sensors for gait recognition and gait security research. In this paper, our method has been tested on a 3D database and the result shows a good adaptability of our method on 3D data.

## III. FEATURES EXTRACTION

### A. Extracting Motion Angles

In this paper, the human model is built according to the Fig.1 [24], and only motion angles of lower limbs are used which affect the gait greatly and are more powerful in distinguishing capability. The thigh and knee angles shown in Fig.1 are used for gait analysis in our method.  $\theta_{thigh1}(t)$  and  $\theta_{thigh2}(t)$  are noted as left thigh angle and right thigh angle respectively.  $\theta_{knee1}(t)$  and  $\theta_{knee2}(t)$  are noted as left knee angle and right knee angle respectively.

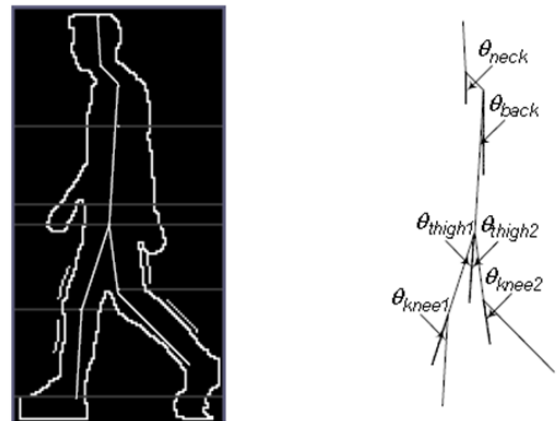


Fig.1 Gait signature and joint angles

In 3D database the angles can be read from the AMC files directly. As for the 2D data, the angles can be extracted from every frame of the gait video by image processing, and the method we used are motivated by [23].

### 1) Estimating motion angles

Let  $\mathbf{F}(\mathbf{X}, \mathbf{Y}, t)$  be the binary frame read from the database directly, and  $(\mathbf{X}_{sil}, \mathbf{Y}_{sil})$  are the pixels in the area of the human silhouette. The lower limbs positions are estimated following the formulas according to [24]:

$$\begin{cases} y_{hip} = \min(\mathbf{Y}_{sil}) + 0.50 \cdot H \\ y_{knee} = \min(\mathbf{Y}_{sil}) + 0.75 \cdot H \\ y_{ankle} = \min(\mathbf{Y}_{sil}) + 0.90 \cdot H \end{cases} \quad (1)$$

where  $H$  is the silhouette's height.

As for the  $t$ th frame, when the legs are not overlapped, there exist two sequence of  $\mathbf{X}_{sil}$  with  $\mathbf{Y}_{sil} = y_{knee}$ . Under this condition the two shins are defined by

$$\begin{cases} \mathbf{Y}_{shinl} = [y_1, y_2, \dots, y_s] \\ \mathbf{X}_{shinl} = [x_{1l}, x_{2l}, \dots, x_{sl}] \end{cases} \quad (2)$$

where  $l = \{1, 2\}$  refers to left or right shin, and

$$\begin{cases} x_{sl} = \frac{\sum_{j=1}^c x_j \cdot F_{shinl}(x_j, y_s, t)}{\sum_{j=1}^c F_{shinl}(x_j, y_s, t)} \\ \mathbf{Y}_{shinl} = [y_{knee}, y_{knee} + 1, \dots, y_{ankle}] \end{cases} \quad (3)$$

with  $c$  is the number of data in each sequence of  $\mathbf{X}_{sil}$ .

Then the shins can be linearly approximated by the first order polynomial with coefficients as follows

$$\mathbf{P}_l(x_{shinl}, t) = p_{l0}(t) + p_{l1}(t) \cdot x_{shinl}(t) \quad (4)$$

The angle between the shins and the vertical axis is

$$\theta_{shinl}(t) = \pi - \tan^{-1} p_{l1}(t) \quad (5)$$

The hip position is defined by

$$\begin{cases} x_{hip} = \frac{1}{p} \cdot \sum_{j=1}^p x_j \\ y_{hip} = y_{knee} \end{cases} \quad (6)$$

where  $p$  is the number of  $\mathbf{X}_{sil}$  with  $\mathbf{Y}_{sil} = y_{hip}$ .

The coefficients satisfy the equations below

$$\begin{cases} y_{hip}(t) = q_{l0}(t) + q_{l1}(t) \cdot x_{hip}(t) \\ y_{knee}(t) = q_{l0}(t) + q_{l1}(t) \cdot x_{l1}(t) \end{cases} \quad (7)$$

Finally, we can calculate the motion angles by

$$\begin{cases} \theta_{thighl}(t) = \pi - \tan^{-1} q_{l1}(t) \\ \theta_{kneel}(t) = \theta_{thighl}(t) + \theta_{shinl}(t) \end{cases} \quad (8)$$

When the legs are overlapped, there exist only one sequence of  $\mathbf{X}_{sil}$  with  $\mathbf{Y}_{sil} = y_{knee}$ . In this case, we ignore the value of the pixel, but treat it as a signal to exchange the positions of right leg and left leg.

### 2) Curve fitting

A sequence of  $\theta_{thighl}(t)$  and  $\theta_{kneel}(t)$  are extracted from the video after motion angles estimation. However the frame rate can't meet the requirement of our recognition method, and more frames are interpolated for one single gait cycle. The following trigonometric-polynomial functions are used to fit the angles [25]:

$$\begin{cases} \theta_{thighl}(t) = b_l + a_{1l} \cdot \sin t + b_{1l} \cdot \cos t \\ \theta_{kneel}(t) = b'_l + \sum_{n=1}^2 [a'_{nl} \cdot \sin(nt) + b'_{nl} \cdot \cos(nt)] \end{cases} \quad (9)$$

Fig.2 and Fig.3 show the fitting results. Then we resample the angles more frequently, and ten times of the original frequency in our experiment.

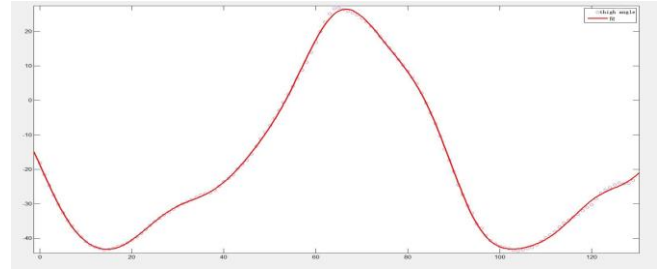


Fig.2 Curve for thigh angle

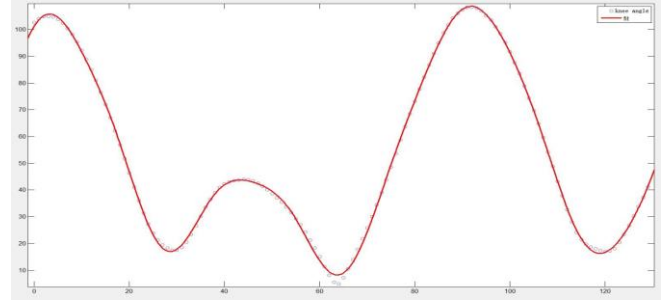


Fig.3 Curve for knee angle

### B. Extracting Feature

Based on the obtained motion angles, the feature will be extracted. In some related work, Fourier Descriptors [26] were used to present the feature of these motion angles. However, the correct classification rate based on Fourier Descriptors is proven to be limited. So in this paper, the feature in time domain is extracted, which is new to gait recognition. To demonstrate the effectiveness of the feature, a series of data are analyzed based on 3D database and 2D database. In the 3D database, all of the data labeled as walk have been analyzed. As for the 2D database we select 50 persons randomly with 6 motion segments each. The analysis method is organized in the following way. The motion angles are extracted as vectors of angle pair

$$\begin{cases} \theta_{thigh} = \{(\theta_{thigh1}(t), \theta_{thigh2}(t))\} \\ \theta_{knee} = \{(\theta_{knee1}(t), \theta_{knee2}(t))\} \end{cases} \quad (10)$$

Fig.4 and Fig.5 show the analysis results on motion angles from 3D database, where x axis is right knee angles and y axis is left knee angles.

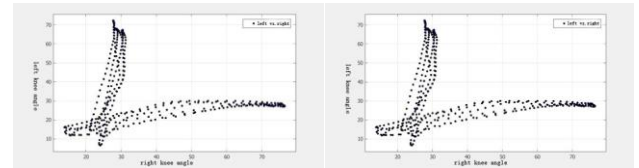


Fig.4 Knee angles trajectory 1

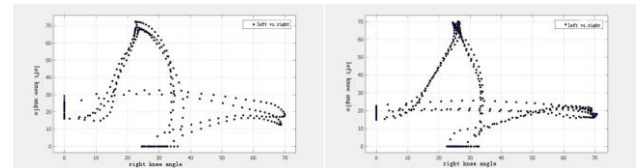


Fig.5 Knee angles trajectory 2

The two trajectories in Fig.4 are formed by the knee angles which belong to one person but different motion segments.

Each point represents one knee angle pair of one frame, and the trajectory represents the spatial distribution of the knee angles. Fig.5 corresponds to another person, where the trajectory is totally different from Fig.4.

Regarding to the motion angles analysis on 2D data, we select 4 out of 50 persons randomly with 5 motion segments each and the corresponding results are shown in Fig.6(a)-Fig.6(d). Due to the raw data noises and computation tolerance, some trajectories of the same person seem a little different, however, most of the data show pattern aggregation and can be used in gait recognition.

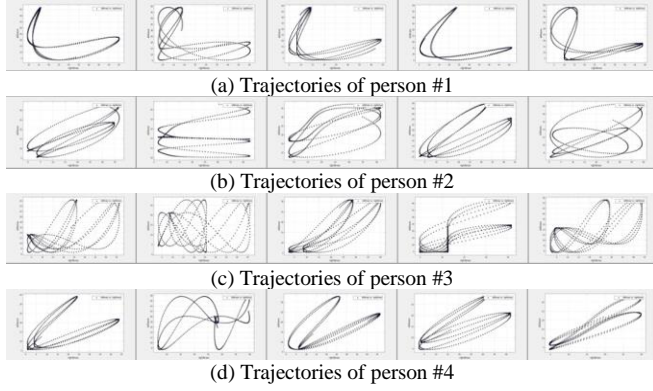


Fig.6 2D examples of knee angle trajectories

Based on the data analysis, the trajectory is a two-value image. In this paper, a histogram is presented and built based on variance of the motion angles to describe the characteristics of the trajectory. The histograms construction is described below:

*Step1:* Obtain the motion angle pairs in Eq.(10).

*Step2:* Build motion angles datasets by

$$\begin{cases} S_{thigh1}(i) = \{\theta_{thigh1}(t') | |\theta_{thigh2}(t') - \theta_{thigh2}(i)|\} \\ S_{thigh2}(i) = \{\theta_{thigh2}(t') | |\theta_{thigh1}(t') - \theta_{thigh1}(i)|\} \\ S_{knee1}(i) = \{\theta_{knee1}(t') | |\theta_{knee2}(t') - \theta_{knee2}(i)|\} \\ S_{knee2}(i) = \{\theta_{knee2}(t') | |\theta_{knee1}(t') - \theta_{knee1}(i)|\} \end{cases} \quad (11)$$

where  $t'$  and  $i$  are the index of the interpolated frames. For example in  $S_{thigh1}(i)$  for every  $\theta_{thigh2}(i)$  we can get a list of  $\theta_{thigh1}(t')$ .

*Step3:* Calculate the variance of each dataset. For example, as for each list of  $S_{thigh1}(i)$ , the variance is calculated by

$$Var = \frac{\sum_{t'=1}^n (\theta_{thigh1}(t') - \overline{\theta_{thigh1}})^2}{n} \quad (12)$$

where  $\overline{\theta_{thigh1}} = \frac{\sum_{t'=1}^n \theta_{thigh1}(t')}{n}$  and  $n$  is the number of data in  $S_{thigh1}(i)$ .

Therefore the four histograms based on left knee, right knee, left thigh and right thigh are obtained, which present the postures of legs during walking and are used as the features for gait recognition. Fig.7 is an example of the obtained histogram.

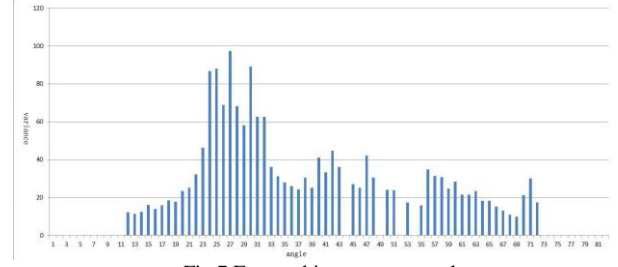


Fig.7 Feature histogram example

## IV. CLASSIFICATION

### A. Distance Measurement

The gait is represented by a feature histogram, so the distance of two gaits can be defined as the similarity between two histograms. Many suitable functions can be used to calculate the similarity. In this paper, the correlation coefficient function,  $\chi^2$  function and  $L_1$  distance are chosen, which have been extensively used for histogram comparison [27].

#### 1) Correlation coefficient function

In statistics the correlation coefficient is a measure of the linear correlation between two variables, and belongs to  $[-1, +1]$  where  $+1$  is total positive correlation,  $0$  is no correlation, and  $-1$  is total negative correlation. We let  $H_1$  and  $H_1'$  denote the two corresponding histograms and  $I$  is the index of the bins. The formula for correlation coefficient is

$$r(H_1, H_1') = \frac{\sum_I (H_1(I) - \overline{H_1})(H_1'(I) - \overline{H_1'})}{\sqrt{\sum_I (H_1(I) - \overline{H_1})^2 \sum_I (H_1'(I) - \overline{H_1'})^2}} \quad (13)$$

where  $\overline{H_1} = \frac{1}{N} \sum_I H_1(I)$  and  $\overline{H_1'} = \frac{1}{N} \sum_I H_1'(I)$ ,  $N$  is the number of bins.

The distance between two histograms  $d_r(H_1, H_1')$  is defined as

$$d_r(H_1, H_1') = \frac{2}{r(H_1, H_1') + 1} - 1 \quad (14)$$

The value of  $d_r(H_1, H_1')$  is between  $[0, +\infty)$ .

#### 2) $\chi^2$ function

$\chi^2$  function is used for statistical test and evaluate the difference between two data sets. We use here a symmetrized approximation of  $\chi^2$ :

$$d_{\chi^2}(H_1, H_1') = \sum_I \frac{(H_1(I) - H_1'(I))^2}{H_1(I) + H_1'(I)} \quad (15)$$

When  $H_1(I)$  and  $H_1'(I)$  are both equal to 0,  $d_{\chi^2}(H_1, H_1') = 0$ .

#### 3) $L_1$ distance

The  $L_1$  distance is also known as Manhattan distance and defined as:

$$L_1(H_1, H_1') = \sum_I |H_1(I) - H_1'(I)| \quad (16)$$

As the distribution interval of the bin-values is not fixed, the normalization is needed for evaluating the distance between two histograms. The normalized formula is

$$d_{L_1}(H_1, H_1') = \sum_I \frac{|H_1(I) - H_1'(I)|}{\max(H_1(I), H_1'(I))} \quad (17)$$

When  $H_1(I)$  and  $H_1'(I)$  are both equal to 0, the value is equal to 0.



In order to compare similarity of the motion segments, the corresponding three distance vectors should be calculated by

$$\begin{cases} D_{X^2}(H, H') = (d_{X^2}(H_{thighl}, H'_{thighl}), d_{X^2}(H_{kneel}, H'_{kneel})) \\ D_{L_1}(H, H') = (d_{L_1}(H_{thighl}, H'_{thighl}), d_{L_1}(H_{kneel}, H'_{kneel})) \\ D_r(H, H') = (d_r(H_{thighl}, H'_{thighl}), d_r(H_{kneel}, H'_{kneel})) \end{cases} \quad (18)$$

One of the three vectors or the combination of them with a Boolean flag will be taken as the input data of the next classification, where the Boolean flag represents whether the two motion segments come from the same person.

### B. Classification

In [27], SVM classifier is proven to be superior in solving the histogram-based classification problem. The SVM is a supervised learning model with associated learning algorithms that analyze data and recognize patterns, and be used for binary classification. The key of using SVM is to select a kernel carefully as an inappropriate kernel may lead to poor performance. In our experiments four types of kernel are selected.  $K_{poly}$  stands for polynomial kernel,  $K_{RBF}$  for a radial basis function (RBF),  $K_{sig}$  for the sigmoid kernel, and  $K_{line}$  for the linear kernel. We assume  $(a_i, b_i)$  to be the input sample, where  $b_i \in \{0,1\}$  is the flag used for labeling the category, 0 stands for different persons and 1 for same person. In this paper, the experiments on the regular expression  $a_i = (D_{X^2}?, D_{L_1}?, D_r?)$  and  $a_i \neq \emptyset$  are implemented. Besides the kernel type and the input vector, the different training sets are also discussed. The classification experiments will attempt to achieve the best combination with the highest correct classification rate.

## V. EXPERIMENTS

### A. Database

2D database and 3D database are used for our experiments.

#### 1) 2D Database

The 2D gait database we used is from CASIA Gait Database Dataset B [5], which was created in 2005 and used a lot in gait recognition research in recent years. There are 124 subjects and the data was captured from 11 views. The angle interval between two adjacent views is  $18^\circ$ , and the view angle is  $0^\circ, 18^\circ, 36^\circ, 54^\circ, 72^\circ, 90^\circ, 108^\circ, 126^\circ, 144^\circ, 162^\circ$  and  $180^\circ$  from left to right. And there are three variations, i.e. view angle, clothing and carrying condition, which are separately considered. The captured sample frames are shown in Fig.8. In this paper, we use the side view frame and the view angle is  $90^\circ$  in the database.

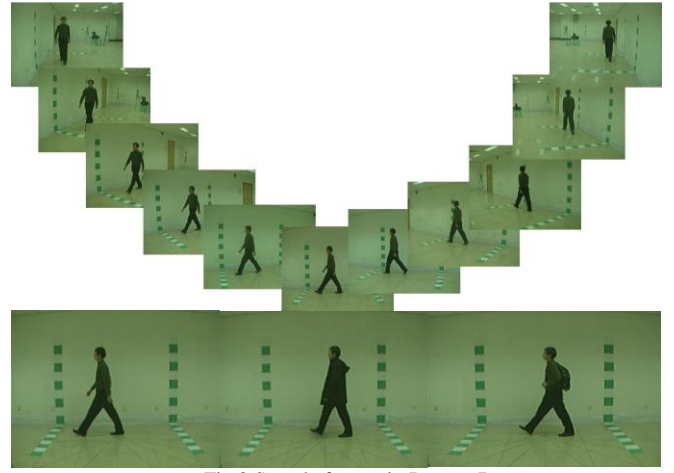


Fig.8 Sample frames in Dataset B

#### 2) 3D Database

3D Data used in our experiment is from CMU motion capture database [6]. The data is captured using 12 Vicon infrared MX-40 cameras, each of which is capable of recording 120 Hz with images of 4 megapixel resolutions. The cameras are placed around a  $3m \times 8m$  rectangular area. Humans wear a black jumpsuit and have 41 markers shown in Fig.9. The skeleton model used in the system is shown in Fig.10 to build 3D capture data. The data is provided in ASF/AMC format with a motion type label, such as walk, run and jump. Our experiments use the motion of walk for gait recognition. The index number of data can tell us which person the motion belongs to.



Fig.9 Marker set in front and back view

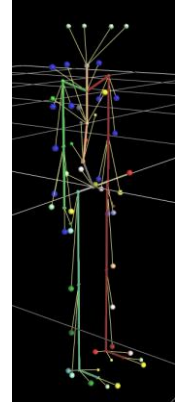


Fig.10 Skeleton model

### B. Experiments on 2D database

In this experiment, the feature is extracted from the CASIA Gait Database Dataset B. This database is widely used in recent years and many papers were published based on this dataset. In order to make comparison with others, only the lateral view result is used in this paper. Based on the walking status, two experiments are designed, where one is conducted on the normal walking data and the other is on the mixed data with normal walking data, walking with a coat and walking with a bag.

### 1) Experiments on the normal walking data

These experiments are designed to demonstrate the validity of our method on normal walking data, meanwhile how the three factors (SVM kernel, distance function and training set) affect CCR will be discussed. Here a total of 124 persons with 6 motion segments each are used for feature extraction. Each pair of them is made comparison calculation, and a total of  $C_{124 \times 6}^2$  comparisons are obtained. In [28], C. Chen discussed several recognition methods and the recognition rate achieved using the same data. The result is shown in Table 1.

Table 1 Performance comparison of LTSM, HMM and DT

Angle	PHMM	Frieze feature			Wavelet feature		
		DT	HMM	LTSM	DT	HMM	LTSM
90	90.3%	58.1	91.1	95.2	61.3	90.3	95.2

In the first experiment, one motion segment is selected randomly and compared with other 457 motion segments in the dataset for analyzing whether they belong to the same person. The regularization parameter (C) is set to 400. The training set keeps same during the experiments. And it contains 30 comparisons in which the segments come from the same person or come from different persons both contained. The performance under different SVM kernels and distance functions are evaluated in order to achieve the highest CCR.

Table 2 shows the CCR results under the same training set and testing set, and different SVM kernel and distance function. It is noted that the highest CCR of 99.12% is reached under the combination of  $K_{sig}$  and  $D_{\chi^2}$ , which performs better than 95.20% in Table 1. The second highest CCR achieved by the single distance is up to 96.06%. The high CCR shows the advantage of our method working on normal walking data. Meanwhile the results in Table 2 also show the fact that increasing input dimensions does not improve the CCR but increase the computing complexity. So combination of the distance input cannot obtain a better performance. As for the normal walking data,  $D_{\chi^2}$  and  $D_{L1}$  are proven to be more effective than other distance vectors. Moreover among all of the kernels,  $K_{sig}$  achieves the highest average CCR. So the next experiments will be designed on the combination of  $K_{sig}$  with  $D_{\chi^2}$  and  $D_{L1}$ .

Table 2 Experimental results on normal walking data

Distance vector	SVM Kernel			
	$K_{poly}$	$K_{RBF}$	$K_{sig}$	$K_{line}$
$D_{\chi^2}$	81.40%	78.34%	99.12%	84.25%
$D_{L1}$	84.46%	90.59%	96.06%	70.20%
$D_r$	67.40%	77.68%	86.21%	91.03%
$(D_{\chi^2}, D_{L1})$	89.93%	77.02%	84.68%	86.87%
$(D_{L1}, D_r)$	62.80%	76.15%	99.12%	75.93%
$(D_{\chi^2}, D_r)$	84.46%	93.22%	82.06%	96.06%
$(D_{\chi^2}, D_{L1}, D_r)$	83.59%	77.46%	98.25%	77.02%

In order to explore that how the training set will influence the CCR, the second experiment is carried out. We select one motion segment randomly and compare with all of the motion segments in the dataset including itself, and obtain a total of  $124 \times 6 = 744$  comparisons. These comparisons form the

testing set in the experiment. The regularization parameter (C) is set to 400.  $K_{sig}$  is taken as the kernel,  $D_{\chi^2}$  and  $D_{L1}$  are used to measure the distance. Four training sets are selected from all the comparisons and follow the rules below.

$ts_1$ : 60 comparisons including 30 data with  $b_i = 0$ .

$ts_2$ : 60 comparisons including 40 data with  $b_i = 0$ .

$ts_3$ : 5000 comparisons including 2500 data with  $b_i = 0$ .

$ts_4$ : 5000 comparisons including 3400 data with  $b_i = 0$ .

The classification results are presented in Table 3. Under the same condition the CCR varies with the training set, which means the quality of training set affects CCR. Among all the training sets,  $ts_3$  achieves the lowest CCR, which indicates that the amount of training set samples is not a necessary factor for a high quality training set. In addition, although the scale of  $ts_4$  and  $ts_2$  different greatly, the two CCRs are close to each other. The CCR of  $ts_4$  is higher than  $ts_2$  under  $D_{\chi^2}$ , however, the case is the contrary as for  $D_{L1}$ . The other aspect to affect the CCR is the proportion of the two kinds of training data ( $b_i = 0$  or  $b_i = 1$ ). The number of the data with  $b_i = 0$  in  $ts_2$  and  $ts_4$  is twice as much as in  $ts_1$  and  $ts_3$ , and the corresponding CCRs are higher. So the proportional relationship of different kinds of samples is the key factor for high quality training set, and when the data with  $b_i = 0$  and  $b_i = 1$  remains a proportion of 2 to 1, the CCR is higher.

Table 3 Classification results on different training set

Distance vector	Training set			
	$ts_1$	$ts_2$	$ts_3$	$ts_4$
$D_{\chi^2}$	93.03%	98.87%	89.45%	99.44%
$D_{L1}$	96.05%	99.62%	90.96%	96.23%

### 2) Experiments on the mixed data

The main purpose of this experiment is to discuss how the clothes affect the recognition results. In previous work clothing types are always treated as an important factor, as they can influence the overall body shape and some certain types of clothing can even affect the way a person walks. In [22] a significant drop in performance (87% to 60%) was reported when the person wore a coat on top of the normal clothes. So that evaluating whether our method is clothes sensitive or not is necessary. CASIA Gait Database Dataset B contains three types of data, i.e. normal walking data, walking data with a bag and walking data with a coat, and a total of  $124 \times 10 = 1240$  motion segments. Here all of these data are mixed together to evaluate our method. Some previous work also made effort on the clothes problem, in [29] several methods were discussed on the CASIA Gait database with clothes type considered. The results in [29] are shown in Table 4.

Table 4 Recognition accuracy using different approaches

Dataset	Recognition accuracy (%)				
	Proposed method	Yan-qin Liu	Su-li xu	Xiaoxiang Li	Khalid Bashir
CASIA dataset B	91.50	83.00	89.70	89.29	55.00



Our experimental results are shown in Table 5, where the training dataset is fixed to 60 motion segments randomly selected from the database. The regularization parameter (C) is set to 400. The testing data are the mixed data including the motion segments randomly selected from the normal walking data, from the walking data with a coat, and from the walking data with a bag respectively. Then all of these three motion segments are compared with whole dataset. So as for each element in Table 5, which is implemented on a total of  $124 \times 10 = 1240$  comparisons.

Table 5 Experimental results on mixed data

Distance vector	SVM Kernel			
	$K_{poly}$	$K_{RBF}$	$K_{sig}$	$K_{line}$
$D_{\chi^2}$	95.94%	98.38%	98.16%	98.16%
$D_{L_1}$	77.66%	78.19%	99.44%	78.94%
$D_r$	52.10%	50.37%	79.12%	55.82%

Table 5 shows our experimental results. It is noted that by using the high quality training set, the CCRs under  $D_{\chi^2}$  all achieve to 95% and above. The best CCR is 99.44% by  $D_{L_1}$  and  $K_{sig}$ . All of these results under  $D_{\chi^2}$  are much higher than the CCR presented in Table 4. The results demonstrate that our method works well on the mixed data and is not clothes sensitive. The experimental results also show that  $D_{L_1}$  and  $D_r$  perform worse than  $D_{\chi^2}$  on the mixed data. Based on the experiments on the 2D database,  $D_{\chi^2}$  is supposed to be used as the distance measurement and can obtain better recognition result.

### C. Experiments on 3D database

In this section the 3D database is used to evaluate the performance of our method. We select 54 motion segments labeled as walk from the database which come from 8 persons. Every pair of the motion segments are compared since it is a small dataset, make it a total of 1458 comparisons in this experiment. The experiments are divided into 2 parts. The first aim to evaluate the effectiveness of our method on the 3D data, and the second are designed to show how the training set would influence the CCR.

#### 1) Experiments of performance on 3D data

A training set containing 150 comparisons is used for these experiments, which is selected randomly from the dataset. And the rest of the 1308 comparisons are treated as the testing set. Four kinds of kernels and seven kinds of distance vectors are implemented as the elements in the experiment for achieving the highest CCR. The results are shown in Table 6.

Table 6 Experimental results on 3D Data

Distance vector	SVM Kernel			
	$K_{poly}$	$K_{RBF}$	$K_{sig}$	$K_{line}$
$D_{\chi^2}$	94.46%	94.53%	91.82%	94.61%
$D_{L_1}$	93.50%	93.23%	77.52%	93.73%
$D_r$	66.44%	67.55%	68.85%	67.55%
$(D_{\chi^2}, D_{L_1})$	94.61%	94.84%	80.20%	95.60%
$(D_{L_1}, D_r)$	72.55%	95.07%	62.08%	81.96%
$(D_{\chi^2}, D_r)$	68.31%	70.72%	77.52%	70.80%
$(D_{\chi^2}, D_{L_1}, D_r)$	71.79%	95.68%	65.41%	82.57%

Table 6 shows a quite different pattern compared with the result on 2D database. The situation of dataset (amount of persons and motion segments), the raw data error and the data format may lead to this difference. However the CCR on 3D data achieves up to 95.6% under the distance vector of  $(D_{\chi^2}, D_{L_1}, D_r)$  with  $K_{RBF}$  and  $(D_{\chi^2}, D_{L_1})$  with  $K_{line} \cdot K_{sig}$  performs the best on 2D data but not on 3D data. Instead of  $K_{sig}$ ,  $K_{RBF}$  become a proper choice for gait recognition in the 3D environment. G. Ariyanto and his fellows achieved to 79.4% correct classification rate in their paper [8]. But the data we used in the experiment is different, while the data collection way is the same.

#### 2) Experiments on different training sets

The experiments on 2D database show that the CCR get influenced greatly by the quality of training set. This experiment aims to evaluate the performance on 3D data of different training set. 1000 comparisons are selected from the dataset randomly as the testing data, and the subset of the rest data forms the training set. The rules below are followed for the four training sets.

$ts_1$ : 150 comparisons, where the data is from 8 persons.

$ts_2$ : 150 comparisons, where the data is from 8 persons but the motion segments are different from  $ts_1$ .

$ts_3$ : 150 comparisons, where the data is from 4 persons.

$ts_4$ : 300 comparisons, where the data is from 4 persons.

In the experiment  $K_{RBF}$  is chosen to be the kernel as its performance is the best according to the former experiments. The results on different training sets are shown in Table 7.

Table 7 Experimental results on different 3D training sets

Distance vector	Training set			
	$ts_1$	$ts_2$	$ts_3$	$ts_4$
$D_{\chi^2}$	94.61%	94.53%	84.25%	82.30%
$D_{L_1}$	92.89%	93.23%	77.91%	72.71%
$D_r$	69.57%	67.55%	56.50%	58.12%
$(D_{\chi^2}, D_{L_1})$	93.12%	94.84%	84.94%	81.95%
$(D_{L_1}, D_r)$	93.27%	95.07%	76.22%	72.71%
$(D_{\chi^2}, D_r)$	71.10%	70.72%	60.63%	62.44%
$(D_{\chi^2}, D_{L_1}, D_r)$	93.43%	95.68%	78.98%	75.82%

The results show that the number of the class in training set is more important than the number of motion segments. More segments from different persons can achieve a higher CCR. The accuracy of  $D_r$  is much less than others, moreover,  $D_{\chi^2}$  and  $(D_{\chi^2}, D_{L_1})$  perform the best in average. The highest CCR is achieved by  $(D_{\chi^2}, D_{L_1}, D_r)$  under the second training set  $ts_2$ , however the low quality of the training set can decrease the CCR more than  $D_{\chi^2}$  and  $(D_{\chi^2}, D_{L_1})$ . Therefore  $D_{\chi^2}$  and  $(D_{\chi^2}, D_{L_1})$  are the better distance vector used in the 3D environment and  $K_{RBF}$  would be the choice of kernel. As for the training set, it should cover as many persons as possible to guarantee the recognition accuracy.

## VI. CONCLUSION

In this paper, a gait recognition method based on the joint distribution of motion angles is proposed. The distribution characteristic of the data in time domain is presented and the feature histograms are built for gait recognition. Three distance measurements are implemented as the input vectors and then a SVM classifier is built to perform the classification. Experiments are conducted both on 2D video gait database and 3D motion capture database to evaluate the performance of our method. The results show a high CCR and the effectiveness of the proposed method both on 2D and 3D databases. In addition, the quality of training set is analyzed and discussed. The results show the importance of proportional relationship of different classes in the training set as for 2D database and the importance of covering more persons on 3D database. This paper only considers the lateral view data in the database, and extending the method to multiple view angles will be our further work.

## ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundations of China (No.61100143, 61370128, and 61272353), Program for New Century Excellent Talents in University (NCET-13-0659), Fundamental Research Funds for the Central Universities (2014JBZ004), Beijing Higher Education Young Elite Teacher Project (YETP0583).

## REFERENCES

- [1] C. Chen, J. Liang, X. Zhu, "Gait recognition based on improved dynamic Bayesian networks," *Pattern Recognition*, Vol. 44, No. 4, pp. 988-995, 2011.
- [2] [http://en.wikipedia.org/wiki/Gait\\_analysis](http://en.wikipedia.org/wiki/Gait_analysis)
- [3] C. P. Lee, A. W. C. Tan, S. C. Tan, "Gait recognition via optimally interpolated deformable contours," *Pattern Recognition Letters*, Vol. 34, No. 6, pp. 663-669, 2013.
- [4] D. Gafurov, E. Snekenes, "Gait recognition using wearable motion recording sensors," *EURASIP Journal on Advances in Signal Processing*, Vol. 2009, No. 7, 2009.
- [5] <http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp>
- [6] <http://mocap.cs.cmu.edu/search.php>
- [7] A. F. Bobick, A. Y. Johnson, "Gait recognition using static, activity-specific parameters," *Computer Vision and Pattern Recognition*, 2001, Vol. 1, pp. 423-430, 2001.
- [8] A. Kale, A. Sundaresan, A. N. Rajagopalan, et al. "Identification of humans using gait," *IEEE Transactions on Image Processing*, Vol. 13, No. 9, pp. 1163-1173, 2004.
- [9] G. Ariyanto, M. S. Nixon, "Model-based 3D gait biometrics," 2011 International Joint Conference on Biometrics, Washington, DC, USA, pp. 1-7, 2011.
- [10] L. Wang, H. Ning, W. Hu, et al. "Gait recognition based on procrustes shape analysis," in *Proc. International Conference on Image Processing*, Vol. 3, pp. 433-436, 2002.
- [11] R. T. Collins, R. Gross, J. Shi, "Silhouette-based human identification from body shape and gait," in *Proc. Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, DC, USA, pp. 366-371, 2002.
- [12] N. Cuntoor, A. Kale, R. Chellappa, "Combining multiple evidences for gait recognition," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, pp. III-33-6, 2003.
- [13] A. Kale, A. N. Rajagopalan, N. Cuntoor, et al. "Gait-based recognition of humans using continuous HMMs," in *Proc. Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, DC, pp. 336-341, 2002.
- [14] A. Kale, N. Cuntoor, B. Yegnanarayana, et al. "Gait analysis for human identification," *Audio- and Video-Based Biometric Person Authentication*, pp. 706-714, 2003.
- [15] M. Hu, Y. Wang, Z. Zhang, et al. "Incremental learning for video-based gait recognition with LBP flow," *IEEE Transactions on Cybernetics*, Vol. 43, No. 1, pp. 77-89, 2013.
- [16] J. W. Davis, A. F. Bobick, "The representation and recognition of human movement using temporal templates," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 928-934, 1997.
- [17] P. Chattopadhyay, A. Roy, S. Sural, et al. "Pose depth volume extraction from RGB-d streams for frontal gait recognition," *Journal of Visual Communication and Image Representation*, Vol. 25, No. 1, pp. 53-63, 2014.
- [18] D. Cunado, J. M. Nash, M. S. Nixon, et al. "Gait extraction and description by evidence-gathering," in *Proc. of the Second International Conference on Audio- and Video-Based Biometric Person Authentication*, Washington, DC, USA, pp. 43-48, 1999.
- [19] D. Cunado, M. S. Nixon, J. N. Carter, "Automatic extraction and description of human gait models for recognition purposes," *Computer Vision and Image Understanding*, Vol. 90, No.1, pp. 1-41, 2003.
- [20] C. Y. Yam, M. S. Nixon, J. N. Carter, "Gait recognition by walking and running: a model-based approach," in *Proc. Asian Conference on Computer Vision*, pp. 1-6, 2002.
- [21] A. Yilma, M. Shah, "Recognizing human actions in videos acquired by uncalibrated moving cameras," *Tenth IEEE International Conference on Computer Vision*, Beijing, China, Vol. 1, pp. 150-157, 2005.
- [22] I. Bouchrika, M. S. Nixon, "Exploratory factor analysis of gait recognition," *8th IEEE International Conference on Automatic Face & Gesture Recognition*, Amsterdam, the Netherlands, pp. 1-6, 2008.
- [23] M. Goffredo, I. Bouchrika, J. N. Carter, et al. "Performance analysis for automated gait extraction and recognition in multi-camera surveillance," *Multimedia Tools and Applications*, Vol. 50, No. 1, pp. 75-94, 2010.
- [24] J. Gu, X. Ding, S. Wang, et al. "Action and gait recognition from recovered 3-D human joints," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 40, No. 4, pp. 1021-1033, 2010.
- [25] J. H. Yoo, M. S. Nixon, "Automated Markerless Analysis of Human Gait Motion for Recognition and Classification," *Journal of the Electronics and Telecommunications Research Institute (ETRI)*, Vol.33, No. 2, pp. 259-266, 2011.
- [26] S. D. Mowbray, M. S. Nixon, "Automatic gait recognition via Fourier descriptors of deformable objects," *Audio- and Video-Based Biometric Person Authentication*, Vol. 2688, pp. 566-573, 2003.
- [27] O. Chapelle, P. Haffner, V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Transactions on Neural Networks*, Vol. 10, No. 5, pp. 1055-1064, 1999.
- [28] C. Chen, J. Liang, X. Zhu, "Gait recognition based on improved dynamic Bayesian networks," *Pattern Recognition*, Vol. 44, No. 4, pp. 988-995, 2011.
- [29] M. K. HP, H. S. Nagendraswamy, "Change Energy Image for Gait Recognition: An Approach Based on Symbolic Representation," *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, Vol. 6, No.4, pp.1-8, 2014.

# Exposing Image Forgery by Detecting Traces of Feather Operation

Jiangbin Zheng, Tingge Zhu, Zhe Li  
Dept. of Computer Science and Engineering,  
School of Computer, Northwestern  
Polytechnical University  
Xi'an, China  
zhengjb0163@163.com

Weiwei Xing  
School of Software Engineering,  
Beijing Jiaotong University  
Beijing, China  
wxing@bjtu.edu.cn

JinChang Ren  
Dept. of Electronic and Electrical  
Engineering, University of  
Strathclyde Glasgow  
G1 1XW, United Kingdom  
jinchang.ren@strath.ac.uk

**Abstract**—Powerful digital image editing tools make it very easy to produce a perfect image forgery. The feather operation is necessary when tampering an image by copy-paste operation because it can help the boundary of pasted object to blend smoothly and unobtrusively with its surroundings. We propose a blind technique capable of detecting traces of feather operation to expose image forgeries. We model the feather operation, and the pixels of feather region will present similarity in their gradient phase angle and feather radius. An effectual scheme is designed to estimate each feather region pixel's gradient phase angle and feather radius, and the pixel's similarity to its neighbor pixels is defined and used to distinguish the feathered pixels from unfeathered pixels. The degree of image credibility is defined, and it is more acceptable to evaluate the reality of one image than just using a decision of YES or NO. Results of experiments on several forgeries demonstrate the effectiveness of the technique.

**Keywords**—*Tampered image; image forgery; feather operation; tampering detection.*

## I. INTRODUCTION

With the great increasing usage of digital photography and the advancing of new technologies, powerful image editing software make it very easy today to create a believable image forgery even for a non-specialist. Today more and more images of high quality are presented on screen. It challenges our ability to tell what's real and what's not. Some samples of image forgeries obtained over the internet are shown in Fig.1.

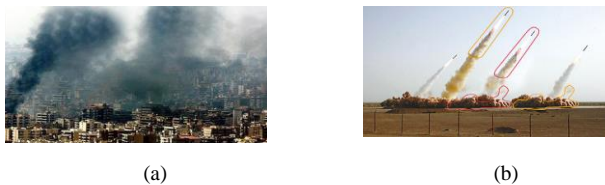


Fig.1.Examples of image forgeries obtained from internet (a) the published Reuters photograph showing the remnants of an Israeli bombing. (b) the published photograph showing four Iranian missiles streaking skyward.

In August of 2006, the Reuters news agency published a photograph [Fig.1.(a)] showing the remnants of an Israeli bombing of a Lebanese town, and in the week that followed, the photograph was revealed by nearly every major news organization to have been doctored with the addition of more smoke. A more recent example of photo tampering came to light in July 2008. A photograph [Fig.1.(b)] showing four

Iranian missiles streaking skyward was first posted on the Web site of Sepah News, the media arm of Iran's Revolutionary Guard. But only three of those rockets actually left the ground; a fourth was digitally added.

The appearance of more and more image forgeries has broken up people's long-term confidence on the reality of image. Image security and authentication play an essential role in people's lives, and they are significant in many social areas such as forensic investigation, criminal investigation, insurance processing, surveillance systems, intelligence services, medical imaging and journalism. As a consequence, it is obvious that we should pay special attention to the field of image authenticity.

This paper is organized as follows. the related prior work is summarized in section 2. The most common tampering steps and models of the feather operation are given in section 3. In section 4 and 5 ,we demonstrate the influence of the feather operation on the gradient phase angle and the feather radius for edge pixels. In Section 6, we explain how to compute the degree of image credibility and how to locate the feather regions. In section 7,we describes the experimental results of the proposed algorithm. Finally, the conclusion is made in Section 8.

## II. RELATED PRIOR WORK

Digital watermarking has been proposed as an active approach for providing image authenticity [1]. The drawback of this approach is that a watermark must be inserted when the image is created. Although image forgeries may leave no visual clues, they may alter the underlying statistics of an image by which we may detect suspicious images, these are passive techniques. Recently, the state-of-the-art digital image forensics in the context of three predominant types of forensic are presented in [2] by Tanzeela Qazi, which include copy or move forgery, image splicing and image retouching.

We classify the passive techniques for image forensics into three categories according to different forensic features: techniques based on the traces left by the tampering process, techniques based on the consistency of imaging equipment, and techniques based on the statistical characteristics of natural images. We will review some typical forensic techniques within each category as follows.

### *2.1 Techniques based on the traces left by the tampering process.*

We try to expose the image forgeries by detecting some forensic characters. The common methods in this category include: copy-paste detection, double JPEG detection, re-sampling detection, light direction detection, sharpness / blurriness detection, splicing detection and so on.

One of the common image tampering is to copy and paste portions of the image to conceal a person or object in the scene. The presence of copy-paste region can be the evidence of tampering [3][4]. In [3] and [4], the image is divided into fixed-size overlapping blocks whose features are represented by the discrete cosine transform (DCT) coefficients. Feature dimension reduction is difference between them. Truncating is used to reduce it in [3], however in [4] each block represented by the quantized DCT coefficients is divided into non-overlapping sub-block, the dimension of the SVD (singular value decomposition) based features from each quantized block is reduced by largest singular value. Then the duplicated regions were detected by lexicographically sorting the features vectors and two similar feature vectors are exported. This approach in [5] was available whether the copied area came from the same image or not, if only source image was JPEG compressed. For JPEG format images, it is likely that the manipulated image is compressed twice, and the double compression introduces specific artifacts not present in singly compressed images. Thus, the presence of these artifacts can be the evidence of tampering [6] [7]. H. Farid described a technique [8] to expose JPEG ghosts by detecting whether part of an image was initially compressed at a lower quality than the rest of the image. The re-sampling operations [9][10], such as scaling, revolving, stretching and so on, are often necessary in tampering, and these operations can introduce specific periodic correlations between neighboring pixels, which can be used to detect tampering. When creating a composite of two or more images, it is often difficult to exactly match the lighting, thus, the lighting inconsistencies can be a useful tool for revealing traces of digital tampering. These methods presented in [11] can detect doctored image based on consistency of shadow. In [11], the author proposes two schemes. The first scheme is based on texture consistency of shadow, another method is based on strength of light source of shadows. In [12], the authors described how to estimate a camera's principal point from the image of a pair of eyes or other planar geometric shapes. They showed how translation in the image plane was equivalent to a shift of the principal point. Inconsistencies in the principal point across an image were then used as evidence of tampering. The authors in [13] proposed a method to detect image tampering operations that involved sharpness/blurriness adjustment, and the estimate of sharpness/blurriness value was based on the regularity properties of wavelet transform coefficients. The method proposed in [14] can detect image splicing by evaluating inconsistencies in motion blur. In [15], we proposed a technique based on the local entropy of the gradient to detect the image forgery, and it can discover the traces of artificial feather operation. In [16], we proposed a technique based on

the wavelet holomorphic filtering to recognize some traces of artificial blur operation.

### *2.2 Techniques based on the consistency of imaging equipment.*

When an image is created, the characters brought by imaging equipment should present consistence in the whole natural image, which can be used as evidence of tampering. The common methods include: CFA interpolation detection, sensor noise detection, chromatic aberration detection and camera response detection.

Most digital cameras employ a single sensor in conjunction with a color filter array (CFA), where the missing color samples are then interpolated from these recorded samples to obtain a three-channel color image. This interpolation introduces specific correlations that are likely to be destroyed when an image is tampered with. The presence of lack of correlations produced by CFA interpolation can be used to authenticate an image. The algorithm described in [17] can detect image forgery by estimating color modification in images. the authors in [18] proposed a method base on the observation that each in-camera and post-camera processing operation left some distinct intrinsic fingerprint traces on the final image. They characterized the properties of a direct camera output using a camera model [19], and considered any further post-camera processing as a manipulation filter. The method could be used to verify whether a given digital image was a direct camera output and identified different types of post-camera processing. The authors in [20] modeled camera processing with an additive and multiplicative noise model. The parameters of the noise model were estimated from the original camera or a series of images originating from the known camera. Correlations between the estimated camera noise and the extracted image noise were then used to authenticate an image. The authors in [21] modeled the camera processing with a generic additive noise model and used statistics from the estimated noise for image forensics. M.K.Johnson and H.Farid in [22] indicated that the chromatic aberration resulted from the failure of an optical system to perfectly focus light of different wavelengths, and these aberrations could be used to detect digital tampering by approximated with a low-parameter model. They developed an automatic technique for estimating the model parameters that was based on maximizing the mutual information between color channels. The authors [23] described how to estimate the mapping, termed a response function, from a single image. The differences in the response function across the image were then used to detect tampering. A automatic splicing detection is proposed in [24], which is based a rigorous camera response function (CRF) consistency checking principle.

### *2.3 Techniques based on the statistical characteristics of natural images.*

Natural images are not simply a collection of independent pixels. The visual structures making them look "natural" are the result of strong correlations among pixels. The techniques in this category try to model statistical regularities within natural image, which is used to discriminate photographic from computer-generated images or tampered ones and to detect hidden messages. Wei Lu [25] proposes a detection scheme for

natural images and fake images, in which the support vector machine (SVM) is used to differentiate true and faked images by feature extracted using multi-resolution decomposition and higher order local autocorrelations (HLACs). The authors in [26] developed an image forensic scheme based on the interplay between feature fusion and decision fusion. The authors in [27] proposed a method for digital image forensics based on Binary Similarity Measures between bit planes used as features. The basic idea was that the correlation between the bit planes as well the binary texture characteristics within the bit planes would differ between an original and a doctored image. This change in the intrinsic characteristics of the image could be monitored via the quintal-spatial moments of the bit planes.

In fact, there are no universally applicable solutions due to multifarious tampering means. Vast observation suggests that the feather operation is nearly inevitably in tampering. A novel approach to expose image forgeries is presented in this paper. It works by determining if feather operation was used at the edge of one object and locating the traces of feather operation. In comparison to the previous work, it is computationally much simpler and does not require a large image database to train a classifier. What's more, the degree of image credibility is defined in our approach to judge the reality of an image. We think it more acceptable than just using a decision of yes or no.

### III. THE FEATHER OPERATION MODEL

One of the most common image manipulations is to copy and paste part of one image to another one, and this operation can conceal or add an important person or object in the scene. When this is done, it seems to be potentially discontinuous or obtrusive along the boundary of the pasted region. The feather operation is necessary, and it helps to create a smooth transition between the pasted region and its surroundings. The common method is shown in Fig.2, in which we can create an image forgery by using feather operation, and it can weaken the traces of splicing.

When a region of one image is copy-pasted to another one, new edge with step shape will incur naturally as shown (c) in Fig.2. New edge  $f(x)$  without any post-processing can be simulated by Eq. (1), and the width of the new edge is 0 pixel (Fig.3).

$$f(x) = \begin{cases} Q_1, & x = A \\ Q_2 = Q_1 + H, & x = A + 1 \end{cases} \quad (1)$$

Where,  $x = A$  denotes the position of the new edge,  $Q_1$  and  $Q_2$  denote the edge pixel values of the background and the pasted region,  $H$  denotes the offset.

Given the specified feather radius  $r$ , the feather operation can make a smooth transition between  $Q_1$  and  $Q_2$ , and create a new edge with the width of  $t$  pixels ( $t > 0$ ). The effect of feather operation is shown in Fig.3.

We have made statistical analysis for the relationship between the new edge's width  $t$  and the feather radius  $r$ . The

relationship between  $r$  and  $t$  is listed in TABLE 1. It shows that the relationship of  $t$  and  $r$  approximately satisfies  $t = 5r$  in Fig.4. That means that the new edge after the feather operation will become wider to make a smooth transition in the feather region.



Fig.2. Example for the most common image manipulations to create an image forgery. (a) is the original Lena image, (b) is the photo of Nicole Mary Kidman. (c) is a composite image forgery by copying Nicole Mary Kidman's face to Lena image. (d) is another composite image forgery, in which the traces of splicing around the face are not clear due to the feather operation.

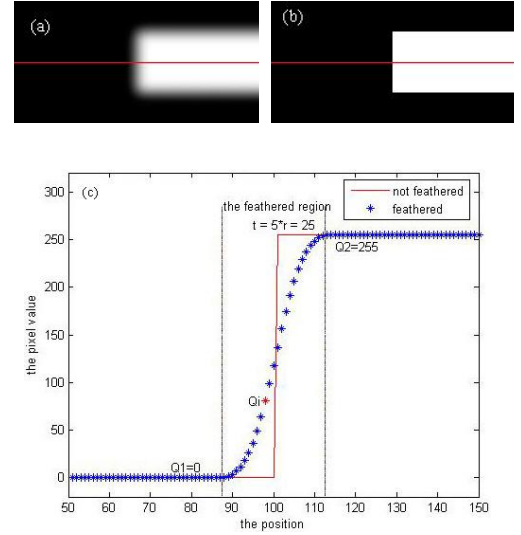


Fig.3. Feathered image (a) and un-feathered edge (b). Profile of the red line in (a) and (b) to show the difference between of the feathered edge and un-feathered edge (c).

TABLE 1. THE RELATIONSHIP BETWEEN  $t$  AND  $r$

$r$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$t$	2	2	2	2	4	4	4	6	6
$r$	2	3	4	5	6	7	8	9	10
$t$	10	14	20	24	32	36	40	46	50
$r$	11	12	13	14	15	16	17	18	19
$t$	56	62	66	70	76	80	86	90	96

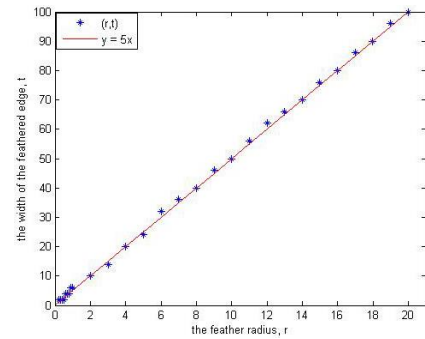


Fig.4. The relationship between  $t$  and  $r$ .

It suggests that each pixel  $Q_i$  in the feather region is acquired via the interpolation between the pixel  $Q_1$  of the background and the pixel  $Q_2$  of the pasted object through vast observation for the feather data. Thus the model of feather operation is described as follows:

$$Q_i = (1 - \frac{i}{5r})Q_1 + \frac{i}{5r}Q_2 \quad i \in \{0, 1, \dots, 5r\} \quad (2)$$

Then the new edge after the feather operation can be described by Eq.(3)

$$f'(x) = \begin{cases} Q_1, & x = A - 2.5r \\ (1 - \frac{x+2.5r}{5r})Q_1 + \frac{x+2.5r}{5r}Q_2, & A - 2.5r < x < A + 2.5r \\ Q_2 = Q_1 + H, & x = A + 2.5r \end{cases} \quad (3)$$

With the increasing of feather radius  $r$ , the new edge  $f'(x)$  with feather operation will become wider and smoother. It is obvious that the new edge  $f'(x)$  is a line function of  $x$  from Eq.(3), its slope is defined as the feather slope  $k$ , which describes the speed of smooth transition in the feather region. With the decreasing of  $k$ , the transition will become smoother, and vice versa.

$$k = \frac{Q_2 - Q_1}{5r} \quad (4)$$

Thus, the feather radius can be estimated by Eq.(5) :

$$r = \frac{Q_2 - Q_1}{5k} = \frac{H}{5k} \quad (5)$$

The accuracy of estimation of feather radius  $r$  is directly determined by the feather slope  $k$ . Then we will demonstrate how to measure the value of  $k$  in section 4.

#### IV. THE GRADIENT OF COLOR IMAGE

The derivative or the first difference of a linear function is a constant at every position. Based on the model of feather operation, the gradient in feather region will present smooth. It is an example that the influence of feather operation on the gradient phase angle is shown in Fig.5. As shown in (e) and (f), the gradient phase angle of the yellow flower's boundary present smooth due to the feather operation. Thus the smooth feature of the gradient phase angle can be used to expose the traces of feather operation.

Our interest is in computing the gradient in RGB color space. The way for RGB images would be to compute the gradient of each component color image and then combine the results. Unfortunately, in consideration of the dependence among the three channels, the gradient using this method is always undesirable. We define one pixel in a color image as a vector  $\mathbf{c}(x, y)$ , and compute the gradient by extending the concept of gradient from scalar function to vector function. Let

$\mathbf{r}, \mathbf{g}, \mathbf{b}$  be unit vectors along the R, G, B axis of RGB color space, and define the vectors

$$\mathbf{u} = \frac{\partial R}{\partial x} \mathbf{r} + \frac{\partial G}{\partial x} \mathbf{g} + \frac{\partial B}{\partial x} \mathbf{b} \quad (6)$$

and

$$\mathbf{v} = \frac{\partial R}{\partial y} \mathbf{r} + \frac{\partial G}{\partial y} \mathbf{g} + \frac{\partial B}{\partial y} \mathbf{b} \quad (7)$$

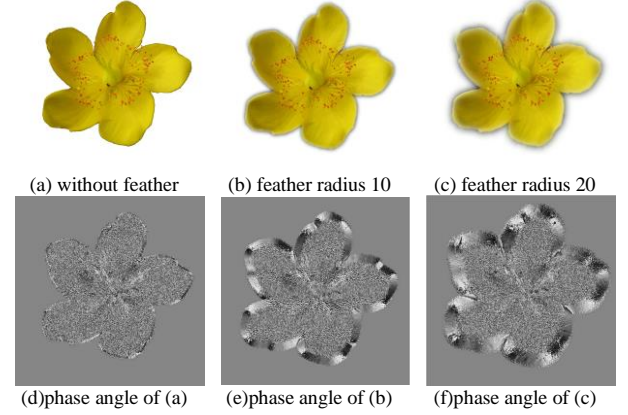


Fig.5. The influence of feather operation on the phase angle. Shown in the first row from left to right are the original image (a), the forgeries with feather radius  $r = 10$  (b) and  $r = 20$  (c). Shown in the second row are the gradient phase angle images respectively.

Let the quantities  $g_{xx}$ ,  $g_{yy}$ , and  $g_{zz}$  be defined in terms of the dot product of these vectors, as follows:

$$g_{xx} = \mathbf{u} \cdot \mathbf{u} = \mathbf{u}^T \mathbf{u} = \left| \frac{\partial R}{\partial x} \right|^2 + \left| \frac{\partial G}{\partial x} \right|^2 + \left| \frac{\partial B}{\partial x} \right|^2 \quad (8)$$

$$g_{yy} = \mathbf{v} \cdot \mathbf{v} = \mathbf{v}^T \mathbf{v} = \left| \frac{\partial R}{\partial y} \right|^2 + \left| \frac{\partial G}{\partial y} \right|^2 + \left| \frac{\partial B}{\partial y} \right|^2 \quad (9)$$

$$g_{xy} = \mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \frac{\partial R}{\partial x} \frac{\partial R}{\partial y} + \frac{\partial G}{\partial x} \frac{\partial G}{\partial y} + \frac{\partial B}{\partial x} \frac{\partial B}{\partial y} \quad (10)$$

$R, G, B$  and consequently the  $g$ 's, are functions of  $x$  and  $y$ . Using this notation, it can be shown [33] that the direction of maximum rate of change of  $\theta(x, y)$  as a function  $(x, y)$  is given by the angle



$$\theta(x, y) = \frac{1}{2} \arctan \left[ \frac{2g_{xy}}{g_{xx} - g_{yy}} \right] \quad (11)$$

and that the value of the rate of change (i.e., the magnitude of the gradient) in the directions given by the elements of  $\theta(x, y)$  is given by

$$F_\theta(x, y) = \left\{ \frac{1}{2} \left[ (g_{xx} + g_{yy}) + 2g_{xy} \sin 2\theta + (g_{xx} - g_{yy}) \cos 2\theta \right] \right\}^{\frac{1}{2}} \quad (12)$$

Note that  $\theta(x, y)$  and  $F_\theta(x, y)$  are images of the same size as the input image. The elements of  $\theta(x, y)$  are simply the angles at each point that the gradient is calculated, and  $F_\theta(x, y)$  is the gradient image.

## V. ESTIMATE THE FEATHER RADIUS

The feather radius is set by the forgery when the feather operation is used, and it is a constant in the same feather region. So we estimate the feather slope by using the least square method, and then compute the feather radius. Based on our model, the pixels in feather region will satisfy a line relationship. We define 16 directions around one pixel, and find the feather direction by detecting if the pixels along the direction can be best fitted by a line function. The feathered pixels along the feather direction can be used to estimate the feather slope.

We define a neighborhood  $\Omega$  size of  $2.5r$  and 16 directions  $D_j$  around the pixel  $q$  as shown in Fig.6.

$$D_j = j\pi/8, \quad j=0,1,\dots,15$$

There are  $5r+1$  pixels along direction  $D_j$ , and the pixel values are  $y_{-2.5r}, y_{-2.5r+1}, \dots, y_{2.5r-1}, y_{2.5r}$ . We assume that the relationship of the pixel values is well represented by a line function, as Eq.(13) shown, by using the least square method.

$$y = kx + b \quad x \in \{-2.5r, -2.5r+1, \dots, 0, \dots, 2.5r-1, 2.5r\} \quad (13)$$

Where,  $k$  is the slope,  $b$  is the intercept,  $y$  represent pixel values, and  $x$  is the position. The slope  $k$  is different from the direction  $D_j$ .  $D_j$  is a direction in a horizontal plane, and  $k$  is the slope of the line function. The slope  $k$  and the intercept  $b$  of the line function can be estimated by Eq.(14) and (15).

$$\hat{k}_{D_j} = \frac{(5r+1)(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{(5r+1)(\sum x_i^2) - (\sum x_i)^2} \quad (14)$$

$$\hat{b}_{D_j} = \frac{(\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{(2r+1)(\sum x_i^2) - (\sum x_i)^2} \quad (15)$$

Where  $\sum$  is equal to  $\sum_{i=-2.5r}^{2.5r}$ . The correlation coefficient  $R_{D_j}$  can be defined as followed.

$$R_{D_j} = \frac{\sum_{i=-2.5r}^{2.5r} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=-2.5r}^{2.5r} (x_i - \bar{x})^2} \sqrt{\sum_{i=-2.5r}^{2.5r} (y_i - \bar{y})^2}} \quad (16)$$

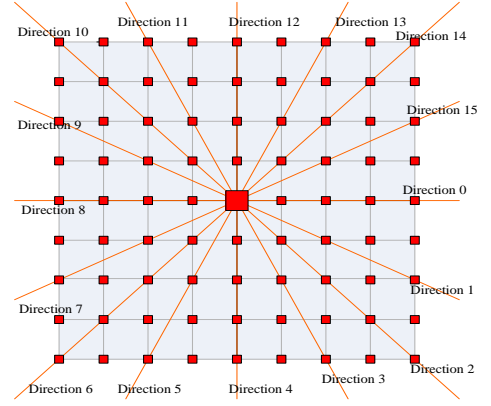


Fig.6. The 16 directions around the center pixel.

Because  $x \in \{-2.5r, -2.5r+1, \dots, 0, \dots, 2.5r-1, 2.5r\}$ , then  $\bar{x} = 0$ , the Eq.(16) can be rewritten as Eq.(17).

$$R_{D_j} = \frac{\sum_{i=-2.5r}^{2.5r} x_i (y_i - \bar{y})}{\sqrt{\sum_{i=-2.5r}^{2.5r} x_i^2} \sqrt{\sum_{i=-2.5r}^{2.5r} (y_i - \bar{y})^2}} \quad (17)$$

Where,

$$\bar{y} = \frac{1}{2W+1} \sum_{i=-2.5r}^{2.5r} y_i \quad (18)$$

The correlation coefficient has a range of  $R_{D_j} \in [-1, 1]$ . With the increasing of  $|R_{D_j}|$ , the linearity between  $x$  and  $y$  gets more consummate, and vice versa.

The standard deviation  $S_{D_j}$  can be computed in the following way.

$$S_{D_j} = \sqrt{\frac{1}{5r+1-2} \sum_{i=-2.5r}^{2.5r} [y_i - (\hat{k} x_i + \hat{b})]^2} \quad (19)$$

If  $S_{D_j} = 0$ , the straight line fits perfectly, namely, every pixel point lies on the straight line. With the increasing of  $S_{D_j}$ , the fit will get worse, because the deviations of points from the straight line become larger.

For all the directions  $D_j, j=0,1,\dots,15$ , we compute the slope  $\hat{k}_{D_j}$ , the correlation coefficient  $R_{D_j}$ , and the standard deviation  $S_{D_j}$  respectively. If the value of  $S_{D_j}$  is minimum, and  $S_{D_j} < \gamma_1$ ,  $|R_{D_j}| > \gamma_2$ , we consider the direction  $D_j$  as the feather direction, and the slope  $\hat{k}_{D_j}$  as the feather slope  $k$ .

$$k = \hat{k}_{D_j}, \text{ if } S_{D_j} < \gamma_1 \text{ and } |R_{D_j}| > \gamma_2 \quad (20)$$

Where  $\gamma_1$  and  $\gamma_2$  are two thresholds. The feather radius can be estimated by Eq.(5).

## VI. DETECTION OF THE FORGERIES

The presence of feather operation can be used as the evidence of tampering. So, in this section, we detect the traces of feather operation by using the smoothness of the gradient phase angle and uniformity of the feather radius. The degree of image credibility is defined to describe the reality of one image.

Let  $q(m,n)$  denote one pixel, and  $\Omega$  is its neighborhood. The most common neighborhood can be defined as shown in Fig.6.

$$q(m \pm i, n \pm j) \in \Omega \quad (i, j = 0, \dots, 5) \quad (21)$$

The similarity of pixel  $q$  among its neighbor pixels can be calculated in the following way.

$$s_q = \alpha \times r_q + \beta \times \theta_q \quad (22)$$

Where,  $\alpha$  and  $\beta$  are weight coefficients.  $r_q$  and  $\theta_q$  are defined as follows.

$$r_q = \frac{\sum_{q_i \in \Omega} C_r(q, q_i)}{N} \quad (23)$$

$$\theta_q = \frac{\sum_{q_i \in \Omega} C_\theta(q, q_i)}{N} \quad (24)$$

Where,  $N$  is the total number of pixels in the neighborhood  $\Omega$ .  $C_r(q, q_i)$  and  $C_\theta(q, q_i)$  are defined as follows.

$$C_r(q, q_i) = \begin{cases} 1, & \text{if } |r_q - r_{q_i}| \leq \lambda_r \\ 0, & \text{if } |r_q - r_{q_i}| > \lambda_r \end{cases} \quad (25)$$

$$C_\theta(q, q_i) = \begin{cases} 1, & \text{if } |\theta_q - \theta_{q_i}| \leq \lambda_\theta \\ 0, & \text{if } |\theta_q - \theta_{q_i}| > \lambda_\theta \end{cases} \quad (26)$$

Where,  $q_i$  is a neighbor pixel in the neighborhood  $\Omega$  of  $q$ ,  $r_q$  is a approximate caculation of feather radius,  $\theta_q$  is a phase angle of the gradient at  $q$ . In the neighborhood  $\Omega$ ,  $C_r(q, q_i)$  denotes number of pixels which are similar to feather radius at  $q$ , and  $C_\theta(q, q_i)$  denotes number of pixels which are similar to the phase angle of the gradient at  $q$ .

For a pixel  $q$ , if  $s_q \geq \delta_s$ , we mark  $q$  as the feathered pixel, if  $s_q \geq \delta_s - \mu$ , we mark  $q$  as the suspicious pixel, where  $\delta_s$  and  $\mu$  are thresholds.

Let  $N_f$  and  $N_s$  denote the number of feathered pixels and suspicious pixels. That is to say:

$$\begin{cases} \text{if } s_q \geq \delta_s & \text{then } N_f + 1 \\ \text{if } s_q \geq \delta_s - \mu & \text{then } N_s + 1 \end{cases} \quad (27)$$

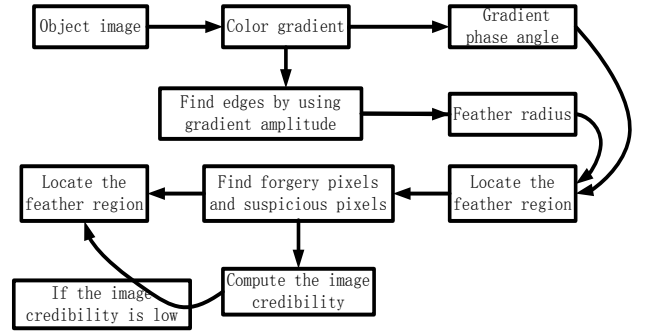


Fig.7 the flow chart of exposing image forgeries by detecting traces of feather operation.

The degree of image credibility can be defined as Eq.(28). We think it more acceptable to judge the reality of one image than just using a decision of YES and NO, which is usually used in many existing works.

$$D_{credibility} = 1 - \frac{N_f}{N_s} \quad (28)$$

The feathered pixels are marked with 255 and others with 0, and the result is presented by a binary image. We process the binary image using erosion and dilation operations to remove the isolated points and make the density points conjoined. If a connected region or a meaningful region can be structured by the white points, we regard it as the suspicious feather region.

The steps of the algorithm can be summarized as follows and the flow chart is shown in Fig.7.

- Find the image's edge by using color gradient amplitude  $F(\theta)$ , and get the gradient phase angle  $\theta$ .
- Estimate the feather radius  $r$  for every edge pixels.
- Estimate the similarity  $s_q$  for each edge pixel by using  $r_q$  and  $\theta_q$ , and find the feathered pixels.
- Locate the forgery region and compute the degree of image credibility.

## VII. RESULTS

### 7.1 Image database

In order to make experiment convenient, we have employed some students, who are unaware of our detection method, to build a database (more than 4000 images) for us. A set of image forgeries undergone the feather operation with various feather radius by using photo-shop have been chosen from our database to test the efficacy of our proposed algorithm. These images span a wide range of indoor and outdoor scenes. Most of them are captured in real scene by several digital cameras (including SONY DSC-T200, Olympus E20, Sony DSC-W220, Cannon 450D, Cannon D40 and Panasonic FS7GK), the rest a few are download from the internet according to the need. They are saved in the JPEG format at the same or different quality. Fig.8 shows some representative examples used in our experiment.

### 7.2 Experiment results

In our experiment, 120 images are chosen to test our algorithm. Tab.2 shows the successful rate, from which we can see that our algorithm is effective to detect the traces of feather operation without reference to the feather radius. The degree of image credibility for the test images are showed in Fig.9, from which we can see that the credibility of original image is usually higher than 85%, and the credibility of forgery is much lower.

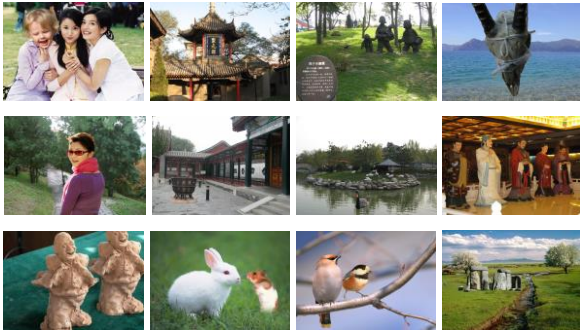


Fig.8. Some representative test images.

Several experiment results of successful detection are shown in Figure 10. Columns (a) and (b) are two original images. Shown in column (c) is an image forgery by copying part of image (b) and pasting to (a), and the artificial feather operation is used around the pasted region. We show the experimental result in column (d), and the traces of feather operation are marked by white points. From column (d) we can see that the pasted regions are marked by white points. The

degree of image credibility is computed for the nine images by using our method, and they are 91.94%, 89.09%, and 92.51% (top to bottom in column (a)), 89.78%, 90.32% and 91.11% (top to bottom in column (b)), 45.90%, 42.43%, 50.31% (top to bottom in column (c)). We can see that the degree of image credibility of forgery image (column (c)) is much lower.

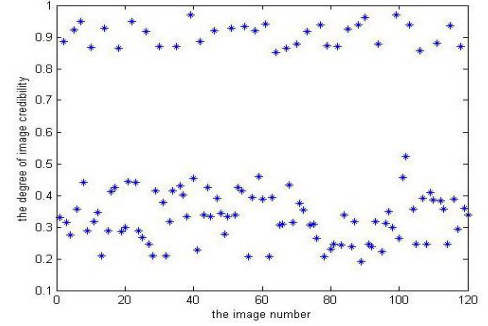


Fig.9. The degree of image credibility for the test images.

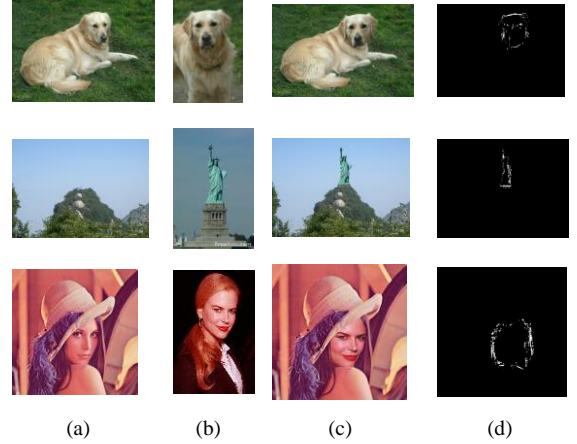


Fig.10. One of the experiment results. (a) and (b) are two original images, (c) is an image forgery, and the feather operation is used around the dog head (Statue of Liberty, or Nicole Mary Kidman' face). (d) is the detecting result using our technique, and the traces of feather operation are marked by white points.

TABLE 2. THE ACCURACY OF DETECTING FEATURE OPERATION

feather radius	number	Accuracy(resize factor)		
		1	0.5	2
No feather	30	29	27	28
$0.2 \leq r < 1$	30	27	26	28
$1 \leq r < 10$	30	28	28	29
$10 \leq r \leq 20$	30	30	30	30

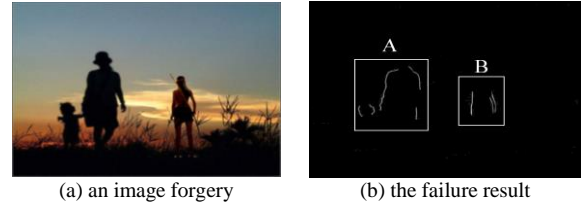


Fig.11. One failure example.

### 7.3 Discussion

We resize the 120 images (resize factor 0.5, 1, 2), and perform the same experiment on the resized images to find the

impact of resizing to our method. The accuracy rates are showed in Table 2, and it suggests that our method performs well even if the presence of resizing.

Although our proposed method produces encouraging results, more effort is still needed to improve the accuracy of our approach. It must be mentioned that the results obtained can be affected by the presence of other post-processing. Fig.11 shows a failure example, in which the right girl is feathered and moved from another image, and the hue adjustment is used for the whole forgery. As shown in B region, only pieces of segment are detected as the feather traces. Our method fails to detect the feather region, because the regularity brought by the feather operation is destroyed by hue adjustment. Also, although with very little possibility, a few original images may be detected as a forgery one, if the image was taken in a specific scene, or the edge of objects in the image was smooth extraordinary. As shown in Fig.11, the edge of the left girl is original but originally smooth, and some segments of its edge are mistaken for the traces of feather operation, shown in region A.

## VIII. CONCLUSION

Photo-shop has become the most attractive image-editing tool, and the feather operation has turned to be almost inevitable when tamping images. We propose a blind and efficient technique capable of exposing image forgeries by detecting the traces of feather operation. The degree of image credibility is defined to describe the reality of the image, which is more acceptable than just using a decision of YES or NO. The experimental results demonstrate that the proposed algorithm is reliable in discovering the traces of feather operation. However, when we don't use feather operation and other post-processing to tamper a image, the detection results would be unsatisfactory. Considerable more work, hopefully, will be done in this area. We expect that the technique described in this paper will lead to the development of image forensics filed, and make it increasingly harder to create convincing image forgeries.

## REFERENCES

- [1] V.M.Potdar, S.Han, E.Chang,2005. survey of digital image watermarking techniques, *IEEE International Conference on Industrial Informatics*, pp.709-716.
- [2] Tanzeela Qazi, 2013. Survey on blind image forgery detection, *IET Image Processing*, Volume 7, Issue 7, pp. 660-670.
- [3] Yanping Huang, Wei Lu, Wei Sun and Dongyang Long, 2011. Improved DCT-based detection of copy-move forgery in images, *Forensic Science International*, Vol.206, pp.178-184.
- [4] Jie Zhao, Jichang Guo, 2013. Passive forensics for copy-move image forgery using a method based on DCT and SVD, *Forensic Science International*, Vol 233, pp. 158-166.
- [5] Weihai Li, Nenghai Yu, and Yuan Yuan, 2008. Doctored JPEG image detection, *IEEE International Conference on Multimedia and Expo*, pp.253-256.
- [6] Qingzhong Liu, Peter A. Cooper, 2013. Detection of JPEG double compression and identification of smartphone image source and post-capture manipulation, *Springer-Applied Intelligence*Vol 39, pp.705-726.
- [7] Qu ZhenHua, Luo WeiQi, Huang JiWu, 2014. A framework for identifying shifted double JPEG compression artifacts with application to non-intrusive digital image forensics, *Science China Information Sciences*, vol.57, pp.028103:1-028103:18.
- [8] H. Farid. 2009.Exposing digital forgeries from JPEG ghosts, *IEEE Transactions on Information Forensics and Security*, vol.4, no.1, pp.154-160.
- [9] Hieu CuongNguyen,Stefan Katzenbeisser,2012.Performance and Robustness Analysis for Some Re-sampling Detection Techniques in Digital Images, *Springer-Verlag Berlin Heidelberg*, pp.387-397.
- [10] Gajanan K. Birajdar,Vijay H. Mankarr,2014. Blind method for rescaling detection and rescale factor estimation in digital images using periodic properties of interpolation, *Elsevier GmbH*.
- [11] Yongzhen Ke, Fan Qin, Weidong Min, Guiling Zhang, 2014.Exposing Image Forgery by Detecting Consistency of Shadow, *The Scientific World Journal*, Vol 2014.
- [12] M.K.Johnson and H.Farid,2008. Detecting photographic composites of people, *Lecture Notes in Computer Science*, Springer-Verlag Berlin Heidelberg, vol.5041, pp.19-33.
- [13] Y.Sutcu, B.Coskun, H.T.Sencar, and N.Memon,2007. Tamper detection based on regularity of wavelet transform coefficients, *IEEE International Conference on Image Processing*, vol.1, pp.397-400.
- [14] Makkena Purnachandra Rao et, 2014. Harnessing Motion Blur to Unveil Splicing, *IEEE Transactions on Information Forensics and Security*, vol. 9, pp.583-595.
- [15] Zhe Li and Jiangbin Zheng, 2009.Blind detection of digital forgery image based on the local entropy of the gradient, *Lecture Notes in Computer Science*, Springer-Verlag Berlin Heidelberg, vol.5450, pp.161-169.
- [16] Jiangbin Zheng and Miao Liu, 2009.A digital forgery image detection algorithm based on wavelet homomorphic filtering, *Lecture Notes in Computer Science*, Springer-Verlag Berlin Heidelberg, vol.5450, pp.152-160.
- [17] Chang-Hee Choi, Hae-Yeoun Lee, Heung-Kyu Lee, 2013. Estimation of color modification in digital images by CFA pattern change, *Forensic Science International*, vol 226, pp. 94-105.
- [18] A.Swaminathan, M.Wu, and K.J.R. Liu, 2008.Digital image forensics via intrinsic fingerprints, *IEEE Transactions on Information Forensics and Security*, vol.3, no.1, pp.101-117.
- [19] A.Swaminathan, M.Wu and K.J.R.Liu,2007. Non-intrusive component forensics of visual sensors using output images, *IEEE Transactions on Information Forensics and Security*, vol.2, no.1, pp.91-106.
- [20] M.Chen, J.Fridrich, J.Lukáš and M.Goljan, 2007.Imaging sensor noise as digital x-ray for revealing forgeries, *In Proc. 9<sup>th</sup> International Workshop on Information Hiding*, Saint Malo, France, *Lecture Notes in Computer Science*, vol.4567, pp.342-358.
- [21] H. Gou, A.Swaminathan, and M.Wu, 2007. Noise feathers for image tampering detection and steganalysis, *In Proc. IEEE International Conference on Image Processing*, San Antonio, vol.6, pp.97-100.
- [22] M.K. Johnson and H. Farid, 2006. Exposing digital forgeries through chromatic aberration, *in Proc. ACM Workshop on Multimedia and Security*, Geneva, Switzerland, pp.48-55.
- [23] Y.F.Hsu and S.F.Chang,2007. Image splicing detection using camera response function consistency and automatic segmentation. *In Proc. IEEE International Conference on Multimedia and Expo*, Beijing, China, pp.28-31.
- [24] Yu-Feng Hsu,2010. *Camera Response Functions for Image Forensics:An Automatic Algorithm for Splicing Detection* *Information Forensics and Security*,IEEE transaction on Signal Processing Society vol5, pp.816-825.
- [25] Wei Lu, Wei Sun,Fu-Lai Chung,Hongtao Lu, 2011. Revealing digital fakery using multiresolution decomposition and higher order statistics, *Engineering Applications of Artificial Intelligence*,vol 24,pp. 666-672.
- [26] S.Bayram, I.Avcibas, B.Sankur and N.Memon,2006. Image manipulation detection, *J. Electron. Imaging*, vol.15, no.4, pp.41102.
- [27] S.Bayram, İ.Avcıbaşı, B.Sankur, and N.Memon, 2005. Image manipulation detection with binary similarity measures, *In Proc. European Signal Processing Conf.*, Turkey.
- [28] S.Di Zenzo,1986. A note on the Gradient of a Multi-Image, *Computer Vision, Graphics and Image Processing*, vol.33, pp.116-125.

# Dual Graph Partitioning For Bottom-Up BVH Construction

Nathan Eloe  
nwe5g8@mst.edu

Joseph Steurer  
asjxc9@mst.edu

Jennifer L. Leopold  
leopoldj@mst.edu

Chaman L. Sabharwal  
chaman@mst.edu

Computer Science Department  
Missouri University of Science and Technology  
Rolla, MO 65409, USA

## Abstract

*Bounding Volume Hierarchies (BVHs) are essential tools in performing collision detection on three-dimensional information. They reduce the number of expensive calculations required to determine whether or not two geometrical entities collide by using inexpensive calculations to rule out parts of the objects that could not possibly intersect. Quickly producing a high quality BVH is an important aspect of three-dimensional multimedia analysis. As such a powerful optimization, efficient and high quality BVHs are still an active area of research. Herein, the authors present a novel BVH representation that reduces the redundancy in the tree structure by allowing a node to contain an arbitrary number of children, as well as compressing non-unique nodes and combining their children. A new partitioning scheme using a graphical representation of the object is also presented to improve the quality of the generated BVH.*

**Keywords** Bounding Volume Hierarchies, Dual Graph, Axis Aligned Bounding Box, Qualitative Spatial Reasoning

## 1 Introduction

The current trend toward big (multimedia) data analysis necessitates the ability to quickly process and analyze the vast amount of three dimensional information that is being generated. Collision detection between rendered static objects is a computationally complex process; even performing a ray-casting collision check can be a time-consuming process if steps are not taken to optimize the number of calculations. A commonly used mechanism to significantly optimize the computation time is a Bounding Volume Hierarchy (BVH) [1]. A BVH subdivides a portion of space into smaller volumes containing objects of interest. Each sub-volume is then adaptively divided until some atomic level is reached. The BVH is used to efficiently pare out parts of a

volume space that could not possibly contribute to the intersection query, resulting in the removal of a large number of potentially expensive calculations.

A BVH is commonly created in one of three different ways: Top-Down, Bottom-Up, or Iterative Insertion. Each of these generation mechanisms has strengths and weaknesses relating to the creation time and the quality of the resulting BVH [2]. One BVH is considered to be of higher quality than another BVH if collision queries can be performed faster. As such, a higher quality BVH tends to minimize the total volume contained in the bounding volumes and be as compact as possible. This is roughly analogous to the *Surface Area Heuristic* (SAH), a mechanism used to determine the expected cost to perform a ray trace [3], though recently additional quality metrics on Bounding Volume Heuristics have been suggested [4]. Because collision detection is frequently used in determining the spatial relation between objects in qualitative spatial reasoning systems such as VRCC-3D+ [5], the implementation of a BVH subsequently can play a pivotal role in a spatial reasoner's information feedback loop and the user's overall experience with an application.

Herein the authors investigate the use of a dual graph representation of a three-dimensional object with triangulated surface boundary to improve on the quality of a BVH generated by using a bottom-up approach. This approach is not designed to replace existing BVH creation algorithms; it should be considered as an enhancement that can be integrated into already existing algorithms that use other optimizations. An implementation of a BVH using Axis Aligned Bounding Boxes (AABB) as the bounding volume and a representation that removes internal redundancy is presented here for use in benchmarking the dual graph partitioning scheme.

The remainder of this paper is organized as follows. Section 2 reviews Bounding Volume Hierarchies and Dual Graphs. Section 3 introduces the internal structure of the AABB Tree used in the implementation of VRCC-3D+ that reduces redundancy and number of nodes. The Dual Graph

partitioning scheme is presented in Section 4. Section 5 outlines the experimental setup; results from these experiments are presented in Section 6. Conclusions drawn from the results are presented in Section 7. Section 8 outlines the future work that will be undertaken.

## 2 Background

### 2.1 Bounding Volume Hierarchies

BVHs are an efficient method for quickly handling ray intersection and collision detection in a three dimensional environment. BVHs can reduce the computation time of ray/collision detection logarithmically, as a child node in the hierarchy doesn't need to be calculated if a parent node is not in the intersection.

There are three kinds of commonly employed methods: Top-Down, Bottom-Up, and Iterative Insertion. Top-down is the most common approach as it is faster than naïve collision detection, easy to implement, and efficient to create [6, 2]. If performed correctly, a bottom-up tree generation is likely to produce a higher quality tree than a top-down approach [7]. Insertion methods are commonly used when not all objects in a scene are visible, as they allow objects to be added dynamically [8]. These approaches are based on the assumption that geometric primitives are predetermined. Trees are sometimes dynamically created based on the complexity and intersection outcome of the objects. In the surface-surface intersection, none of these methods is sufficient [9].

The Top-Down method works by wrapping a scene or an object in a bounding volume, typically the root of a tree. This bounding volume is subdivided into smaller volumes, with these volumes becoming the children of the root of the tree. This process is recursively performed until some stopping criterion is met. Common stopping criteria are minimum bounding volume, minimum number of primitives contained in a tree node, or maximum depth of the tree [10].

The Bottom-Up creation algorithm begins with the primitives and groups them, creating a bounding volume around each grouping. This process is repeated, treating each bounding volume as a new primitive. This process minimizes the unnecessary space contained within the internal nodes, and frequently produces a higher quality tree than a top-down method provided all the primitives are predetermined [11].

The third method, Iterative Insertion, begins with an empty tree, and inserts objects into the tree as they become visible. This allows for a dynamic scene to be processed using a BVH [10].

Herein the authors focus on a modification of the Bottom-Up approach. By using a dual graph (see Section 2.2) to represent a three-dimensional object, primitives

can be grouped such that they are spatially close to each other, resulting in a more optimal tree than a naïve grouping. A novel representation of a BVH that reduces the internal size of the tree by removing redundancy is used as a benchmark; a more thorough examination of the implementation is presented in Section 3.

### 2.2 Dual Graphs

In graph theory, the dual graph of graph  $G = (V, E)$ , is denoted as  $G^*$ . In  $G^* = (V^*, E^*)$ , a vertex  $v^* \in V^*$  represents a face in  $G$ . An edge  $e^* \in E^*$  exists between any two vertices  $v_1^*$  and  $v_2^*$  if the faces they represent in  $G$  share at least one edge [12].

A common representation of three-dimensional objects in CAD/CAM software is the ANSI B-rep model utilizing the triangulation of the boundary surface. The object boundary is specified as a mesh of vertices in space, with edges connecting those points, and triangular faces enclosed by the edges. This is directly analogous to a graphical representation of the object. By considering every vertex as a node in the graph, and edges between vertices as edges in the graph, the mesh representation of the object boundary becomes a graph. The triangles that comprise the approximation of the surface are called the faces of the object.

A dual graph directly follows; by allowing each triangular face in the object to be represented as a vertex in  $G^*$ , edges are defined in  $G^*$  as any two faces that share exactly two vertices  $v_i, v_j \in V$  of  $G(V, E)$ . The dual graph allows visualization of how faces are arranged, and insights into the structure of the object can be gleaned from a simple breadth-first or depth-first search of the graph. In using the geometric Dual Graph to partition the faces of an object, the breadth-first search is used to form a spanning tree of half of the faces so as to keep the groupings as compact as possible.

## 3 BVH in VRCC-3D+

VRCC-3D+ [13, 5, 14] is the implementation of a region connection calculus that qualitatively determines the spatial relations between three dimensional objects, both in terms of connectivity and obscuration. The VRCC-3D+ connectivity relations are calculated in three dimensions, and include: disconnected (DC), externally connected (EC), partial overlap (PO), equality (EQ), tangential proper part (TPP), non-tangential proper part (NTPP), tangential proper part converse (TPPc), and non-tangential proper part converse (NTPPc); see Fig. 1. A composite VRCC-3D+ relation specifies both a connectivity relation and an obscuration relation. Obscuration is considered from a 2D projection and a depth relation. There are fifteen different obscuration relations defined in VRCC-3D+; each is a refinement of basic concepts: no obscuration, partial obscuration, and



complete obscuration. For a more complete discussion of VRCC-3D+, see [13, 5, 14].

The implementation of VRCC-3D+ utilizes a large number of collision detection determinations to calculate the connectivity relationships between objects, and ray-casting techniques to determine how objects obscure one another from a given vantage point. If the size or location of an object in a scene changes, the object's spatial relationships with other objects in the scene may change. In turn, each recalculation of a VRCC-3D+ relation requires creation and/or queries of BVHs.

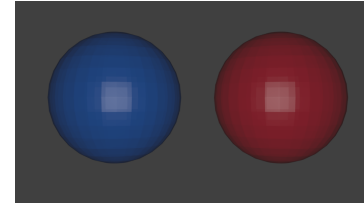
The VRCC-3D+ BVH implementation uses Axis Aligned Bounding Boxes (AABBs) as the bounding volume. The creation of an AABB is efficient, requiring only the minimum and maximum coordinate in each dimension. This operation is linear in the number of points that are being bounded if there is no known ordering of the points. Intersection tests involving AABBs are simple in their logic and execution for several types of primitives, including line segments, rays, and other AABBs. For these reasons, libraries such as CGAL [15] have AABB trees implemented as part of the standard set of tools used in computational geometry.

As with any BVH, the AABB has some drawbacks. There are many cases where the AABB might not be an optimal (tight fitting) bounding volume. Other bounding volumes, like Oriented Bounding Boxes [16] (OBBs) may give tighter bounds. However, their creation is more complex. Other bounding volumes, such as Bounding Spheres are easier to test for intersection with. Bounding spheres are easy to create, but the creation process is an approximation; it is unlikely that a Bounding Sphere will be tightly fitting around the points it encloses.

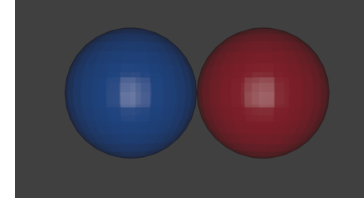
For the initial implementation, it was decided that the AABB was the best way to proceed. The simple creation of the bounding volume and the ease of representing the intersection calculations allowed for fast implementation and testing. The BVH in VRCC-3D+ is designed such that a different bounding volume can be used without significant effort. Also, because the boxes are aligned along the global axes, multiple groupings of points are more likely to have identical bounding boxes. This property is exploited in the representation introduced below.

The BVH used in VRCC-3D+ is in the worst case a binary tree. Algorithm 1 shows the tree creation pseudocode; the nodes are partitioned into two sets, and then trees are created from the partitions. When implemented as a tree, the result is a binary tree. However, in the initial implementation, several redundancies were noticed: the same bounding box appeared multiple times as an internal node.

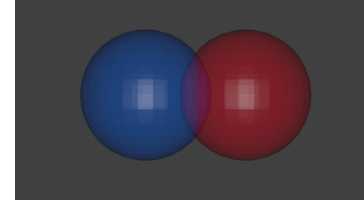
As such, the implementation was modified in a way inspired by the work presented in [17]. The underlying tree structure was changed to be a dictionary (or hash map), in



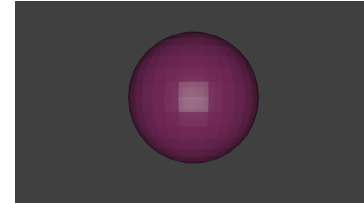
(a) DC(red, blue)



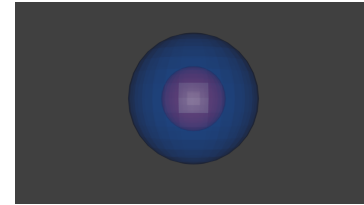
(b) EC(red, blue)



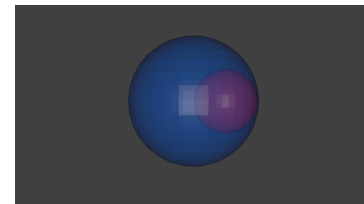
(c) PO(red, blue)



(d) EQ(red, blue)



(e) NTPP(red, blue) = NTPPc(blue, red)



(f) TPP(red, blue) = TPPc(blue, red)

Figure 1: The VRCC-3D+ Connectivity Relations. The centers of all spheres lie in the same YZ plane; only the X coordinate changes.

which the key-value pairs represented a node and a list of its children. Instead of setting a node's left and right child, the list of children is extended. This implementation removed redundancy from internal nodes, reducing the height of the tree and preventing repeated queries against the same bounding box. Fig. 2 shows a scenario in which a bounding box appears multiple times in the tree. The bounding box for triangle b completely contained the box for triangle d, and as such the bottom-up method for tree construction created a parent node for both box B and box D that was identical to box B. This redundancy initially expressed itself as a bug where nodes were deleted from the tree as it was being created, causing inconsistencies and incorrect query results.

This tree representation lends itself to a clean implementation. Algorithm 1 shows the pseudocode for generating the BVH. The pseudocode (using some Python notation) assumes the following:

- `root` is the root of the tree object being generated.
- `tree` is the internal representation of the tree and is a hash map, in which the key is an AABB, and the value is a list of either AABBs or triangular faces.
- The AABB constructor takes a list of faces or boxes and generates an AABB containing all objects passed to it.
- `faces` is a set (only unique faces).

---

**Algorithm 1** VRCC-3D+ Tree Creation: Bottom Up

---

```

function MAKE TREE(faces)
  if faces.length=1 then
    face ← faces.pop()
    root ← AABB(face)
    tree[root] ← face
    return root
  end if
  (left, right) ← PARTITION(faces)
  lbox ← MAKE TREE(left)
  rbox ← MAKE TREE(right)
  root ← AABB(lbox, rbox)
  tree[root].append(lbox)
  tree[root].append(rbox)
  return root
end function

```

---

## 4 Dual Graph Partitioning (DGP)

In a three-dimensional object represented by the ANSI boundary representation, every face contains three vertices connected by three edges. This is directly analogous to a

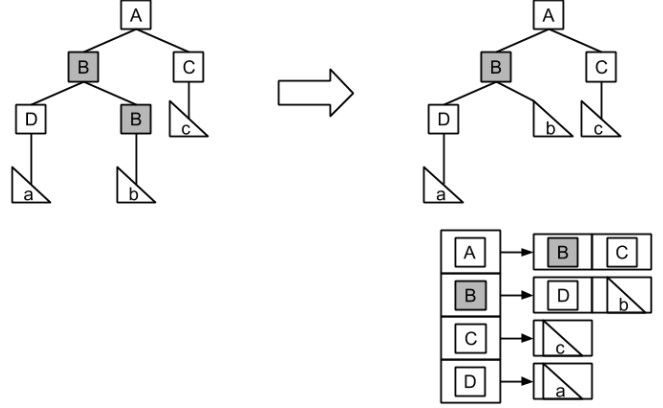


Figure 2: Redundancy in a BVH. In a bottom up approach, AABB B is a box that contains face b, but also fully encloses bounding box D. As such, it appeared multiple times in the tree, introducing redundancy and unnecessary calculations.

graph  $G = (V, E)$ , in which the set of vertices in the graph is the vertices of the triangular faces in 3-space, and an edge in the graph is a connection between two vertices in the facial boundary representation of the object. The dual graph  $G^*$  can be constructed as a graph in which each face is represented by a vertex, and an edge represents two faces that share two vertices in  $V$ ; connectivity in  $G^*$  guarantees that two faces are spatially close.

The dual graph can be used to partition the faces into groups of faces that are spatially close to each other. By choosing an arbitrary starting face in a set of faces, a breadth-first search to create a spanning tree that contains half of the vertices in the graph ensures that all faces in the partition are connected in the dual graph, and as such are spatially close relative to each other.

It is known in this implementation that the dual graph will be a sparse graph; every vertex has exactly three neighbors because every vertex represents a triangular face. As a sparse graph, it is efficient to represent it as an adjacency list. Algorithm 2 shows how the dual graph partitioning generates two partitions of faces.

## 5 Experimental Design

The purpose of this experiment is to determine the effect that the dual graph partitioning scheme has on the runtime and quality of the tree. An algorithm in which *partition()* was implemented to split the set of faces in half by index was used as a control. It was then modified to accept and use the adjacency list of the dual graph to determine how the faces should be split when recursively creating the tree. This implementation has some interesting features that

---

**Algorithm 2** Dual Graph Partitioning

---

```

function DUAL PARTITION(faces, adj)
  right  $\leftarrow$  set()
  currFace  $\leftarrow$  faces.pop()
  nextF  $\leftarrow$  queue()
  right  $\leftarrow$  {currFace}
  goal  $\leftarrow$  faces.length/2
  while right.length < goal do
    adjFaces  $\leftarrow$  faces  $\cap$  adj[currFace]
    adjFaces  $\leftarrow$  adjFaces - right
    numToAdd  $\leftarrow$  goal - right.length
    if numToAdd > 3 then
      numToAdd  $\leftarrow$  3
    end if
    toPush  $\leftarrow$  set(adjFaces[0 : numToAdd])
    right  $\leftarrow$  right  $\cup$  toPush
    faces  $\leftarrow$  faces - toPush
    ENQUEUE(nextF, toPush)
    currFace  $\leftarrow$  DEQUEUE(nextF)
  end while
  return faces, right
end function

```

---

evolved as the VRCC-3D+ implementation was growing, specifically that the bounding hierarchy is represented as a dictionary/hash map, and each internal node has a set that contains the children. This has an interesting side-effect: while the tree is constructed as a binary tree, if two grouped faces have the same bounding box, then the internal node in the tree will expand and have more than two children. This directly affects the size and average height of the tree, making those aspects of the tree smaller.

All timing was performed on an Intel Core i5 processor, running at 3.2GHz, with 16GB of RAM. All implementations are written in Python, using the *timeit* module to collect runtime information. Timing does not include the creation of the adjacency list unless otherwise specified; this is a one time cost that can be amortized over every tree creation.

## 5.1 Configuration 1

For the purposes of testing VRCC-3D+, a collection of 68 .obj files was created, each file depicting one of the VRCC-3D+ relations between two objects. Each of these files portrays two polyhedrons ranging from tetrahedrons (four triangular faces) to spheres (with 2000 triangular faces per polyhedron). These files were used for testing each of the BVH creation implementations.

The BVH creation time was averaged over 100 executions on each of the 68 sample files using both partitioning schemes. The number of internal nodes and average box

volume was collected when using each partitioning scheme for every file.

## 5.2 Configuration 2

For a BVH, the act of determining which primitives a geometry can intersect is called querying the tree. Determining which faces from an object could intersect with faces from another object is called a box-box query (because boxes from one BVH are queried against boxes in the second).

Every .obj file contains two discrete objects. The box-box query time between the two objects was averaged over 100 runs for the same file set as in Configuration 1 for both partitioning schemes by implementing the algorithm presented in Algorithm 3.

## 5.3 Configuration 3

The need to create the adjacency list is an additional overhead that could impact performance. However, because the faces of the objects should never change their position relative to each other, the adjacency list can be precalculated and stored. This has a known memory cost (linear in the number of faces in the object). If memory is a problem, this list could be generated every time the BVH is generated. In this configuration, the time to create the adjacency list is averaged over 100 runs on the same set of files as in Configuration 1.

## 5.4 Configuration 4

To test scalability, each algorithm was executed 100 times on a file with two objects with significantly larger numbers of faces (9144 and 14624). The same statistics as in Configuration 1 were collected.

# 6 Results

All times in these results are reported in seconds. All volume measurements are presented as cubed world units. Box and whisker plots are used to report quartile values over the input file set.

## 6.1 Configuration 1

Table 1 shows that, on average, the time to create a tree using DGP is about 15% longer than using a naïve partitioning scheme. Fig. 3 shows that for smaller objects, in the best case, DGP can run as quickly as a naïve method.

The average box volume is also shown in Table 1 and Figure 5; note that the volume is, on average 5% smaller

---

**Algorithm 3** Box-Box Query

---

**function** BOX BOX QUERY(*tree1*, *node1*, *tree2*, *node2*)*isect*  $\leftarrow$  dict() $\triangleright$  Dictionary of (node, list of nodes) pairs**if** INTERSECT(*node1*, *node2*) **then**    **for** *n1*  $\in$  *tree1*[*node1*] **do**        **for** *n2*  $\in$  *tree2*[*node2*] **do**            **if** ISAABB(*n1*) and ISAABB(*n2*) **then**                *isect*  $\leftarrow$  *isect* + BOX BOX QUERY(*tree1*, *n1*, *tree2*, *n2*)            **else if** ISAABB(*n1*) **then**                *isect*  $\leftarrow$  *isect* + BOX BOX QUERY(*tree1*, *n1*, *tree2*, *node2*)            **else if** ISAABB(*n2*) **then**                *isect*  $\leftarrow$  *isect* + BOX BOX QUERY(*tree1*, *node1*, *tree2*, *n2*)            **else**                *isect*[*node1*].append(*node2*)            **end if**        **end for**    **end for****end if****return** *isect***end function**

---

when the tree is generated using DGP. Table 1 and Fig. 4 reflect the size of the tree through the number of internal nodes. A BVH created using DGP is, on average, slightly smaller than one created with a naïve partitioning method. In some cases, an increase is seen, but the average and median volumes are smaller when using DGP.

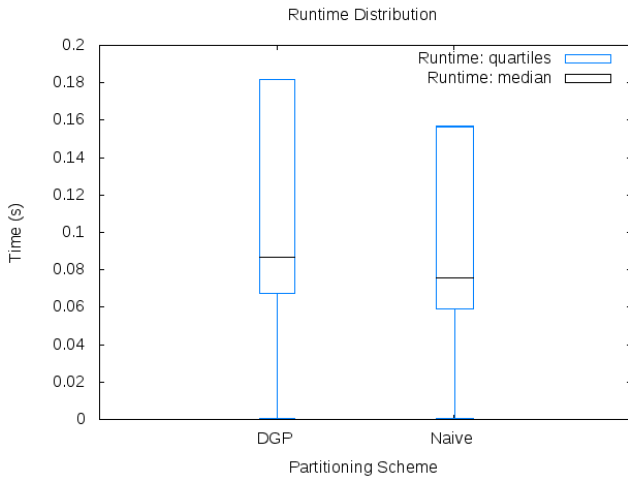


Figure 3: Range of Runtime for DGP and Naïve Partitioning

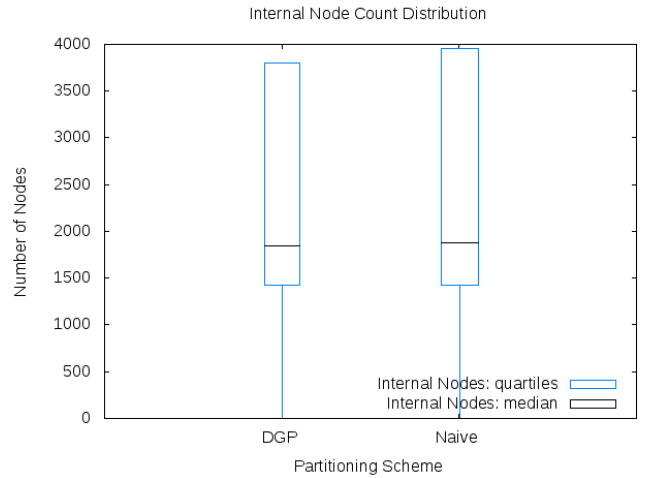


Figure 4: Range of Number of Nodes for DGP and Naïve Partitioning

## 6.2 Configuration 2

Table 2 shows the average execution time of a box-box query between two objects. Again, DGP shows an improvement over the naïve partitioning process.

Table 1: Average and Standard Deviation: Configuration 1

	$\mu$ : runtime	$\sigma$ : runtime	$\mu$ : box volume	$\sigma$ : box volume	$\mu$ : node count	$\sigma$ : node count
DGP	0.105168340627	0.0714088494841	1.8130619326	3.24021971564	2202.60294118	1491.67271031
Naïve	0.090840193454	0.0613155721599	1.96238023672	3.16972935732	2275.45588235	1556.73862926

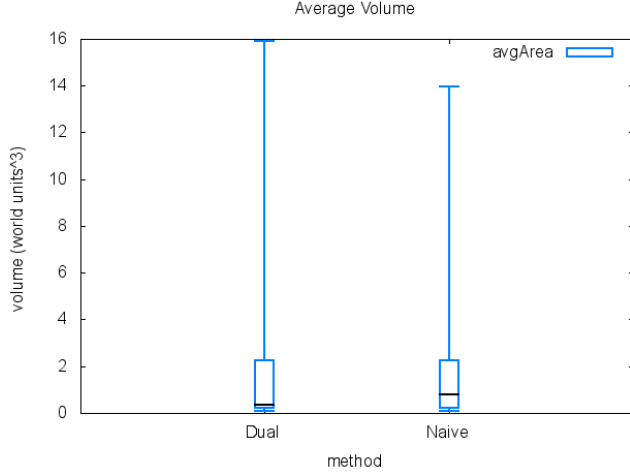


Figure 5: Range of Box Volume for DGP and Naïve Partitioning

Table 2: Average and Standard Deviation of Query Runtime Over Test Files

	$\mu$ : runtime	$\sigma$ : runtime
DGP	0.528159871522	0.574440745841
Naïve	0.641162978411	0.51794431845

### 6.3 Configuration 3

Table 3 shows that the overhead of creating the list is small; it is less than half of the time required to generate the tree. If memory is a limitation, the adjacency list can be generated at tree creation, though the runtime will be half again as long.

Table 3: Average and Standard Deviation of Adjacency List Creation Over Test Files

Average	Standard Deviation
0.0472587228873	0.0422061185621

### 6.4 Configuration 4

The average creation time, number of nodes, and box volume taken from trees created on significantly larger objects are shown in Table 4. Note the mixed results; for one object, DGP generates fewer boxes but has a higher box volume, while for the other object, DGP results in more boxes with a smaller average box volume.

## 7 Conclusions

### 7.1 Configurations 1 and 2

The results from the first and second testing configurations are encouraging for the DGP scheme. Using DGP introduces a slight increase in the average BVH creation time (approximately 0.014 seconds, a 15% increase on the tested input object files). However, the query time to determine whether two objects could possibly intersect by performing a box-box query decreases by more than .11 seconds; this is almost an 18% improvement in runtime. The gain in query performance is almost eight times the slowdown introduced in the BVH creation.

To put this in perspective, consider a scene with  $n$  objects, requiring exactly  $n$  AABB creations. In the worst case, the VRCC-3D+ implementation would perform  $\binom{n}{2}$  box-box queries ( $O(n^2)$ ) to exactly determine the three-dimensional spatial relationships. Determining the obscuration relation requires additional queries to the tree; one of the predicates used to determine the obscuration is whether an object is in front of another.

In determining the obscuration between two objects, the implementation of VRCC-3D+ projects the three dimensional objects onto the view plane. If the two projections overlap, a number of evenly distributed points is chosen from the overlap. A ray starting at the camera and passing through each of these points is queried against the BVH to determine which object is passed through first. The precision and correctness of this operation increases as the number of points distributed across the overlapping projection increases. If  $p$  points are chosen from the intersection of the projection, then the number of ray queries made against the tree is  $O(p)$ .

It can be seen that the number of queries made to the tree quickly surpasses the number of tree creations:  $O(n) <$

Table 4: Creation Time, Tree Size, and Box Volume on Larger Objects

	Faces	Average Creation Time	Average Number of Nodes	Average Box Volume
DGP	9144	0.8753760600090027	15768	271249.03166667151
Naïve	9144	0.7410803198814392	15712	290342.60701300012
DGP	14642	1.4204304003715515	25688	285103.56916561484
Naïve	14642	1.193346061706543	25863	265531.88658817636

$O(n^2) + O(p)$ . As such, the time savings are significant. Not even the time overhead to create the adjacency list on the fly is enough to negate the benefit of using DGP over a naïve partitioning mechanism.

The box volume tells an interesting story. The average and median box volumes are smaller when using DGP. However, on occasion, DGP produces a larger total bounding volume than the naïve partitioning. While using the spanning tree of the dual graph ensures that the bounding volumes are spatially close, the starting point used in generating the tree can have an effect on the overall shape of the enclosing bounding volumes.

## 7.2 Configuration 3

As shown in the testing results from Configuration 3, the time it takes to generate the adjacency list is smaller than the time of tree generation. In cases where there is not enough memory to create and store the adjacency list for all objects on initialization, the list may be generated as the tree is being built. The additional runtime does not cause the increase in the creation time to overshadow the benefits to the query time. Indeed, the total time to create the adjacency list and generate the tree is smaller than the time savings in a single query.

## 7.3 Configuration 4

The results from testing Configuration 4 show that the DGP scales as well as a naïve partitioning scheme. The increase in runtime remains a constant factor ( $\sim 18\%$ ). The trees generated show some inconsistencies; for one of the objects, the tree created using DGP has more nodes than the one created with the naïve partitioning method, but has a significantly smaller average box volume. For the other, larger object, this relationship is flipped: DGP results in a larger average box volume, but a smaller number of nodes. This suggests that for any partitioning, the shape of the object affects the results. Other work has been done in which the primitives are partitioned along the longest axis of the overall bounding box [7]; DGP would perhaps benefit here from a different starting point for the breadth-first search.

As a field, computer science is about problem solving with an emphasis on efficiency and elegance. A given solution may not work well on all representations of a problem; multiple approaches to a solution are beneficial as it allows flexibility in the application of the solution to different problem representations. Computer graphics and multimedia are very broad subjects: objects can be represented in a myriad of ways, making a single approach to analysis and computation tools nearly impossible. Bounding Volume Hierarchies are powerful tools; they allow for efficient collision queries by exploiting the logarithmic complexity of tree structures to quickly prune expensive calculations. Dual Graph Partitioning is an approach to generating these tree structures that can be applied to representations of scenes that do not lend themselves well to other tree creation algorithms. It is another tool in the large toolbox of powerful multimedia analysis mechanisms available to the modern computer scientist.

## 8 Future Work

Improving the quality and efficiency of the BVH implementation in VRCC-3D+ is an ongoing process. Optimizing the improvements gained by introducing DGP and the hash-map based tree representation will continue with the integration of other bottom-up tree improvements, such as those suggested in [7].

Areas for improvement include analyzing the effect that different breadth-first search starting points has on the quality of the resulting BVH and determining the feasibility and practicality of using threading to build pieces of the tree in parallel.

## 9 Acknowledgments

Thanks to the students in the CS 390 Undergraduate Research course at Missouri S&T: Spencer Cramm, Emerson Lentz, and Emilio Navarro.



## References

- [1] I. Wald, S. Boulos, and P. Shirley, “Ray Tracing Deformable Scenes using Dynamic Bounding Volume Hierarchies,” *ACM Transactions on Graphics*, vol. 26, no. 1, 2007.
- [2] C. Lauterbach, M. Garland, S. Sengupta, D. Luebke, and D. Manocha, “Fast BVH construction on GPUs,” *Computer Graphics Forum*, vol. 28, no. 2, pp. 375–384, 2009.
- [3] J. D. MacDonald and K. S. Booth, “Heuristics for Ray Tracing Using Space Subdivision,” *The Visual Computer*, vol. 6, no. 3, pp. 153–166, 1990.
- [4] T. Aila, T. Karras, and S. Laine, “On Quality Metrics of Bounding Volume Hierarchies,” in *Proceedings of the 5th High-Performance Graphics Conference*. ACM, 2013, pp. 101–107.
- [5] C. L. Sabharwal, J. L. Leopold, and N. Elloe, “A More Expressive 3D Region Connection Calculus,” in *Proceedings of the 2011 International Workshop on Visual Languages and Computing (in conjunction with the 17th International Conference on Distributed Multimedia Systems (DMS 11))*, Aug. 2011, pp. 307–311.
- [6] I. Wald, W. R. Mark, J. Günther, S. Boulos, T. Ize, W. Hunt, S. G. Parker, and P. Shirley, “State of the Art in Ray Tracing Animated Scenes,” in *Eurographics 2007 State of the Art Reports*, 2007.
- [7] Y. Gu, H. Yong, J. Fatahalian, and G. Bluelloch, “Efficient BVH Construction via Approximate Agglomerative Clustering,” in *Proceedings of the Fifth High Performance Graphics 2013 (HPG ’13)*, 2013, pp. 81–88.
- [8] J. Haber, M. Stamminger, and H.-P. Seidel, “Enhanced automatic creation of multi-purpose object hierarchies,” in *Proceedings of the Eighth Pacific Conference on Computer Graphics and Applications*, 2000. IEEE, 2000, pp. 52–437.
- [9] E. G. Houghton, R. F. Emmett, J. D. Factor, and C. L. Sabharwal, “Implementation of a Divide-and-Conquer Method for Intersection of Parametric Surfaces,” *Computer Aided Geometric Design*, vol. 2, no. 1, pp. 173–183, 1985.
- [10] T. Larsson, “Adaptive Bounding Volume Hierarchies for Efficient Collision Queries,” Ph.D. dissertation, Mälardalen University, 2009.
- [11] B. Walter, K. Bala, M. Kulkarni, and K. Pingali, “Fast Agglomerative Clustering for Rendering,” in *IEEE Symposium on Interactive Ray Tracing, 2008 (RT 2008)*. IEEE, 2008, pp. 81–86.
- [12] D. B. West, *Introduction to Graph Theory*, 2nd ed. Prentice Hall, 2001.
- [13] J. Albath, J. L. Leopold, C. L. Sabharwal, and A. M. Maglia, “RCC-3D: Qualitative Spatial Reasoning in 3D,” in *Proceedings of the 23rd International Conference on Computer Applications in Industry and Engineering*, Nov. 2010, pp. 74–79.
- [14] C. L. Sabharwal and J. L. Leopold, “Smooth Transition Neighborhood Graphs For 3D Spatial Relations,” in *IEEE Symposium Series Workshop on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP)*, 2013, pp. 8–15.
- [15] “CGAL, Computational Geometry Algorithms Library,” <http://www.cgal.org>.
- [16] S. Gottschalk, M. C. Lin, and D. Manocha, “OBB-Tree: A Hierarchical Structure for Rapid Interference Detection,” in *23rd annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 1996, pp. 171–180.
- [17] T. Kim, B. Moon, D. Kim, and S. Yoon, “Random-Accessible Compressed Bounding Volume Hierarchies,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 2, pp. 273–286, 2010.

# Path Planning Directed Motion Control of Virtual Humans in Complex Environments

Song Song, Weibin Liu,  
Ruxiang Wei

Institute of Information Science  
Beijing Jiaotong University  
Beijing Key Laboratory of Advanced  
Information Science and Network  
Technology  
Beijing 100044, China  
e-mail: wblu@bjtu.edu.cn

Weiwei Xing

School of Software Engineering  
Beijing Jiaotong University  
Beijing 100044, China

Cheng Ren

Safety Equipment Research Institute  
China Coal Research Institute  
Beijing 100013, China

**Abstract**—Natural motion synthesis of virtual humans have been studied extensively, however, motion control of virtual characters actively responding to complex dynamic environments is still a challenging task in computer animation. It is a labor and cost intensive animator-driven work to create realistic human motions of character animations in a dynamically varying environment in movies, television and video games. To solve this problem, in this paper we propose a novel approach of motion synthesis that applies the optimal path planning to direct motion synthesis for generating realistic character motions in response to complex dynamic environment. In our framework, SIPP (Safe Interval Path Planning) search is implemented to plan a globally optimal path in complex dynamic environments. Three types of control anchors to motion synthesis are for the first time defined and extracted on the obtained planning path, including turning anchors, height anchors and time anchors. Directed by these control anchors, highly interactive motions of virtual character are synthesized by motion field which produces a wide variety of natural motions and has high control agility to handle complex dynamic environments. Experimental results have proven that our framework is capable of synthesizing motions of virtual humans naturally adapted to the complex dynamic environments which guarantee both the optimal path and the realistic motion simultaneously.

**Keywords:** *motion synthesis; motion control; virtual human; path planning; complex dynamic environments*

## I. INTRODUCTION

Animating and controlling virtual humans realistically in complex dynamic environments are a crucial problem in computer animation, which depends on both the path virtual

humans choose and the character motions synthesized. Many efforts have been made on the two aspects of path planning and motion synthesis respectively and great advancements have been achieved. However, it is still a challenging task in computer animation to create realistic and natural motions of character animation actively responding to complex dynamically varying environments.

To solve this problem, a novel comprehensive framework has been proposed in this paper which directs motion synthesis of virtual humans in complex dynamic environment by a global optimal path planned.

For path planning, SIPP algorithm [1] is used as path planner to find a globally optimal path from current location in a complex dynamic environment to another where virtual character want to go in the presence of dynamic obstacles, which adds time as an additional dimension into the search-space explored by the path planner to properly handle moving obstacles and introduces the concept of safe time intervals to greatly reduce the number of states that need to be searched. As the control information for directing motion synthesis of virtual character that drives virtual human follow the path planned naturally and realistically, three types of control anchors are defined, including turning anchors, height anchors and time anchors, and extracted from the path planned.

For motion synthesis, a popular traditional approach is motion graphs, which gains the motion that satisfies user demand by compressing connections among different pieces of motions in database and searching this graph. Although this is conceptually intuitive and broadly applicable, graph-based methods lack the adequate flexibility and controllability to synthesize agile reactivity of the animated character to the dynamically changing environment due to the sparse and finite states in the graph. In our framework, motion fields [2] approach is used as motion synthesizer, which organizes motion dataset into a high-dimensional

---

This research is partially supported by National Natural Science Foundation of China (61370127, 61100143), Program for New Century Excellent Talents in University (NCET-13-0659), Fundamental Research Funds for the Central Universities (2014JBZ004), Beijing Higher Education Young Elite Teacher Project (YETP0583). { Corresponding author: Weibin Liu, wblu@bjtu.edu.cn }

generalization of a vector field of state space, then generates an animation by freely flowing through the motion field in response to interactive controls. Since each state in motion field has a set of candidate actions and Reinforcement Learning is adopted to find an action that leads to a desirable control, these enable real-time controlled highly responsive motion synthesis for motion field instead of waiting for pre-determined transition points for motion graph. So motion field is able to control character motion at arbitrary state in a continuous state space for motion synthesis, which allows highly agile motion controls with control anchors for adapting to dynamically varying environment and guaranteeing the synthesized character motion well-fitting to the global optimal path planned.

## II. RELATED WORKS

Our work is mainly related to two fields: path planning and motion synthesis for character animations. Many previous works have been done in both fields.

Path planning approaches usually can be divided into undirected and directed. Undirected approach seeks to go blindly through a maze, which has two main approaches including Depth-first search and Breadth-first search [3]. These approaches expand searched nodes without directing, so that they may be unable to find a way out. Some measures of assessing have been introduced to solve this problem, which introduce directed approach. Dijkstra's algorithm [4] and A\* algorithm [5,6] estimate the cost of moving by measuring the distance between the nodes to direct which node should be expanded. Although this approach can find a path with the lowest cost, the path is not always the most efficient solution [7]. In addition, these approaches can't solve the cases with dynamic obstacles due to pre-processing computation, where the agent may wait for all dynamic obstacles passing or collide with dynamic obstacles. SIPP [1] is proposed to plan an optimal path in complex dynamic environments with moving obstacles by introducing the concepts of the safe interval and the earliest arrival time, which is also implemented as the path planner in our work to handle the moving obstacle in complex dynamic environment.

Over past decade, creating character animations from raw mocap data have been studied extensively in research and industry, which may offer many advantages over traditional computer animation. Since motion capture system is not cheap and already captured motion data don't always reflect our needs, it is necessary to reuse and edit the existing motion capture data for synthesizing new required motion rather than capturing new motion whenever needed. To improve reusability of motion capture data for reusing and editing have become a challenging problem in mocap data-driven motion synthesis.

Motion graphs [8-10] were proposed simultaneously by several groups in 2002 and have been emerged as a primary technique for automatic synthesis of character motions at runtime, which employ the bag-of-clips data structures and generate motion by concatenating short motion clips to accomplish the desired task. In the subsequent years, many

approaches have been developed to extend and augment motion graphs. Shin et al. [11] proposed parameterized motion graphs; Safonova et al. [12] applied interpolated motion graph and optimal search; Ikemoto et al. [13] adopted multi-way motion blending to generate transitions quickly; Ren et al. [14] proposed an optimization-based graph that combines continuous constrained optimization with graph-based motion synthesis; Zhao et al. [15,16] searched motion graph to extract a minimum size sub-graph and build interpolated segments of motion with same contact pattern to keep best transition. So the motion synthesized is constrained by the mocap dataset that the graph is built from, and the structural control of motion graph lacks the controllability to synthesize agile response to a dynamically changing interactive environment so that synthesized motions of virtual humans can't respond quickly to changes of walking direction or unexpected disturbances adapting to the complex and dynamic environment.

Some researchers have been conducted on remedying the drawback of space discontinuous in motion graph. Wang et al. [17] introduce Gaussian process dynamical models (GPDMs) for learning models of human pose and motion from high-dimensional motion capture data, and high-quality motions can be synthesized from the learned model. Ye and Liu [18] described an optimal feedback controller for motion tracking that allows for on-the-fly re-planning of long term goals and final completion time. This technique is able to synthesize motion from arbitrary frame and the controller can handle perturbations and changes in the environments, but it implements only an approximation of the optimality condition for the synthesized motion of virtual character at current state. Instead of modeling the most possible single motion, motion field approach [2] modeled a set of candidate motions at each motion state and selected a motion by employing an optimal control framework based on Reinforcement Learning, which freed the character from simply replaying the motion data and allowed controlling virtual humans interactively in a fully continuous motion state space. Since motion field models a series of candidate actions at each state, virtual humans can respond to motion controls immediately rather than waiting for pre-determined transition points as in motion graph, which is significantly more agile in incorporating the dynamic varying environment. So in our work, motion fields approach is employed to synthesize motions directed by the control anchors obtained in path planning, which enables agile motion control and well-fitting to the planned path to adapt to complex dynamic environments.

## III. PATH PLANNING AND CONTROL ANCHORS EXTRACTING

In a complex dynamic environment, SIPP algorithm as path planner is responsible for determining an optimal collision-free path toward where the virtual human to move and generating intermediate targets on the path to a final goal. Control anchors are then extracted from the obtained path for directing motion synthesis of virtual human in Section 4, which guarantee the synthesized character motion

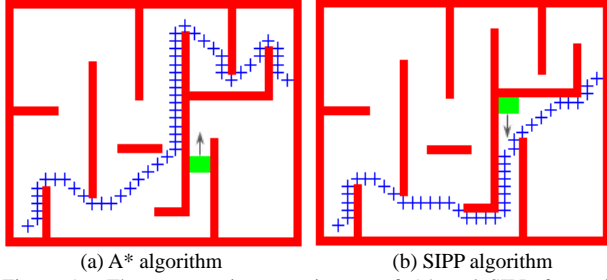


Figure 1. The comparative experiments of A\* and SIPP for path planning in an example dynamic environment. (a) Red lines represent obstacles, the coordinates of start and goal points are (3, 3) and (29, 22) respectively. The dynamic obstacle (green rectangle) in the tunnel moves back and forth. A\* algorithm generates a suboptimal path (blue crosses), because it treats the dynamic obstacle as a static one which results in an inefficient solution. (b) In the same environment, SIPP generates a more optimal path by waiting and then moving.

following the planned path naturally and realistically.

#### A. SIPP Path Planning

Although the problem of path planning in static environments is solved successfully, planning in complex dynamic environments is still a difficult problem to handle the moving obstacles. In our work, SIPP (Safe Interval Path Planning) algorithm [1] is used as path planner to find an optimal path in dynamic environment. When planning path in dynamic environments, adding time as an additional dimension to the search-space is needed in order to handle moving obstacles which will lead to the large increase in the number of states to be searched and much longer the planning time. To solve the computation problem, SIPP uses safe intervals to compress the dimensionality of time-based path planning. Safe interval is defined as a contiguous period of time for a given spatial configuration during which there are no collisions, and it is in collision one timestep prior and one timestep after the period. SIPP builds on the observation that there are usually significantly fewer safe intervals than time discretization that compose those intervals, so it constructs a search space with a few states defined by safe time interval and configurations to plan an optimal path in dynamic environment. In additional, for a static environment with no moving obstacles, SIPP has the same planning result as A\*. Fig. 1 shows a comparative experiments of A\* and SIPP for path planning in an example dynamic environment, which shows SIPP is capable of handling moving obstacles in dynamic environment. The searching procedure of SIPP is similar to A\* algorithm except how it generates successors of a state and updates time variable for states

#### B. Control Anchors Extraction

To generate a realistic animation that drives virtual humans to follow the path, we extract control information for motion synthesis from the path SIPP planned, which is defined as control anchors including turning anchors, height anchors and time anchors.

**Turning Anchors** In complex environments, an optimal collision-free path usually is not a straight line, so path guidance-based motion synthesis needs virtual human make

proper turning action at the inflexion points, which first needs to confirm the position of inflexion points along the path. We search the inflexion points along the path by checking whether the slope of line segments connecting each pair points on the path has changed or not. If the slope has changed, the searching function marks the point as an inflexion point, records the coordinates of the inflexion point  $(x, z)$ , and computes the turning angle  $\theta$  by the slopes of two adjacent line segments passing the inflexion point. A triple  $((x, z), \theta)$  is obtained for the inflexion point, termed turning anchor. A triple sequence is so obtained by checking the slopes of all the line segments between each pairs of adjacent points, which direct the virtual human to make proper turning along the planned path, as shown in Fig. 2(a).

**Height Anchors** In real situation, for getting the most efficient path, some obstacles are not necessary only be bypassed, such as very small creek, step tool and so on. In our work, we define this type of obstacle as passable obstacle, and the passable obstacles are treated as passable area in SIPP path planning. For motion synthesis, the control anchors are introduced by marking the passable properties to control the motion synthesis of virtual human in response to these passable obstacles. When encountering these height anchors, virtual humans should be informed to synthesize suitable motion like jumping, rather than only bypassing them. By considering the size of obstacle and virtual human's height, all the obstacles in the environments can be classed into two types: passable obstacle and impassable obstacle. If the obstacle is passable, we record the position  $(x, z)$  that the planned path intersects the passable obstacle, and the length  $l$  that the planned path traverses the obstacle. A triple  $((x, z), l)$  is obtained for the passable obstacle, termed height anchors. When meeting the height anchors, the virtual human can pass the obstacles by choosing proper motion behaviors, like jumping over the creek or going up stair, as shown in Fig. 2(a).

**Time Anchors** In dynamic environments, virtual humans need waiting and then moving for avoiding dynamic obstacles, it is necessary to introduce the safe waiting position and the waiting time for motion synthesis control of virtual human. SIPP introduces timestep parameter which defines the safe time interval and the early

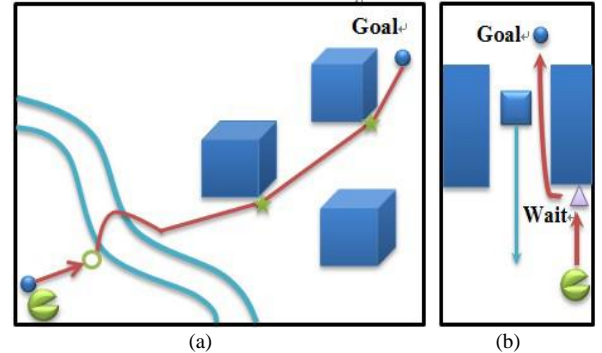


Figure 2. An illustration for control anchors. (a) An example environment with a small creek and some boxes. Mark turning anchors (green stars) and height anchor (green circle) for the optimal path (red line). (b) SIPP finds a solution by waiting for the obstacle to pass and then proceeding. We mark time anchor (purple triangle) for the path in dynamic environment.

arrival time to measure the time configuration. So we extract the position of waiting point  $(x, z)$  and the waiting time  $h$  from SIPP planning, a triple  $((x, z), h)$  is obtained for each waiting point, termed time anchor. A triple sequence is so obtained and introduced into motion synthesis, and then direct virtual human to perform waiting and moving behaviors to avoid dynamic obstacles, as shown in Fig. 2(b).

#### IV. PATH DIRECTED MOTION FIELD-BASED MOTION CONTROL

Driving virtual humans to follow the planned path requires that virtual humans can quickly respond to motion controls. In our work, motion field approach is implemented as motion synthesizer, which combines continuous state space for synthesis and highly agile interactive control that adapt to complex environments.

##### A. Motion Fields Constructing

Motion fields[2] organize frames of motions into a high-dimensional vector field and synthesize motion by flowing through the motion field according to the control commands. In motion field, each frame of motion is expressed as motion state  $m$ . It is composed of pose  $x$  and velocity  $v$ . For a pose  $x = (x_{\text{root}}, p_0, \dots, p_n)$ ,  $x_{\text{root}}$  is a 3d root position vector,  $p_0$  is a root orientation quaternion and  $p_1, \dots, p_n$  are joint orientation quaternions. And a velocity can be computed by a pair of successive poses  $x$  and  $x'$ , which is represented as  $v = (x' \ominus x)$ . Finally, a motion state can be defined as  $m = (x, v) = (x, x' \ominus x)$ . All of motion states form a high dimensional continuous motion state space, in which each state represents a single frame and a path through this state space represents a continuous motion sequence.

New state is generated by a set of neighbors of current state in motion field. Given a motion state  $m$ , we calculate  $k$  most similar states  $N(m) = \{m_i\}_{i=1}^k$  of state  $m$  via a  $k$ -nearest neighbor [19] and the distance measure. Since virtual humans may deviate from motion states in motion database, one appeal solution is to interpolate data by using the neighborhoods  $N(m)$  of the current state and similarity weights  $\omega_i$ . To generate next state, it also requires a 'direction' from current state to next state, termed action. Therefore, define the value of motion state  $m$  in motion field as a set of control actions  $\mathcal{A}(m)$ , which determines next state.

##### B. Motion Synthesis Directed by Control Anchors

Considering long term consequence of action virtual humans chooses, MDP (Markov Decision Processes)[2] is used to model decision processes. In order to control virtual humans, we introduce a task parameter  $\theta_T$  to track how well virtual humans responds to control commands. Then the form of motion state is represented as task state  $s = (m, \theta_T)$  and each task state  $s$  in motion field has a set of actions  $\mathcal{A}(s)$ . Using a task state  $s$  and an action  $a$  can synthesize next task state  $s' = I_s(s, a) = I_s(m, \theta_T, a) = (I(m, a), \theta'_T)$ . In addition, for making virtual humans respond control command better, we introduce a reward function  $R(s, a)$ , which is used to measure the reward for performing action  $a$  at state  $s$  [20]. Reinforcement learning is used to find the

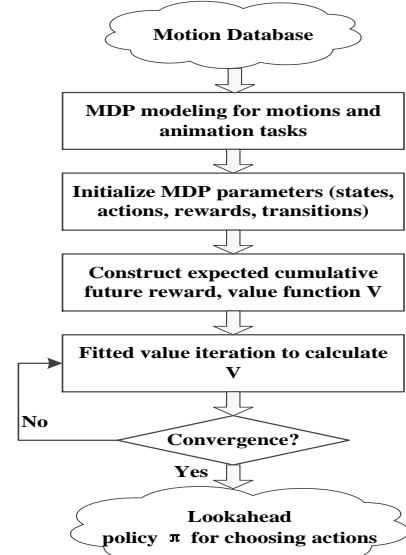


Figure 3. The process of constructing control policy.

optimal policy for choosing which action to perform at current motion state. It considers the effect of the current action on the future reward by using a look-ahead policy, which is computed by the cumulative rewards of the current and the future.

Unfortunately, the value function is difficult to calculate precisely because there are infinite task states. Instead by storing values of finite task states  $s_i$  and interpolating values to approximate the real value function, these finite task states are selected by sampling task parameters and taking the Cartesian of motion state  $m_i$ . And the value out of database is estimated by interpolating neighbors of current motion state with the similarity weights and the task parameters. Given an MDP model of motion field and task parameters, the value function is approximated via fitted value iteration [21]. The process of constructing control policy [2] is shown as Fig. 3.

Directed by turning anchors, we construct a motion field for walk and turn behaviors. The reward function  $R$  is defined as:  $R_{\text{turning}}(m, \theta_d, a) = -\omega_c |\theta - \theta_d|$ . Where  $\theta$  is the virtual character's actual movement direction,  $\theta_d$  is the desired orientation of movement, and  $\omega_c$  is coefficient.

For our experiment, we show two examples for synthesizing turning motions. For the walk to turn example, our database consists of a straight walk motion and a right turn walk motion. We use motion field to synthesize a motion which is from straight walk to turn, as shown in Fig. 4. Fig. 4(b) and 4(c) use the equal time interval sampling, as we can see, to move the same distance 4(c) spends shorter time than 4(b), so virtual human with synthesized motion turns more sharply than the original motions in database. For the run to turn, our database consists of a straight run motion and a left turn run motion. Fig. 5 shows the synthesized motion which is from straight run to turn. These two examples demonstrate that virtual human can be controlled agilely by using motion field and the synthesized



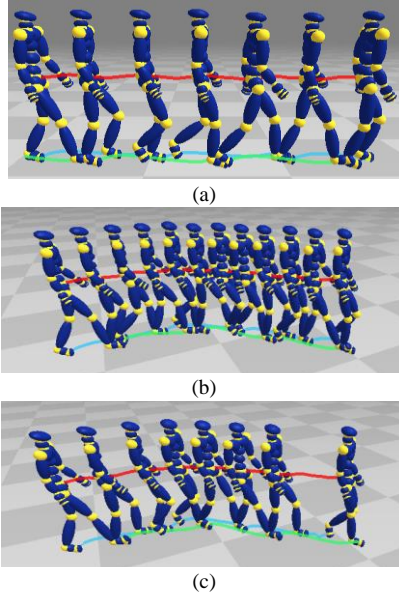


Figure 4. An example for synthesized walk to turn motion. (a) and (b) show original motion capture data; (a) shows a walking motion and (b) shows a walk to turn motion. (c) shows the synthesized motion by using motion field, which is from straight walk to turn right. (b) and (c) use the equal time interval sampling. The curves in red, green and blue respectively represent the trajectory of root joint, left foot and right foot of virtual human.

motion is natural. On the other hand, the synthesized motion is not explicitly specified in database.

For synthesizing motions between different kinds of motions, if we construct only one motion field then the synthesized motion may be in an infinite loop, so this problem can be transformed into the transition of different motion controllers. Since the optimal action depends on the

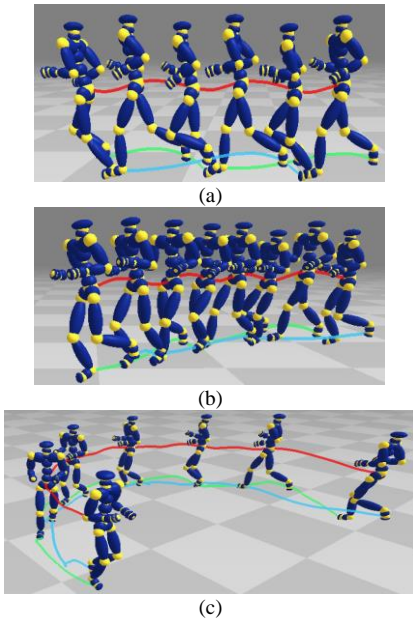


Figure 5. An example for synthesized run to turn motion. (a) and (b) show original motion capture data; (a) shows a straight run motion and (b) shows a left turning run motion. (c) shows the synthesized motion by using motion field, which is from straight run to turn left.

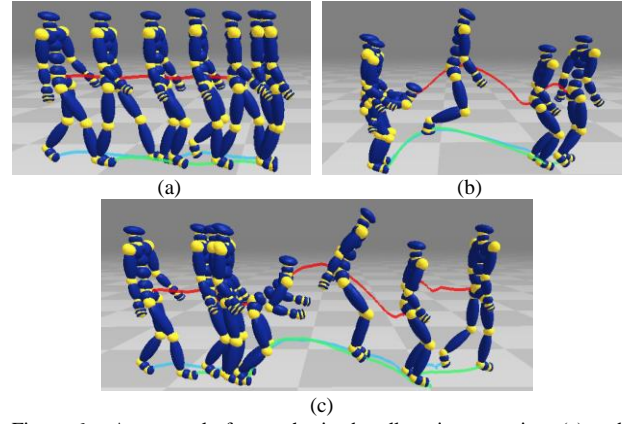


Figure 6. An example for synthesized walk to jump motion. (a) and (b) show original motion capture data; (a) shows a straight walk to stop motion and (b) shows a stop to jump motion. (c) shows the synthesized motion by using motion field, which is from straight walk to jump.

reward function and value function of target controller only, and the transition between motion clips is allowed, so the transition between different motion controllers is realizable. This means that we can realize it by constructing a transition controller to synthesize various motions [22]. We construct transition controller by defining score metric and iteration motion selecting process. The reward function of transition controller is defined as the reward function of target controller and using reinforcement learning algorithm to find a path between transition controller and target controller, i.e. optimal policy, which can be expressed as value function. In our experiment, we construct motion transition controllers for walk and jump motions. We define the reward function of walk controller as  $R = -|\theta_c - \theta_d| - |v_c - v_d|$  and the reward function of jump controller as  $R = -|\theta_c - \theta_d| - \omega J(h, k)$ , where  $\theta_c$  and  $\theta_d$  express the practical and target direction of virtual human,  $v_c$  and  $v_d$  express velocity,  $J(h, k)$  is the reward function of rebound landing of virtual human.

We show two examples for synthesizing inhomogeneous motions. For the example of the walk to jump, our database consists of a straight walk to stop motion and a stop to jump motion. Directed by height anchors, we use motion fields to synthesize a motion which is from straight walk to stop and then to jump, as shown in Fig. 6. In this example, stop motion as a transition clip to synthesize the motion which is from walk to jump. Synthesizing inhomogeneous motions can be realized by using transition controllers. Moreover, comparing the synthesized motions with motions in database, we can find that the process of synthesizing motion by motion field can not only be cutting and pasting the motions in database, but also generate motions differently from database to enrich motion data.

Directed by time anchors, we firstly construct a motion field for walk and stop behaviors. When virtual human moves to point (a, b), then motion field call the reward function  $R$  is defined as:  $_{time}(\mathbf{m}, \mathbf{v}_d, h, \mathbf{a}) = -\omega_d |\mathbf{v} - \mathbf{v}_d|$ . Where  $\mathbf{v}$ ,  $\mathbf{v}_d$  are the actual and desired movement velocity, and  $\omega_d$  is coefficient. This controller can realize the movement from walk to stop. Then we synthesize states



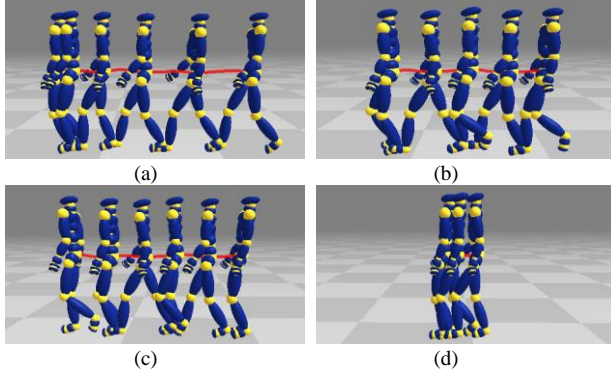


Figure 7. An example for synthesized stop to walk then walk to stop motions. (a) and (b) show original motion capture data; (a) shows a walk to stop motion and (b) shows a walk motion. (c) shows the synthesized motion by using motion field, which is from stop to walk; and (d) shows the synthesized motion, which is from walk to stop.

with no direction to wait the dynamic obstacle passing. Finally, we construct a motion field for stop and walk behaviors to control virtual character to move on when the obstacle passed.

Fig. 7 shows an example for synthesizing motions where walking and stopping switches each other. Some motion clips containing walk-stop switches are selected from the database, and motion field is built to synthesize the switching motions. The synthesized switching motions are natural and response more quickly than the origin motions in the database.

## V. EXPERIMENTS

In this section, we present the results and analysis of motion synthesis in complex environment by combining

path finding and motion fields. We begin with a path planning example.

### A. Path Planning in Complex Environment

We show an example for comparing the paths plan by A\* and SIPP in complex environment. Due to SIPP algorithm takes time factor into consideration by introducing concepts of the safe time interval and the earliest arrival time, when in the case of dynamic environment, virtual human walks along the path that can avoid dynamic obstacles by wait and move, as shown in Fig. 8(a). However, A\* algorithm treats dynamic obstacle as static, which results in having no solution or suboptimal solution, as shown in Fig. 8(b). Obviously, in our experiment, the path SIPP plans is shorter and more reasonable than the path A\* plans. The example demonstrates that SIPP algorithm may be used to plan an optimal path in both static and dynamic environment. Furthermore, the process of SIPP planning an optimal path in dynamic environment is shown in Fig. 8(c)-(j).

### B. Marking Control Anchors

Since that virtual character responses to control commands requires a certain extent of response time, even though it is usually short. To get rid of virtual human deviating from the path seriously, we define turning point of the path as the center and half step of virtual character as radius, when the coordinate of root joint in this circular area, virtual human turns corresponding angles. The turning anchor describes the turning position and turning angle. The time anchor characterizes the position of waiting and how long to wait which directs virtual character to avoid collisions with dynamic solid obstacles. And the height anchor marks the passable obstacles.

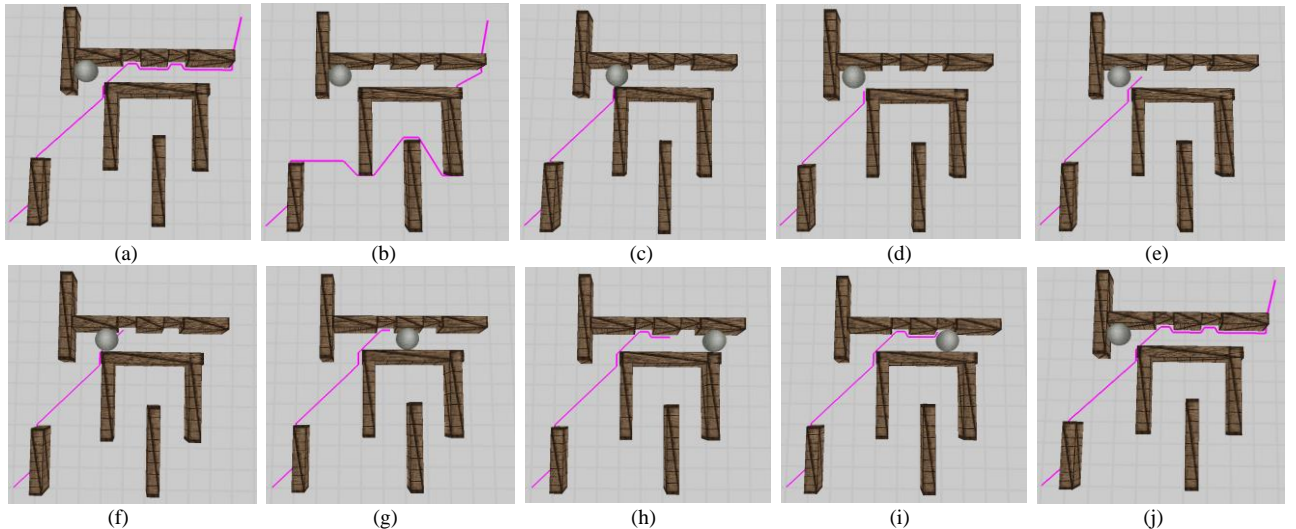


Figure 8. The comparison of A\* and SIPP algorithms and the process of SIPP planning a path in dynamic environment. (a) The optimal path which is planned by SIPP. (b) The suboptimal path A\* generates. (c)-(j) The process of planning the path with SIPP algorithm. (c) The initial environment with virtual human trying to get to the target point. The dynamic ball in the tunnel moves back and forth. The dynamic ball can move three cells each timestep and virtual human can move one cell each timestep. (d) The ball moves to the left, while virtual human waits 2s for the ball passing. (e)-(g) The ball on the left moves to the right. When virtual human is about to colliding with the ball, it walks into the first gap and moves on. (h)-(i) The ball on the right moves to the left. When virtual human is about to colliding with the ball, it moves into the second gap until the ball passed and moves to the target point (j)

### C. Motion Control in Complex Environment

Finally, we show an example of motion control in complex dynamic environment which is 9m\*9m. We discretize the environment into 450\*450 grids. Our motion database, which is shown in Table 1, consists of various motions in CMU [23]. All of the motion data in our experiments are collected at the rate of 120 frames/second.

TABLE I. MOTION CAPTURE DATABASE

Motion Type	Name	Motion Type	Name
Normal straight walk	16_15.amc	Slow straight walk	16_31.amc
	16_16.amc		16_32.amc
	16_21.amc		
Straight walk, turn left 45 °	16_11.amc	Straight walk, turn right 45 °	16_13.amc
	16_12.amc		16_14.amc
Straight walk, turn left 90 °	16_17.amc	Straight walk, turn right 90 °	16_19.amc
	16_18.amc		16_20.amc
Slow Walk, stop	16_33.amc	Stop, jump	16_09.amc
	16_34.amc		16_10.amc

For a complex example environment which is shown in Fig. 9, wooden boxes and solid spheres are impassable obstacles and the creek is passable obstacle. The initial point and two target points are specified by user. Then virtual character immediately chooses an optimal path from the initial point to the target point to follow. Firstly, virtual human moves from the initial point on the left to the first target point on the right by avoiding collisions with static obstacles, including wooden boxes and creek. Directed by turning anchors, virtual human bypasses the wooden boxes, as shown in Fig. 10 and directed by the height anchors, it chooses jumping to pass the creek, as shown in Fig. 11. When virtual human arrives at the first target point, then define this target point as the initial point and moves on.

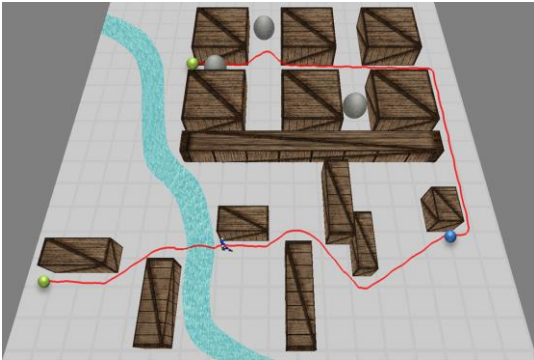


Figure 9. An example complex environment with some wooden boxes, a small creek and three dynamic solid spheres. The left bottom point in green is the initial point, the right point in blue is the first target point and the other one in green is the target point.



Figure 10. Virtual character avoids collision with static obstacles by bypassing them.

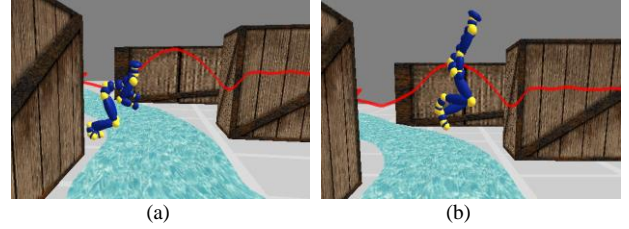


Figure 11. Virtual character crosses the creek by jumping. (a) Virtual human chooses jump movement to overleap the passable obstacle. (b) Virtual human jumped over the creek.

While virtual character moves into the gallery, directed by time anchors, it avoids collisions with dynamic obstacles by wait and move until to the target point. The initial position of solid spheres is shown in Fig. 9. The process of virtual character avoiding the dynamic solid spheres is shown in Fig. 12. This example demonstrates that virtual character can choose the optimal path to follow and select an appropriate behavior to respond control anchors in environment. Furthermore, the synthesized motion is natural and different from the motion in database.

## VI. CONCLUSION

This paper presented a comprehensive framework for virtual human animation and control in complex dynamic environment by planning an optimal path to direct motion synthesis of virtual character. We have shown how to plan the feasible path in complex dynamic environment by using SIPP algorithm. For getting the control information from the optimal path, we defined types of control anchors for the path as control commands. Directed by these control anchors, motion field algorithm is used to synthesize motions of virtual character in complex dynamic environment. Since it is continuous, it can overcome the drawbacks of graph-based approach, such as responsiveness and agility. Our method is effective to address the issue that motion synthesis of virtual human in complex dynamic environment. Furthermore, the ability of our approach to construct animations of virtual human easily and without expertise would enable many novice users and compelling applications.

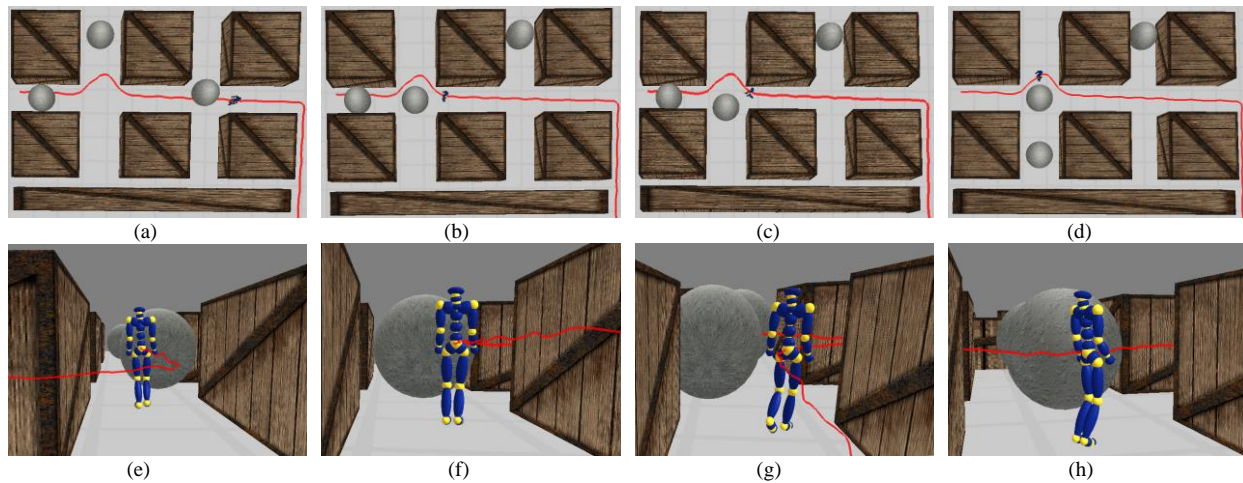


Figure 12. Virtual human avoids all collisions by wait and move in complex dynamic environment. The top column is planform and the bottom column is the counterpart in environment. The dynamic solid sphere at the bottom moves up, the one at the upper moves down and the other on the left moves to the right. The first two solid spheres move 30 cells each second and the last solid sphere moves 15 cells each second. (a), (e) Virtual human waits two seconds for the first solid sphere passing and moves on. (b), (f) Virtual human waits two seconds for the second solid sphere passing and (c), (g) moves up with no waits. (d), (h) Virtual human waits one second for the last solid sphere passing and moves down and then moves to the left.

Because SIPP algorithm plans a path in dynamic environment with timestep parameter, which leads to unrealistic locomotion of virtual character. Additionally, motion field approach is limited by the given motion data, for instance, when the virtual human deviates from the data, the realism of the synthesized motion degraded. For future work, we will optimize the path SIPP plans in dynamic environment and introduce physical dynamics into motion synthesis to extend the physical plausibility.

## REFERENCES

- [1] M. Philips and M. Likhachev, "SIPP: Safe Interval Path Planning for Dynamic Environment," Proc. IEEE Intl Conf. on Robotics and Automation (ICRA), Shanghai, China, pp.5628-5635, May. 2011.
- [2] Y. Lee, K. Wampler, G. Bernstein, et al., "Motion fields for interactive character locomotion," ACM Trans. on Graph. (TOG), vol. 29, no. 6, Dec. 2010.
- [3] S. J. Russel and P. Norvig, Artificial Intelligence A Modern Approach. Prentice-Hall, Inc, 2002
- [4] "Dijkstra's algorithm," [http://en.wikipedia.org/wiki/Dijkstra's\\_algorithm](http://en.wikipedia.org/wiki/Dijkstra's_algorithm), Sep.2012.
- [5] P. E. Hart, N. J. Nilsson and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," IEEE Trans. on Systems Science and Cybernetics, vol. 4, no. 2, pp. 100-107, Jul. 1968.
- [6] P. Lester, "A\* Pathfinding for Beginners," <http://www.policyalmanac.org/games/aStarTutorial.htm>, Jul, 2005.
- [7] R. Graham, H. McCabe and S. Sheridan, "Pathfinding in Computer Games," ITB Journal, no.8, pp. 57-81, Dec. 2003.
- [8] L. Kovar, M. Gleicher and F. Pighin, "Motion graphs" ACM Trans. on Graph. (TOG), vol. 21, no. 3, pp.473-482, 2002.
- [9] J. Lee, J. Chai, et al., "Interactive control of virtual human animated with human motion data," ACM Trans. on Graph. (TOG), vol. 21, no. 3, pp.491-500, Jul. 2002.
- [10] O. Arikian and D. A. Forsyth, "Interactive motion generation from examples," ACM Trans. on Graph. (TOG), vol. 21, no. 3, pp. 483-490, Jul. 2002.
- [11] H. J. Shin and H. S. Oh, "Fat graphs: constructing an interactive character with continuous controls," Proc. ACM SIGGRAPH/Eurographics symposium on Computer animation., pp. 291-298, Vienna, Austria, 2006.
- [12] A. Safonova, J. K. Hodgins, "Construction and optimal search of interpolated motion graphs," ACM Trans. on Graph. (TOG), vol. 26, no. 3, Jul. 2007.
- [13] L. Ikemoto, O. Arikian and D. Forsyth, "Quick transitions with cached multi-way blends," Proc. 2007 symposium on Interactive 3D graphics and games, pp. 145-151, Seattle, WA, USA, 2007.
- [14] C. Ren, L. Zhao and A. Safonova, "Human motion synthesis with optimization-based graphs," Comput. Graph. Forum, vol. 29, no. 2, pp. 545-554, May 2010.
- [15] L. Zhao, A. Normoyle., S. Khanna, et al., "Automatic construction of a minimum size motion graph," Proc. the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. pp. 27-35, New Orleans, LA, USA, 2009.
- [16] L. Zhao and A. Safonova, "Achieving good connectivity in motion graphs," Graphical Models ACM SIGGRAPH / Eurographics Symposium on Computer Animation, vol. 71, no. 4, pp.139-152, Jul. 2009.
- [17] J. M. Wang, D. J. Fleet, And A. Hertzmann, "Gaussian process dynamical models for human motion," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 283-298, Feb. 2008.
- [18] Y. Ye and C. K. Liu, "Optimal Feedback Control for Character Animation Using an Abstract Model," ACM Trans. on Graph. (TOG), vol. 29, no. 4, Jul. 2010.
- [19] D. M. Mount and S. Arya, "Ann: A library for approximate nearest neighbor searching," Jan. 2010.
- [20] W. Y. Lo and M. Zwicker, "Real-time planning for parameterized human motion," Proc. the 2008 ACM SIGGRAPH/Eurographics symposium on Computer animation, pp.29-38, Dublin, Ireland, 2008.
- [21] E. Damien, G. Pierre, W. Louis, et al., "Tree-based batch mode reinforcement learning," Journal of Machine Learning Research, vol. 6, pp.503-556, Apr. 2005.
- [22] Y. Lee, "Decision Mechanisms for Real-time Character Animation," University of Washington, 2010.
- [23] CMU, "Carnegie-Mellon mocap database," <http://mocap.cs.cmu.edu/>, 2006.



# Athena: Capacity Enhancement of Reversible Data Hiding with Consideration of the Adaptive Embedding Level

Ya-Chi Hsu, Bo-Chao Cheng

Dept. of Communications Engineering  
National Chung Cheng University  
Chia-Yi, Taiwan

alicehsu00@gmail.com, bcheng@ccu.edu.tw

Huan Chen\*

Dept. of Computer Science and Engineering  
National Chung Hsing University  
Tai-Chung, Taiwan  
huan@nchu.edu.tw

Yuan-Sun Chu

Advanced Institute of  
Manufacturing with High-tech  
Innovations (AIM-HI)  
National Chung Cheng University  
Chia-Yi, Taiwan  
chu@ee.ccu.edu.tw

**Abstract**—This paper provides a novel method of reversible data hiding by adaptively adjusting the embedding level to achieve minimal distortion and to attain the embedded data confidentiality via data reversion key generation. The peak point queue (PPQ) is used to adaptively determine the embedding level and achieve superior quality in the marked image. In respect to data reversion key for the double protection, our approach, named Athena, makes better utilization of the overhead information and allows users to exchange the key based on a public key infrastructure (PKI). Our experimental results show that the method is capable of providing a better quality image for a range of different test images of various sizes.

**Keywords**—*Embedding level; Image histogram; Public Key Infrastructure (PKI); Reversible data hiding*

## I. INTRODUCTION

In the current explosion of the information era, there is a significant volume of data being exchanged between people. So far, there are several techniques that can protect data and prevent modification, such as encryption, signatures, and watermarks. It is impossible for us to encrypt any publications. We can encrypt personal information and exchange them with other people, yet in the case of public work, this is not viable. It has the same problem as the signature technique. However, we not only need protection, but also some notation to indicate our ownership and detect whether or not it has been modified [1]. A simple technique is a stamp to declare that the author is authentic, yet we have to make the stamp invisible without changing anything. Data hiding is the most typical technique.

Data hiding embeds minimal data into an image for rights protection, authentication, secret sharing, etc. It is used in military images, medical images, artwork, law enforcement, and document preservation. In general, the method embeds notation into an image and then publishes it. In order to verify the image, we can extract notation to confirm whether or not it is modified. While we can extract secret data from a marked

image, we expect to be able to recover the original image without distortion. In the case of medical images, minute changes are unacceptable and can cause potential risks in the future. For law enforcement, an authentication message is embedded. It is also important to be able to recover the cover image after the authentication message is extracted. Furthermore, in artwork protection, a watermark is embedded for ownership protection [2]. It is necessary to recover the original artwork after the watermark is extracted.

The method of data hiding with reversibility has the advantage of being able to recover the cover image without distortion, yet its hiding capacity is limited [3]. To obtain reversible data hiding with a higher hiding capacity and maintain good quality images, a histogram-based scheme using pixel differences was proposed [2]. The generated residual image was then employed to embed the secret data. The residual histogram of the values in the residual image was calculated according to the secret data size, using a binary structure to satisfy the requirements adaptively and exchange pairs of peaks with the recipient as well. Upon completion, a histogram shifting technique prevents overflow and underflow [4].

However, the requirement of the peak point value increases steeply, which influences the quality of the marked image. In this paper we will present techniques to maintain the same hiding capacity while achieving a higher quality marked image. The peak point queue (PPQ) is proposed. This method allows adjustment of the value of the peak point with the requirement of the secret data size. The data reversion key is generated after completion of the embedding process. A single image corresponds to a unique key. The sender authenticates the identity of the receiver and the key will be sent. The receiver can then extract the message using the corresponding key.

The remainder of this paper is organized as follows. In Section II, we review the reversible scheme developed earlier. Section III contains a detailed explanation and derivation of the proposed algorithm. We experimentally investigate the relationship between the capacity and distortion in Section IV. The paper concludes in Section V.

---

\*Corresponding author: Huan Chen. The work was supported by the Research Grant NSC 102-2221-E-194-006 and NSC 102-2220-E-005-008 from the National Science Council, Taiwan.

## II. RELATED WORKS

A reversible data hiding method is one where an embedded message can be extracted and the image completely restored to its original state. There are eight bits representing color for each pixel. Therefore, the range of the value of each pixel is 0 to 255. The value 0 is white, while the value 255 is black. In order to achieve reversibility, there are many limitations to overcome. The histogram-based reversible data hiding technique was proposed by Ni et al. in 2006[3], who proposed a reversible data hiding algorithm which is now considered a significant step forward in the data embedding research area. They used the zero point and peak point of the histogram of an image and slightly modified the pixel values to embed secret data into the image. In the embedding process, they searched the peak and zero points and then shifted the histogram in order to generate free space to embed data. Only the peak points were used to hide data while the others were only modified. Their work is guaranteed to have a peak signal to noise ratio (PSNR) above 48 dB. Although Ni et al. managed to achieve a short execution time and a high PSNR value, their work suffers due to its low hiding capacity.

Therefore, Tsai et al. [2] proposed a reversible image hiding scheme based on histogram shifting for medical images, and is based on the Ni et al. method [3]. In their work, they divided the original image into blocks of  $n \times n$  pixels, where the center pixel in the block is selected as the basic pixel and all pixels in the block are processed using a linear prediction technique to generate the residual values. Upon completion of the linear prediction procedure, the histogram of the residual image is generated and divided into two parts: non-negative histogram (NNH) and negative histogram (NH). They determined the peak and zero pairs in the NNH and NH, and then hide the secret data in the peak point. Their work doesn't generate overhead information and provides overlapping embedding to enlarge the hiding capacity. However, the PSNR value drops dramatically although the capacity is enlarged.

There is another method which can also improve upon the Ni et al. method [3]. Tai et al. [4] proposed a reversible data hiding scheme based on histogram modification of pixel differences. They used a binary tree structure to store the communication pairs of peak points. In the pre-processing stage, they used the pixel difference to consider the difference between adjacent pixels instead of a simple pixel value, and adapted the histogram shifting method to prevent overflow and underflow of the image. Due to the use of the pixel difference techniques, they are able to find more free space to hide the secret data in the cover image (a.k.a. original image). They claimed that their work can provide a large hiding capacity while minimizing distortion. However, the PSNR values drop rapidly according to their experimental results. This is a major flaw that we wish to address in our work.

Lee et al. [5] proposed a reversible image authentication technique which is based on histogram modification similar to Ni et al.[3], and claims to provide lower distortion. In their work, they used the difference-histogram techniques which are not regular in shape and provide a much higher peak point. The difference-histogram techniques are generated by projecting into the two dimensional histogram which are the odd- and

even- line fields. The secret messages are embedded in the odd-line field by calculating the value of the pixels. They also used the MD5 as a hash function in order to produce a 128 bit array as output. In their experimental results, the Lena image's PSNR was 52.21 dB and the capacity was 26900 bits. This shows that their capacity is quite small, and is unsuitable for modern applications.

There are some methods which have been proposed that have the ability to adjust the embedding level (peak point value). Jung et al. [6] suggested one such approach where a novel histogram modification based reversible data hiding technique using human visual system (HVS) characteristics is used to reduce the distortion of the embedded image. Application of the HVS characteristics means that it is necessary to take the edge and the just noticeable difference (JND) of pixels in the cover image into consideration. First, an edge can be decided as the boundary between a smooth area and a complex area. A smooth area change is smaller than a complex area change. In this way, the method is able to achieve a higher quality in the marked image. Second, JND is used to determine the appropriate embedding level for data embedding. In comparison to previous works, the Jung et al. method provides a sharper image with better quality. However, their method requires expensive pre-processing since it must compute the embedding level for each pixel. Therefore, we provide a queue structure which only needs to compute once at the beginning and determine a better embedding level for data hiding. According to TABLE I, it can be seen that in the early work of reversible data hiding ([3],[5]) they only focused on how to increase the PSNR value, with no regard for the consideration of other aspects.

TABLE I. COMPARISON WITH VARIOUS APPROACHES

	Tsai [2]	Ni [3]	Tai [4]	Lee [5]	Jung [6]	Proposed Athena
Enlarge capacity	O		O		O	O
Exchange peak points			O	O	O	O
Compress overhead information			O		O	O
No Need to determine threshold/block size		O	O	O		O
Adjust embedding level adaptively					O	O
Generate/ exchange security key				O		O

However, our approach and the work of ([2], [4], [6])not only provides a higher PSNR value, but also enlarges the hiding capacity. Moreover, the peak points exchange is provided in later approaches with the exception of [2] and [3]. The problem of dealing with the overhead information is another issue. Hence, [4] provides this ability as well as [6] and our method. Without threshold and block size determination, we risk reducing the performance. Despite this, the operation is in the requirement of [2] and [6]. Embedding level adjustment and data reversion key protection are the major contributions of our approach. Athena and [6] allow adaptive adjustment of the

embedding level. Using this, we are able to lower the distortion of the pure image, while the others [2]-[5] must modify an entire image to hide only a small amount of data and risk causing serious distortion. Although [6] also provides this ability, it must compute for every pixel and requires significant computational resources. Both [5] and our method provide key protection but use different generations. In regards to security, Athena and [5] exchange the key when we decide to transmit our embedded image. This technique means that we can not only reinforce the security issue, but also prevent a hacker easily cracking our data.

### III. APPROACH

This paper provides two major contributions to reversible data hiding (1). Through adaptively adjusting the embedding level, Athena achieves minimal distortion and (2). Athena makes use of overhead information and the embedding level value to generate the data reversion key providing embedded data confidentiality. The marked image and data reversion key are produced after the cover image is embedded the secret data (as shown in Figure 1). While the extraction procedure is processed, the marked image is extracted with the corresponding key/overhead information, and the cover image can be recovered without any distortion. This is what our reversible data hiding algorithm does.

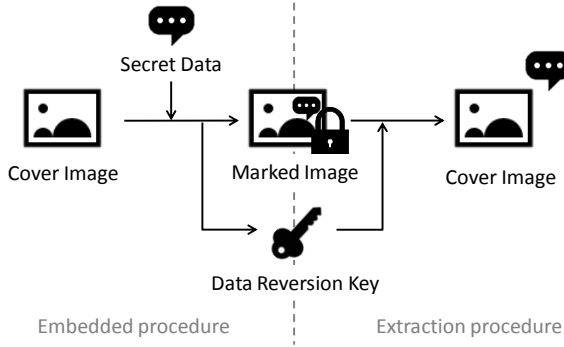


Figure 1. The reversible data hiding algorithm

A method for adjusting the embedding level is proposed here, and uses a queue structure to calculate the hiding capacity for comparison with the secret data size. In the method proposed by [4], the binary structure and level of a binary tree are used to estimate the hiding capacity, increasing it rapidly when the level is pulsed by one. If the reversed range of pixels is too large, it would cause heavy distortion.

Athena provides a more adaptive way to choose the embedding level. It only needs to compute at the beginning of the reversible data hiding process, rather than compute for each pixel one by one. Upon completion of the embedding procedure, the overhead information and the value of the embedding level are combined into a key and shared using public key encryption. The following gives the detailed procedure of Athena. TABLE II. shows some symbols that are used in the algorithm. We assume a BMP image format, gray scale image tone, 512 x 512 image size, and 8 bits per pixel.

TABLE II. SYMBOL TABLE

Symbol	Description
$\mathcal{P}$	The queue of peak points and the elements is $p_i$ which satisfies $0 \leq p_i \leq 255$ .
$\mathcal{E}$	Embedding level and satisfies $\mathcal{E} \in \mathcal{P}$ .
$\mathcal{S}$	The size of the secret data.
$x_i$	The pixel value of the cover image and satisfies $0 \leq x_i \leq 255$ .
$d_i$	The value of the histogram after pixel difference and satisfies $0 \leq d_i \leq 255$ .
$y_i$	The value of the histogram after data embedding.
$w_i$	The value in the histogram after histogram shifting and satisfies $0 \leq w_i \leq 255$ .
$\ell_i$	The value of the location map and satisfies $s_i \in 0,1$ .
$\beta$	The bit value of the secret data and satisfies $b \in 0,1$ .
$\mathcal{K}$	The data revision key of the marked image.

#### A. Peak Point Queue (PPQ)

In the proposed method, the PPQ is provided for the communication of multiple peak points. In [2] and [3], the efficient transmission of peak and zero point(s) information is a problem. However, we begin our embedding from  $p_i = 0$  in the queue and the hiding capacity is counted. We assume that the size of the secret data is  $\mathcal{S}$ . The number of peak points used to embed it is also  $\mathcal{S}$ .

$\mathcal{P}$  is the queue (set) of the peak point.  $p_i$  is the peak point value and should satisfy  $0 \leq p_i \leq 255$ . The reason we call this a queue is due to the use of the peak point being one by one; it cannot be skipped. The number of the peak point which is equal to zero will be counted one by one in the pixel difference histogram and also when the value is equal to one, two, three, etc. The number for each peak point value is then summed up one by one until the number is equal or bigger than  $\mathcal{S}$ . If the capacity is insufficient, the space of the next peak point will be included. After peak point elimination, we let  $\mathcal{E}$  equal  $p_i + 1$ . In this way, the most suitable peak point is

$$\mathcal{P} = \{1, 2, 3, 4, \dots, p_i, p_{i+1}, \dots, 255\}, \text{ where } \forall i \in \mathbb{Z} \text{ and } 0 \leq i \leq 255. \quad (1)$$

We will narrow the histogram into a row to simplify the example. The histogram is as shown in Figure 2 after the pixel difference process is complete. There are three bits of secret data that are going to be embedded in the cover image.

	0	1	2	3	4	5	6
$d_i$	125	125	2	249	2	3	2

Peak point value ( $p_i$ )	# of position
0	0
1	0
2	3
...	...
125	2

Figure 2. The result of the pixel difference and the amount of the peak points



After the pixel difference is complete and the number of pixel points is counted, it is easy to tell that when  $p_i = 0$ , there is no space for embedding. It is the same situation when  $i = 0, 1$ , which means that when a value in the histogram  $d_i$  is equal to 0 or 1, it can be embedded in a secret data bit. When  $i = 0, 1, 2$ , there are three spaces for embedding which is enough. Therefore, we let  $\mathcal{E} = i + 1 = 3$ , which is the final result of the embedding level. It can be embedded when  $d_i < \mathcal{E}$ .

Larger payloads require the use of a higher embedding level [4]. Furthermore, in the queue structure, the level will increase relatively slowly and will not cause heavy distortion. Upon completion of the data hiding, this embedding level will be combined with the overhead information to generate the data reversion key.

### B. Embedding procedure

The embedding procedure is described in detail in this section, and the related notations are denoted in TABLE II. The secret data,  $\beta$ , is assumed as  $\{1, 0, 1\}$ .

Step 1. With the exception of the first column, the cover image  $x_i$  is scanned in raster-scan order as shown in Figure 4 ([7], [8], and [9]) and the pixel difference  $d_i$  with  $x_i$  and  $x_{i-1}$  is calculated.

$$d_i = \begin{cases} x_i & , \text{ if } i = 0 \\ |x_{i-1} - x_i| & , \text{ otherwise} \end{cases} \quad (2)$$

$x_i$	125	0	2	251	249	252	254
	0	1	2	3	4	5	6
$d_i$	125	125	2	249	2	3	2

Figure 3. The pixel value of the cover image and pixel differences

Step 2. Determine the embedding level  $\mathcal{E}$  using the PPQ which was described in previous Subsection A. The  $d_i$  is the same as the example in Subsection A. Hence, the embedding level,  $\mathcal{E}$ , is equal to 3.

Step 3. In the proposed scheme,  $d_i$  is classified into two cases which are  $d_i < \mathcal{E}$  and  $d_i \geq \mathcal{E}$ , where the front case can carry one secret data bit, and the latter case cannot. Scan the whole image in the same raster-scan order (as shown in Figure 4). The  $x_i$  in the first column does nothing.

Case 1. If  $d_i < \mathcal{E}$ ,  $x_i$  is employed to carry one secret data bit  $\beta$

$$y_i = \begin{cases} x_i + (d_i + \beta), & \text{ if } x_i \geq x_{i-1}, \\ x_i - (d_i + \beta), & \text{ if } x_i < x_{i-1}. \end{cases} \quad (3)$$

Case 2. If  $d_i \geq \mathcal{E}$ , the pixel is derived with no secret data, yet the pixel movement should be expanded to discriminate this pixel from the embedded pixels. Therefore, the pixel value is obtained by

$$y_i = \begin{cases} x_i, & \text{ if } i = 0, \\ x_i + \mathcal{E}, & \text{ if } x_i \geq x_{i-1}, \\ x_i - \mathcal{E}, & \text{ if } x_i < x_{i-1}. \end{cases} \quad (4)$$

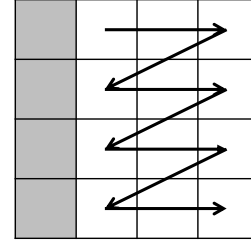


Figure 4. Raster-scan order

Steps 2 and 3 are repeated until all pixels are scanned, and the data hiding procedure is complete. However, since the pixel is shifted by processing, underflow or overflow may be caused. That is the reason why an illegal pixel value will be revised in the next Subsection C.

	0	1	2	3	4	5	6
$y_i$	125	-3	5	254	247	255	257

Figure 5. The pixel value with the embedded data.

### C. Revise the overflow/underflow

In order to revise the overflow and underflow pixel value, Athena scans the image with the embedded secret data pixel by pixel. If there are any pixels smaller than zero, the pixel value will be pulsed by  $\mathcal{E}$ . On the other hand, pixels that exceed 255 will be reduced by  $\mathcal{E}$ . According to the embedding result,  $y_i$ , (in Figure 5) Some illegal pixels with values, either greater than 255 or smaller than 0, should be further processed.

$$\begin{aligned} & \text{if } y_i < 0, \quad w_i = y_i + \mathcal{E} \ \& \ \ell_i = 1, \\ & \text{if } y_i > 255, \quad w_i = y_i - \mathcal{E} \ \& \ \ell_i = 1, \\ & \text{otherwise, } \ell_i = 0. \end{aligned} \quad (5)$$

	0	1	2	3	4	5	6
$w_i$	125	0	5	254	247	255	254

Figure 6. The pixel value revised while overflow and underflow

In this way, two pixels ( $i = 1, 6$ ) can be corrected (as shown in Figure 6). However, we need to record the position of the modified pixel as overhead information. Hence, Athena adopts the technique in [4] using a location map whose size is equal to the cover image. First, we initial the location map,  $\ell_i$ , as  $\{0, 0, 0, 0, 0, 0, 0\}$  is the same size as the cover image. While the illegal pixel value is revised, we mark a 1 at the corresponding position in the location map,  $\ell_i$ . Otherwise, it will be assigned a 0. In the previous papers [3]-[9], this was overhead information. However, we achieve better utilization of the overhead information which will be introduced in the following subsection.

#### D. Key management

In order to use information efficiently, we integrate the overhead information  $\ell_i$  and the value of the embedding level  $\mathcal{E}$  to generate a key which can provide a second protection level. Except in extreme images, the illegal values are few. Hence, run-length coding can efficiently reduce the length of the overhead information. Firstly, the location map is compressed using lossless compression, the run-length coding algorithm. Let A represent 0 and B, 1. For instance, the location map  $\ell_i$  which is shown in Figure 7 is compressed by the run-length coding algorithm, with the result also shown in Figure 7. The  $\mathcal{E}$  is connected right after the overhead information. There are three spaces reserved for  $\mathcal{E}$  since the maximum value is 255.

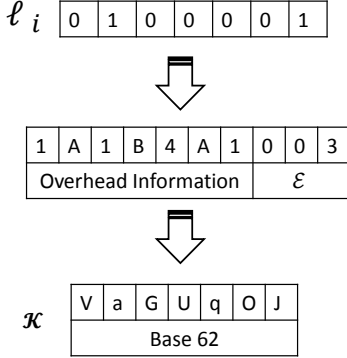


Figure 7. The method for generating the key of the marked image

Upon completion, the overhead information and  $\mathcal{E}$  are modeled as hexadecimal (or base 16), and then converted to base 62 which is composed of 0 to 9, a to z, and A to Z. The result is shown in Figure 7. The sender shares this data reversion key as well as the marked image with the recipient. Only in this way can the recipient extract the secret data successfully.

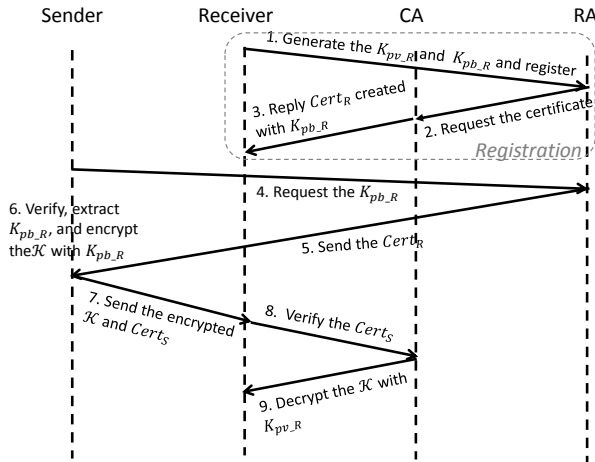


Figure 8. The public key infrastructure

However, the method for exchanging the key with the recipient correctly and securely is another key point in Athena. Following the flowchart shown in Figure 8, we use the Public Key Infrastructure (PKI) to maintain the security key

transmission which can provide authentication, confidentiality, non-repudiation, and integrity of the messages exchanged[10][11]. The certificate authority (CA) and registration authority (RA) are the trusted authorities that provide the security policies. At the beginning, the receiver needs to perform registration to create the keys, the public key  $K_{pb,R}$ , the private key  $K_{pv,R}$ , and certificates  $Cert_R$ . According to the procedure of the step 7 in the Figure 8, the data reversion key ( $\mathcal{K}$ ) of the marked image can then be shared securely.

#### E. Extraction procedure

In this process, we extract the secret data from the marked image and recover the image without any distortion. Assume that the recipient receives the marked image and corresponding data reversion key. At the beginning, we must decode the data reversion key to retrieve the necessary information. Without these, a recipient cannot successfully extract the secret data. This is why we can claim that this approach can provide a second level of protection. The key is converted from base 62 into base 16 where the last three positions are  $\mathcal{E}$ , and then the rest of the data is overhead information. When the key,  $\mathcal{E}$ , and the marked image are ready, the process can then begin.

Step 1. Before the extraction procedure, the values which have been revised must be shifted back. The location map is scanned from left to right and up to down. When the value 1 is visited, the marked image will be modified to its original state by (6). Otherwise, it will do nothing.

$$\begin{aligned} y_i &= w_i - m, \text{ if } y_i < 2m \\ y_i &= w_i + m, \text{ if } y_i \geq 2m \\ &\text{otherwise, no modification.} \end{aligned} \quad (6)$$

Step 2. After the whole image is processed by step 1, scan the image in the same way, in raster-scan order. If  $|y_i - x_{i-1}| < 2m$ , using this formula, the secret data can be extracted as

$$\beta = \begin{cases} 0, & \text{if } |y_i - x_{i-1}| \text{ is even,} \\ 1, & \text{if } |y_i - x_{i-1}| \text{ is odd.} \end{cases} \quad (7)$$

Step 3. Scan the image in the same way, raster-scan order, and recover the image using

$$x_i = \begin{cases} y_i + \left\lceil \frac{|y_i - x_{i-1}|}{2} \right\rceil, & \text{if } |y_i - x_{i-1}| < 2\mathcal{E} \text{ and } y_i < x_{i-1} \\ y_i - \left\lceil \frac{|y_i - x_{i-1}|}{2} \right\rceil, & \text{if } |y_i - x_{i-1}| < 2\mathcal{E} \text{ and } y_i > x_{i-1} \\ y_i + \mathcal{E}, & \text{if } |y_i - x_{i-1}| \geq 2\mathcal{E} \text{ and } y_i < x_{i-1} \\ y_i - \mathcal{E}, & \text{if } |y_i - x_{i-1}| \geq 2\mathcal{E} \text{ and } y_i > x_{i-1} \\ y_i, & \text{otherwise.} \end{cases} \quad (8)$$

In fact, steps 2 and 3 can be processed in parallel. The pixel secret data bit detection and restoration occur simultaneously. Finally, the embedded secret data can be extracted and the cover image is restored.

#### IV. EVALUATION

To obtain a better understanding of how different host images impact the performance of the proposed reversible data hiding scheme, we performed computer simulations on several test images. The test images are from the USC SIPI Image Databases and are 512 x 512 color images. We converted the color images into gray scale with 8 bits per pixel. The secret data was generated by a random function and was composed of binary 1 and 0. We list the experimental environment TABLE III.

TABLE III. EXPERIMENTAL ENVIRONMENT

SW/HW	Description
Processor	Intel Core i7 3.40 GHz
RAM	4 GB
Operating System	Microsoft Windows 7
Programming Language	Java (OpenCV model)

The peak signal to noise ratio (PSNR), shown in Eq. (9) and (10), was employed as a measure to compare the visual quality between the cover image and the marked image[12]. The pure payload capacity ([2]-[9]) was evaluated using the embedding rate  $\mathfrak{R}$  (bpp), shown in Eq. (11)

$$MSE = \frac{1}{MN} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2, \quad (9)$$

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX^2}{MSE} \right), \quad (10)$$

$$\mathfrak{R} = \frac{C_h - |O|}{M \times N}. \quad (11)$$

where M and N are defined as the width and height of the cover image, respectively,  $C_h$  is the hiding capacity, and  $|O|$  is the quantity of overhead information. To be fair, in this experiment, we assumed that the overhead information wasn't embedded in the image. However, in[4], they embedded the secret data as well as the overhead information.

##### A. Hiding capacity versus distortion

Figure 9 and TABLE IV. show experimental results with severe distortion due to the large embedding size. The cover images are hiding 31.68 KB of secret data. In the optimal case, the host image should provide the same size. However, according to the Lena image, there is a capacity of 31.68 KB provided in Athena while [4] provides 31.92 KB. The extra spaces are embedded with a 0 bit since it will not cause further distortion, yet other pixels will be shifted by a larger peak point value and will cause heavy distortion. It is easy to see that there is a range of damage on Lena's forehead. It can also be seen that the nose in the Baboon image is also distorted. Nevertheless, these two images have the embedded secret data, and the distortion is more severe in the Baboon since the Baboon is a complex and sensitive image.

Next, we let the Lena image be embedded with different sizes of secret data using a distinct embedding rate R. In Fig. 10, it is easy to tell that we have the same quality or better based on the same embedding rate. For the small amount of secret data, it is difficult to see the difference in the distortion

since, in [4], the peak point value of the binary tree is small in the beginning. However, the value increases rapidly (for instance, 1, 2, 4, 8, 16, etc.). As the size of the secret data increases, the amount of the peak point also increases. On the other hand, the value of the peak point becomes large as well and causes a large scale shift in the embedding procedure. Hence, in the end, the differences in the PSNR values increase.

TABLE IV. ATHENA VERSUS TAI ET AL.

		Athena	Tai et al. [4]
Lena	PSNR (dB)	30.46	24.77
	Hiding Capacity (KB)	31.68	31.92
Baboon	PSNR (dB)	19.80	14.37
	Hiding Capacity (KB)	31.69	31.90

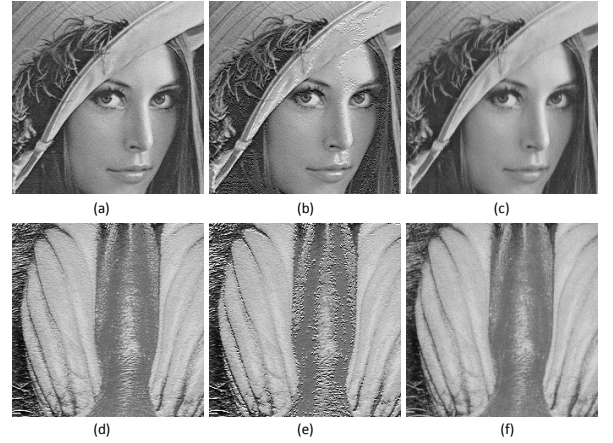


Figure 9. Part of the marked images: (a) (d) Athena; (b) (e) Tai et al.; (c) (f) Original.

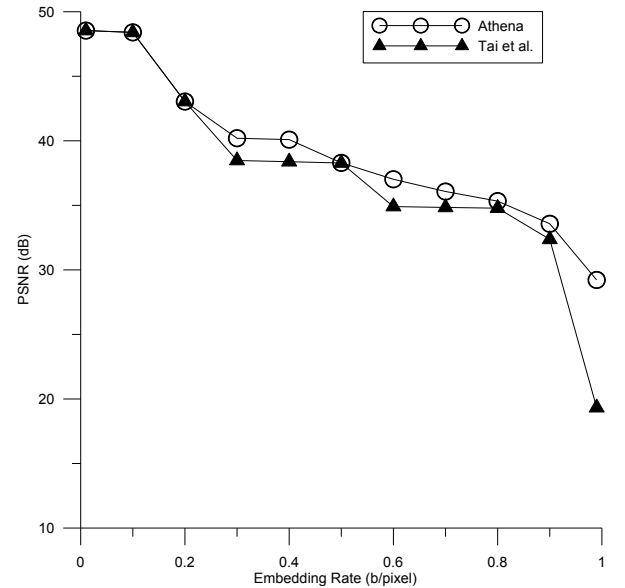


Figure 10. The embedding rate versus PSNR

##### B. Embedding rate for the test images

According to the result in TABLE V. , we can use different images to embed the amount of secret data to prove that it can

achieve almost the same performance irrespective of whether the image is smooth or complex. When  $R=0.01$ , the average PSNR of Athena and [4] was 48.52 and 48.5, respectively. However, this increases with an increased embedding rate. When  $R=0.99$ , the average PSNR of Athena and [4] was 26.51 and 22.39, respectively. When  $R=0.6$ , some of the performance was the same. Nonetheless, this situation is normal since the result of the peak point determination may be multiples of 2 and the value of the peak point will be the same. The overall performance is still better than [4].

TABLE V. PERFORMANCE EVALUATION IN DIFFERENT IMAGES

Embedding rate ( $\mathcal{R}$ )	Athena			Tai et al. [4]		
	0.01	0.6	0.99	0.01	0.6	0.99
Lena	48.61	37.01	30.47	48.54	34.91	24.77
Peppers	48.44	35.21	29.44	48.44	33.84	26.08
F16	48.87	38.99	27.79	48.87	38.99	24.67
Sailboat	48.41	33.69	25.06	48.41	32.92	22.08
Baboon	48.25	26.78	19.79	48.25	26.78	14.37
Average	48.52	34.34	26.51	48.50	33.49	22.39

### C. Image size versus Distortion

To maintain a user-friendly application, we must make sure that it can preserve the same level of performance at a distinct image size. (as shown in Figure 11) We converted the Lena images to different sizes, for example, 128 x 128, 256 x 256, 512 x 512, 768 x 768, and 1024 x 1024. We then set up the same embedding rate ( $\mathcal{R} = 0.6$ ) of the secret data. The distortion was serious when the size of the image was 128 x 128. The reason for this is that the number of pixels used to present one color is much smaller, meaning that the difference in adjacent pixels is large. In other words, as the size of the image becomes larger, the adjacent pixels increase in similarity, meaning that the difference will be much smaller. Hence, the requirement of the embedding level will be smaller and is beneficial to the image quality.

## V. CONCLUSION

In this paper, we proposed a method, Athena, which maintains the advantage of existing reversible data hiding techniques as well as improving the issue of embedding level adjustment. It can determine a better embedding level using the PPQ. Athena also provides a novel method to manage the overhead information. With a combination of overhead information and the value of the embedding level, it can generate a unique data reversion key for a corresponding marked image. The data that is required for extraction can be transmitted at the same time as the key as it provides double protection for the private message. In the future, it is expected that this technique can be deployed for communication applications, not only for copyright protection.

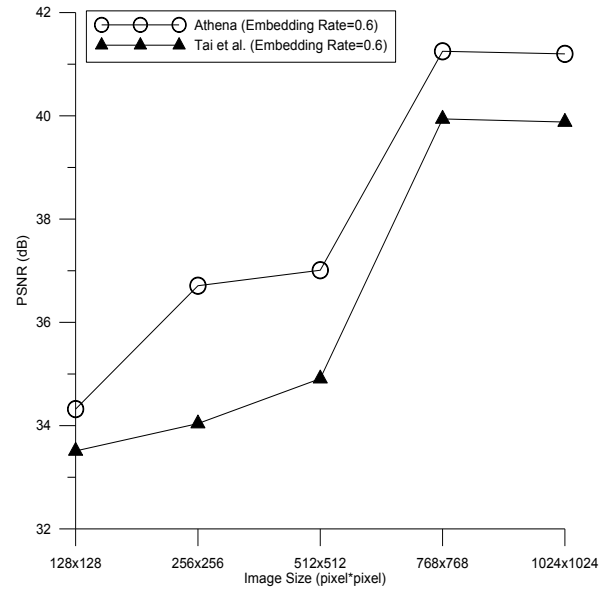


Figure 11. The image size versus PSNR.

## REFERENCES

- [1] M. Wu and B. Liu, "Data hiding in binary image for authentication and annotation," *IEEE Transactions on Multimedia*, 6(4), pp. 528-538, 2004.
- [2] P. Tsai, Y.-C. Hu and H.-L. Yeh, "Reversible image hiding scheme using predictive coding and histogram shifting," *Signal Processing*, 89(6), pp. 1129-1143, 2009.
- [3] Z. Ni, Y.-Q. Shi, N. Ansari and W. Su, "Reversible data hiding," *IEEE Transactions on Circuits and Systems for Video Technology*, 16(3), pp. 354-362, 2006.
- [4] W.-L. Tai, C.-M. Yeh and C.-C. Chang, "Reversible data hiding based on histogram modification of pixel differences," *IEEE Transactions on Circuits and Systems for Video Technology*, 19(6), pp. 906-910, 2009.
- [5] S.-K. Lee, Y.-H. Suh and Y.-S. Ho, "Reversible Image Authentication Based on Watermarking," *IEEE International Conference on Multimedia and Expo (ICME 2006)*, pp. 1321-1324, 2006.
- [6] S.-W. Jung, L. T. Ha and S.-J. Ko, "A new histogram modification based reversible data hiding algorithm considering the human visual system," *IEEE Signal Processing Letters*, 18(2), pp. 95-98, 2011.
- [7] H.-W. Tseng and C.-P. Hsieh, "Prediction-based reversible data hiding," *Information Sciences: an International Journal*, 179(14), pp. 2460-2469 2009.
- [8] X. Zeng, L. Ping and Z. Li, "Lossless Data Hiding Scheme Using Adjacent Pixel Difference Based on Scan Path," *Journal of Multimedia*, 4(3), pp. 145-152, 2009.
- [9] C.-F. Lee, H.-L. Chen and H.K. Tso, "Embedding capacity raising in reversible data hiding based on prediction of difference expansion," *Journal of Systems and Software*, 83(10), pp. 1864-1872, 2010.
- [10] R. Housley, W. Polk, W. Ford and D. Solo, "Internet X. 509 public key infrastructure certificate and certificate revocation list (CRL) profile," *RFC 3280*, 2002.
- [11] S. Harris, *CISSP All-In-One Exam Guide Exam guide*, 4 ed., McGraw-Hill Osborne Media, 2007, ISBN 0071497870.
- [12] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, 44(13), pp. 800-801, 2008.

# User Features-aware Trust Measurement of Cloud Services via Evidence Synthesis for Potential Users

Hua Ma<sup>a,b</sup>, Zhigang Hu<sup>a,\*</sup>, Liu Yang<sup>a</sup>, Tie Song<sup>a</sup>

<sup>a</sup> School of Software,  
Central South University  
Changsha, China  
zghu@csu.edu.cn

<sup>b</sup> School of Information Science and Engineering,  
Hunan International Economics University,  
Changsha, China  
hua.ma@csu.edu.cn

**Abstract**—The cloud computing paradigm can provide elastic and dynamic resources on demand, which facilitates service providers to make profits resulting from the long tail effect. So it is vitally important to ensure that cloud services can be acceptable to more and more potential users. However, it is challenging for potential users to discover the trustworthy cloud services due to the deficiency of usage experiences and the information overload of QoE evaluations from consumers. This paper proposed a user features-aware trust measurement approach for potential users. In this approach, the influence factors of QoE are systematically analyzed based on user feature model and the quantitative computation methods are designed to measure the user feature similarity. In addition, employing FAHP method identifies the user feature community. To enhance the accuracy of trust measurement, the false evidences in the QoE evaluations are iteratively filtered with dynamic mean distance threshold. Finally, the service trust is measured via evidence synthesis combining user feature similarity. The experiments show that this approach is effective to improve the quality of trust measurement, which is helpful to solve the cold start problem and data fusion problem with false evidences in cloud computing paradigm.

**Keywords**—cloud services, trust measurement, user features, potential users, evidence synthesis

## I. INTRODUCTION

Recently the cloud computing paradigm has gained enormous momentum. More and more services are published into clouds, such as cloud storage and file synchronization, virtual office, remote virtual desktop, video/audio streaming and cloud gaming [1]. The growing presence of cloud services creates new problems, and a major problem the cloud computing ecosystem faces is about the quality experienced by those using services [2]. If the *quality of experience (QoE)* does not reach expectations, users will not trust the service. Therefore, ensuring *quality of services (QoS)* close to user expectation and discover trustworthy services will be the key factors for promoting cloud computing [3]. Especially, this issue becomes more challenging for multimedia cloud services.

In the last few years, Internet companies have reaped huge profits by utilizing long tail effect [4]. Considering that the cloud computing paradigm can provide elastic and dynamic cloud resources on demand, the long tail effect would play a more important role in cloud market. Therefore, it has become

the latest trend to discover and recommend trustworthy services for potential users [5]. It faces some important challenges.

- To take full advantage of the long tail effect, it is necessary to help cloud service provider (CSP) to promote services and help potential users to find suitable services. But potential users don't know whether the *QoS* of one service is trustworthy due to the deficiency of usage experiences. So the *QoE* evaluations from consumers are beneficial for potential users [6]. Nonetheless, the information overload problem of *QoE* evaluations makes it very difficult to acquire highly valuable information about service trust.
- It is a trend to employ the personalized recommendation technologies to solve information overload problem [5]. However, most of consumers do not make evaluations usually in practice. Thus, the cold start problem will become more serious for potential users in the cloud systems. As a result, the existing methods [3][7][8][9] based on collaborative filtering algorithm (CFA) cannot provide the high-quality recommendation results.
- The cloud services are mainly provided via Internet, and the dynamic network environment cannot ensure that all users experience one service with identical *QoS*. It causes that the *QoE* evaluations from deferent users cannot be uniform. Therefore, it is a significant problem to analyze the differences of users and to identify their user features for helping potential users to obtain valuable information about the trust of cloud services.

To address these issues, this paper proposes a user features-aware trust measurement approach of cloud services via evidence synthesis. It can provide new ideas for solving the information overload problem of *QoE* evaluations and the cold start problem of trustworthy service recommendation for potential users in the cloud computing paradigm.

The rest of the paper is organized as follows. Section II introduces the related work. Section III describes a user feature model, and presents the similarity measurement method of user features. Section IV proposes the trust measurement approach for cloud service based on FAHP and evidence theory. In section V, the experiments and results are given. Finally, conclusions and discussion are given in section VI.

## II. RELATED WORK

In order to reduce or avoid the negative influence on *CSP* from service consumer with poor reputation, some researchers put forward the reputation share mechanism [10], reputation reporting mechanism [11]. In addition, for purpose of helping service consumer to select trusted service and detecting deception from *CSP* providing malicious or false services, some researchers proposed dynamic scheduling model based on MAS [12], heuristic algorithm [13] and so on. However, *CSP* usually provides several types of cloud services, and it is improper to measure service trust with entity trust of *CSP*. Sherchan et al. [14] proposed a fuzzy approach to analyze user rating behavior, detect deception and identify user preferences. However, the inference rules will become very complex when there are many attributes of *QoS*.

It's a main research approach to recommend the trusted cloud services based on *CFA*. Zheng et al. [8] employed *CFA* to predict the reliability of service, computing user similarity and item similarity based on *PCC* (Pearson correlation coefficient). However, the *PCC*-based similarity results probably show apparent linear dependence of two users although their absolute scores are totally different [4]. Chen et al. [7] used a region model to study service recommendation based on hybrid *CFA*. However, in practice, two physical locations are not necessarily near each other even though their IP addresses are quite similar. Tang et al. [9] presented a location-aware prediction method of the quality of service based on autonomous system (AS), employing *CFA* to recommend services for users in same AS. But it is very probable that two locations even in the same AS are geographically far away. Rosaci et al. [3] proposed an agent-based architecture to compute recommendations of multimedia web services by employing the content-based recommendations and *CFA*. However, this paper mainly focuses on experiment analysis for lack of theoretical basis.

Some methods are usually used to solve the cold start problem, such as simple average method [8] and expectation maximization algorithm [15]. These existing methods cannot provide the high-quality recommendation result. Recently the popularity and entropy are employed to improve recommendation quality [16][17]. In addition, some information from users themselves can be used to identify preferences, such as sex, age, education background and interest [4]. Nori et al. [18] argued that users' behaviors in one system can be used to predict their behaviors in another system.

In order to obtain valuable *QoE* evaluations about service trust, it is very important to identify the influence factors of *QoE*. Rosaci et al. [3] argued that the recommendation of multimedia web services should take into account the effect of the device exploited by the user. On the basis of experimental observation, Zheng et al. [8] indicated that users from different countries might have quite different *QoE* on the same web services, influenced by the network connections. Casas et al. [2] discussed the impact of network *QoS* features, including *RTT*, bandwidth, based on lab experiment and field trial experiments. Lin et al. [19] discussed the evaluation model of *QoE*, and argued that the influence factors of *QoE* consist of services factors, environment factors and user factors.

## III. USER FEATURES MODEL

*QoE* is influenced by many factors, which can be summarized as subjective factors and objective factors. In this paper these factors are collectively called user feature. And the user feature model is defined to analyze the relationships between related concepts.

### A. Structure of User Features Model

The user feature model is shown as Fig. 1.

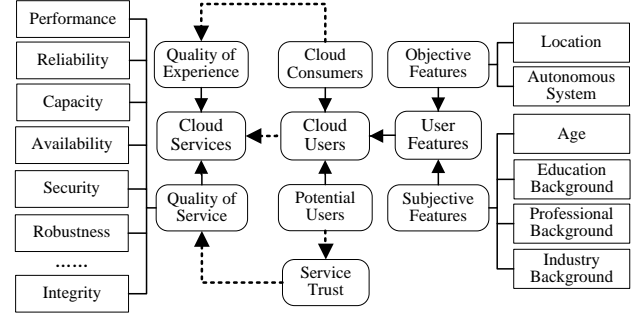


Figure 1. User features model.

The key concepts are described as follows: (a) Cloud consumer (*cc*): it refers to users who paid one specified service and used it. *cc* may make *QoE* evaluations of cloud service based on his/her usage experiences. (b) Potential user (*pu*): it refers to users who have not used directly one specified cloud service. (c) Cloud user (*cu*): it refers to the collection of *cc* and *pu*. (d) Quality of service (*QoS*): According to W3C Working Group Note [20], the *QoS* requirements include performance, reliability, integrity, and so on, which can be measured with some parameters. For example, performance can be measured by response time, throughput and latency. (e) Quality of experience (*QoE*): it refers to the overall acceptability of a service subjectively by *cc*. *QoE* is closely related to *QoS* and the user expectations. (f) Service trust: it refers to the *pu*'s confidence degree to *QoS* claimed by *CSP*, which is uncertain influenced by some objective factors and subjective factors. (g) User feature: it influences greatly the *QoE* of *cc*, consisting of objective features and subjective features.

The objective features include location and network autonomous system (AS). The location feature recognizes the service level of local ISP (Internet service provider) and administrative controls condition of local government. The AS feature concerns the routing condition and communication quality of network. The subjective features include age, professional background, education background and industry background. These subjective features can influence people's expectation and evaluation criterion deeply [19], and lead to the differences of *QoE* evaluations between users for same thing.

If the user features of both users are similar, then the users will tend to reach common understandings on *QoS*, as well as give similar *QoE* evaluations. That is, if *cc* is similar with *pu*, then *cc* can provide valuable information about the service trust to *pu*. In practice, *pu* prefers to learn the service trust probably experienced by themselves in the future. Therefore, it's crucial for cloud system to compute user feature similarity between users and to help *pu* to measure the service trust accurately.



### B. Similarity Measurement of User Features

The location feature can be noted as a five-tuple:  $fl = (\text{country, state/province, city, county or district, subdistrict})$ . It's not necessary to use all five elements to describe the location.  $fl_j^i$  represents the  $j$ -th location information of the  $i$ -th user. The binary location feature coding as  $loc = b_1 b_2 \dots b_{n-1} b_n$  is proposed to measure location similarity between users.  $b_1$  represents the highest level of administrative unit, and  $b_n$  represents the lowest level of it. Comparing the location features of  $cc$  with  $pu$ ,  $loc$  of every  $cc$  can be computed. Assume the location code of  $pu$  is  $loc^{pu} = \underbrace{11\dots1}_n$ , and the code of  $cc$  can be obtained as (1).

$$loc_i^{cc} = \begin{cases} 1, & fl_i^{pu} = fl_i^{cc} \\ 0, & fl_i^{pu} \neq fl_i^{cc} \end{cases} \quad (1)$$

where  $fl_i^{pu}$  and  $fl_i^{cc}$  represent the feature values of the  $i$ -th location information of  $pu$  and  $cc$  respectively.  $\sum_{j=1}^n loc_j^{cc} = 0$  if  $loc_i^{cc} = 0$ . The location code of  $cc$  can be normalized as (2).

$$s^{loc} = (b_1 b_2 \dots b_{n-1} b_n)_2 / (2^n - 1) \quad (2)$$

Let  $as^{pu}$  and  $as^{cc}$  represent AS number of  $pu$  and  $cc$  respectively,  $0 \leq as^{pu}, as^{cc} \leq 2^{32} - 1$ , AS similarity is computed with (3). Let  $a^{pu}$  and  $a^{cc}$  represent the age of  $pu$  and  $cc$  respectively, the age similarity is computed as (4).

$$s^{as} = \begin{cases} 1, & as^{pu} = as^{cc} \\ 0, & as^{pu} \neq as^{cc} \end{cases} \quad (3)$$

$$s^a = \begin{cases} 1 - |a^{pu} - a^{cc}| / a^{pu}, & |a^{pu} - a^{cc}| \leq a^{pu} \\ 0, & |a^{pu} - a^{cc}| > a^{pu} \end{cases} \quad (4)$$

The feature set of education background is defined as  $fe = \{e_1, e_2, \dots, e_{n-1}, e_n\}$ . The similarity of  $e_i$  and  $e_j$  can be measured with the relevance metric as (5). In (5),  $0 \leq e_{ij} \leq 1$ ,  $e_{ij}$  represents the similarity of the  $i$ -th education background and the  $j$ -th one. Obviously,  $e_{ij} = e_{ji}$ .

$$S^e = \begin{pmatrix} e_{11} & e_{12} & \dots & e_{1n} \\ e_{21} & e_{22} & \dots & e_{2n} \\ \vdots & \dots & e_{ij} & \vdots \\ e_{n1} & e_{n2} & \dots & e_{nn} \end{pmatrix} \quad (5)$$

The feature set of professional background is defined as  $fp = \{p_1, p_2, \dots, p_{n-1}, p_n\}$ . The feature set of industry background is defined as  $fi = \{i_1, i_2, \dots, i_{n-1}, i_n\}$ . According to international standard [21], the industry types include 21 categories. The feature similarity between various professional background and the feature similarity between various industries can be also measured with the relevance metric as (4).

## IV. USER FEATURES-AWARE TRUST MEASUREMENT

### A. Computing Weights of User Features based on FAHP

Even for one user, the weight of each user feature varies widely with application scenarios. Therefore, it's not appropriate to synthesize the similarity values of user features with weighted mean method. And the fuzzy analytic hierarchy process (FAHP) method [22] can overcome the shortcomings of AHP, which is suitable for solving the multiple attribute decision making problems. Therefore, FAHP method is used to compute the weights of user features.

Supposing  $B = (b_{ij})_{n \times n}$  is a fuzzy judgment matrix with  $0 \leq b_{ij} \leq 1$ ,  $n$  is the number of user features, and  $b_{ij}$  is the importance ratio of the  $i$ -th user feature and the  $j$ -th one. If  $b_{ij} + b_{ji} = 1$  and  $b_{ii} = 0.5$ ,  $B$  is a fuzzy complementary judgment matrix. Giving an integer  $k$ , if  $b_{ij} = b_{ik} - b_{jk} + 0.5$ ,  $B$  is a fuzzy consistency matrix. Firstly, for transforming  $B$  into a fuzzy complementary judgment matrix, the sum of each row of this matrix is defined as  $b_i$ , and mathematical manipulation is performed with (6)[23]. Secondly, a new fuzzy matrix  $C = (c_{ij})_{n \times n}$  can be obtained, which is a fuzzy consistency judgment matrix. Thirdly, the sum of each row is computed and standardized. Finally, the weight vector is calculated by (7).

$$c_{ij} = 0.5 + (b_i - b_j) / 2(n - 1) \quad (6)$$

$$w_i = \frac{1}{n(n-1)} \sum_{j=1}^n c_{ij} + \frac{n}{2} - 1 \quad (7)$$

The user feature similarity is defined as matrix  $S$  by (8).

$$S = (s^{loc} \ s^{as} \ s^a \ s^e \ s^p \ s^i)^T = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1k} \\ s_{21} & s_{22} & \dots & s_{2k} \\ \vdots & \dots & s_{ij} & \vdots \\ s_{61} & s_{62} & \dots & s_{6k} \end{pmatrix} \quad (8)$$

In  $S$ ,  $s_{ij}$  represents the similarity value of the  $i$ -th feature for the  $j$ -th consumer. The comprehensive value of user feature similarity for the  $j$ -th consumer is computed by (9).

$$sim_j = \sum_{i=1}^n s_{ij} \times w_i \quad (9)$$

The similarity threshold is noted as  $s^{th}$ . If  $sim_j$  is less than  $s^{th}$ , the  $j$ -th consumer cannot provide valuable evaluations. The user feature community is denoted as  $UFC = \{u_1, u_2, \dots, u_z\}$ , and  $sim_j$  of  $u_j$  is bigger than  $s^{th}$ . These cloud consumers in  $UFC$  can provide reliable information about service trust.

### B. Computing Service Trust via Evidence Synthesis

Influenced by the dynamic Internet and subjective factors of consumers, the trust evaluations of cloud service are uncertain. And evidence theory has unique advantages in expression of uncertainty, and has been widely used in expert system and multiple attribute decision making fields [24]. Thus,

evidence theory can be employed to synthesize the fuzzy evaluations from user feature community.

The fuzzy trust evaluation set is defined as  $VS=\{vt, vl\}$ , which describes the evaluation results of service trust.  $vt$  represents trust,  $vl$  represents distrust. And fuzzy evaluation  $v=(v^t, v^l)$  is the fuzzy subset of  $VS$  with  $v^t+v^l=1$ . For example, the fuzzy evaluation of service given by a cloud consumer is  $v^t=(0.91, 0.09)$ , indicating that the consumer thinks the trust degree of this service is 0.91 and the distrust degree is 0.09. Denote the identification framework as  $\Theta=\{T, F\}$ , where  $T$  represents service is trustworthy, and  $F$  represents it's trustless.  $\Theta$  is mapped to  $VS$ . The power set of  $\Theta$  is  $2^\Theta=\{\Phi, \{T\}, \{F\}, \Theta\}$ . And the basic trust distribution function  $m$  is defined as a mapping from  $2^\Theta$  to  $[0, 1]$  with  $m(\varphi)=0$  and  $\sum_{A \subseteq \Theta} m(A)=1$ .  $m$

can be measured by the trust evaluation.

Due to the possibilities of evaluation forgery and network anomaly, there might be a few false evidences in trust evaluations, which will lead to the poor evidence synthesis result. So it's vital to filter false evidences for ensuring the accuracy of data synthesis. Suppose the basic trust distribution function of evidence  $E_1$  and  $E_2$  are  $m_1$  and  $m_2$  respectively, and the focal elements are  $A_i$  and  $B_j$  respectively. Equation (10) is used to calculate the distance between  $m_1$  and  $m_2$ .

$$d(m_1, m_2) = \sqrt{\frac{1}{2} (\|m_1\|^2 + \|m_2\|^2 - 2\langle m_1, m_2 \rangle)} \quad (10)$$

The distances between evidences are small if they support each other, and the distances will become large if there are some false evidence. Therefore, the false evidences can be identified according to the mean distance of evidence. Suppose  $\bar{d}_i$  represents the mean evidence distance of between  $i$ -th evidence and other  $n-1$  evidences.

Some researches [25] used static mean distance threshold to filter false evidence, which cannot adapt to different situations. This paper proposes a dynamic function to create the mean distance threshold, and employ iteration filtering to improve the accuracy of filtering. The function is shown as (11).

$$\alpha = \frac{1}{n} (1 + \beta) \times \sum_{i=1}^n \bar{d}_i \quad (11)$$

where  $\beta$  represents the threshold coefficient. The ideal value rang of  $\beta$  is from 0.05 to 0.30.  $\beta$  should be a greater value if the distances between evidences are quite large. The distance threshold is adaptable because it is obtained based on the mean distances of all evidences, which ensure that every iteration only filters those most likely to be false evidences.

In practice, a rational distance between evidences should be allowed. The lower limit of mean distance is denoted as  $\zeta$ . The filtering operation is executed when  $\alpha > \zeta$ , and the  $i$ -th evidence will be removed if  $\bar{d}_i \geq \alpha$ . The operation continues until  $\alpha \leq \zeta$ . The remaining evidences are viewed as real evidences. The set of users providing real evidences is defined as  $Ref$ . Considering the evidences about trust evaluations are interrelated [26], they cannot be synthesized directly with D-S method, unless the condition of idempotence is met [27]. In this paper, an evidence fusion method with user feature weights is proposed as (12).

$$m(A) = m_1(A) \oplus m_2(A) \cdots \oplus m_{|FC|}(A) = \sum_{i=1}^{|Ref|} m_i(A) \bullet fw_i \quad (12)$$

$$fw_i = \frac{1}{1 - sim_i} \times \frac{1}{\sum_{j=1}^{|Ref|} \frac{1}{1 - sim_j}}$$

where  $fw_i$  is the feature weight of the  $i$ -th evidence, and represents the importance of the  $i$ -th evidence for potential user. According to (12), the synthesis result of interrelated evidences satisfies the idempotence. Considering that the user feature similarity is introduced into, the evidence synthesis can provide highly valuable reference for potential user.

## V. EXPERIMENTS

In this section, some experiments are carried out to verify the effectiveness of our approach.

### A. Experiment Setup

We used the WS-DREAM datasets [28], which collected the real-world *QoS* evaluation results from Planet-Lab. Considering the deficiency of user feature information, we extended WS-DREAM and supplemented some real data about user feature by using QueryIP services [29] and Hurricane Electric Internet services [30]. These data includes AS number, city, network description or area, and so on. And the detailed raw data used by our experiments are provided online [31]. The extended user information is shown as TABLE I.

TABLE I. USER INFORMATION IN EXTENDED WS-DREAM DATASET

UID	IP Address	Country	City	Network Description/Area	AS Number	Industry background
0	12.108.127.138	United States	Pittsburgh	AT&T Services, Inc.	AS7018	3
2	122.1.115.91	Japan	Hamamatsu	NTT Communications Corp.	AS4713	3
3	128.10.19.52	United States	West Lafayette	Purdue University	AS17	1
...	...	...	...	...	...	...

In TABLE I, according to network description, the industry background can be identified. Thus, a three-level location code, consisting of country, city and area, can be created. Considering the users from Planet-lab have the similar features

in age, education background, professional background, these features are assigned with small weighted values in FAHP. WS-DREAM provides response time matrix of service invocations. Considering the timeout value denoted with -1

cannot be used to measure service trust, Equation (13) is employed to transform them into positive numbers.

$$rt^{time-out} = p \times \left( \left( \sum_{i=1}^n rt_i \right) + k \right) \frac{1}{n-k} \quad (13)$$

where  $n$  is the total number of users;  $k$  represents the number of users who experience timeout of service;  $rt_i$  represents the response time experienced by user  $i$ ;  $p$  represents the penalty index with  $p \geq 10$ . In addition, the response time is viewed as the important indicator to measure *QoE*. Comparing the response time provided by consumers with user expectation value, the trust degree of *QoS* is calculated with (14).

$$v'_i = \begin{cases} 1 & , rt_i < \xi \\ 1 - \delta \times (rt_i - \xi) / \xi, \xi \leq rt_i \leq \xi(1 + \delta) / \delta & \\ 0 & , \xi(1 + \delta) / \delta < rt_i \text{ or } rt_i = -1 \end{cases} \quad (14)$$

where  $\xi$  is the user's expectation value of the response time. If  $rt_i \leq \xi$ , the  $i$ -th user thinks this service completely trustworthy.  $\delta$  is the adjustment factor, which determinates the range of response time. According to the 2-5-10 principles [32] of response time in software testing analysis,  $\xi = 2s$ ,  $\delta = 0.25$ .

### B. Experiment Result and Analysis

To verify our trust measurement method via evidence synthesis based on user feature weights named as UFWM, we establish five experiments scenarios, and compare UFWM with other methods, including Hybrid [8], distance-based weights method [33] denoted as DWM, AS distance weights method [9]

denoted as ASDWM. These scenarios are as follows: (a) Scenario 1#: no identifying user features; no filtering false evidences; (b) Scenario 2#: identifying user features; no filtering false evidences; (c) Scenario 3#: filtering false evidences based on static mean distance threshold; no identifying user features; (d) Scenario 4#: filtering false evidences based on dynamic mean distance threshold; no identifying user features; (e) Scenario 5#: identifying user features; filtering false evidences based on dynamic mean distance threshold. And the parameters are set as  $p=10$ ,  $s^{th}=0.70$  and  $\zeta=0.2$ .

#### 1) Comparison study based on one specified service

The experiments are performed based on the service 645# and the potential user coming from Technical University of Berlin in AS680, and the results are shown as Fig. 2.

Comparing Fig. 2 (a) with Fig. 2 (b), the analysis is given as follows: (a) As the number of evaluation declines sharply in scenario 2#, it is inevitable that the negative effects caused by false evidences will be enhanced. Thus, all measurement values of four methods drop markedly. (b) DWM always gets the greatest value of data fusion because evidence weight is proportional to the distance of evidence. Then, if most of evidences are true, these evidences can weaken the effects of false evidences. (c) ASDWM obtains the same results in two scenarios. The reason is that ASDWM only uses data provided by those from AS680. (d) In scenario 2#, the performance of UFWM is poor. The reason is possibly that users highly similar with potential user provided false evidences. As a result, it is important to filter false evidences to improve the quality of trust measurement.

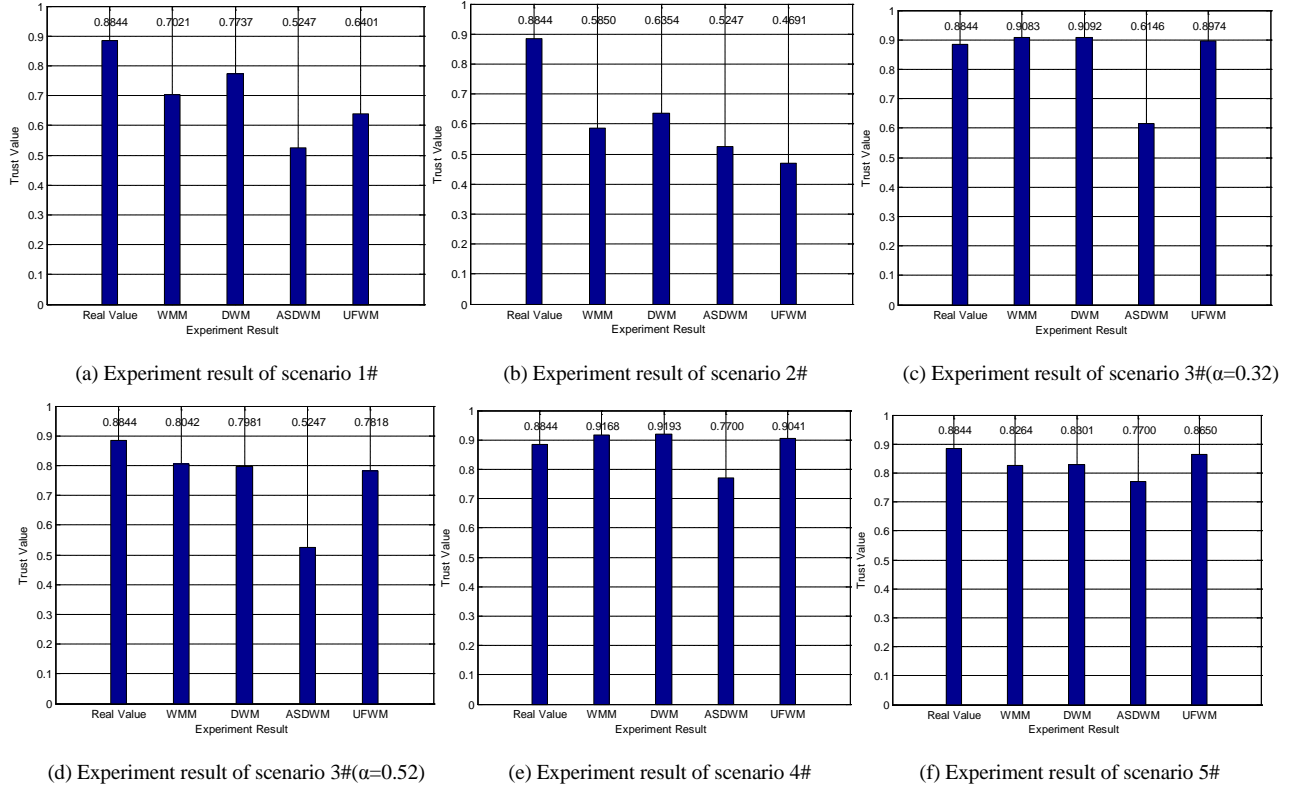


Figure 2. Comparing experiment results in deferent scenarios..

In scenario 3#, the results are shown as Fig. 2 (c) and Fig. 2 (d) when  $\alpha=0.32$  and  $\alpha=0.62$ . Obviously, the qualities of data fusion are quite different. The analysis is given as follows: (a) Most of evidences will be mistakenly identified as false evidences if  $\alpha$  is given a smaller decimal, which must lead to lower precision ratio of false evidences. According to Fig. 2 (c), WMM, DWM and UFWM obtained better quality of trust measurement in contrast to scenario 1# and scenario 2#. The reason is that many small trust evaluation values have been filtered. (b) Only a few false evidences can be identified if  $\alpha$  is given a greater decimal, which must lead to the lower recall ratio of false evidences. Therefore, It is very difficult to assign appropriate value to static mean distance threshold.

In scenario 4#, WMM, DWM and UFWM identified 109 false evidences, and ASDWM found 10 false evidences. Thus, compared with scenario 3#, scenario 4# can make four methods provide data fusion value closer to real value shown as Fig. 2 (e). The reason is that many evidences provided by those consumers not very similar with potential user are synthesized indistinguishably in scenario 4#, and these evidences just believe this service is trustworthy. Likewise, the result of

calculation will become lower when there are many poor ratings. And the result given by ASDWM is just an example.

In scenario 5#, the user feature community is identified including 37 users. After multiple iterations, WMM, DWM and UFWM found 15 false evidences, and ASDWM identified 10 false evidences. Thus, scenario 5# reflects the important significance of user features. According to Fig. 2 (f), the trust measurement result based on user feature community is closer to real value after filtering false evidences, and UFWM can provide the best result in contrast to the other three methods.

## 2) Comparison study based on all services

The following experiments are performed based on all 5825 services from dataset. In experiments, the potential user is selected randomly from AS680, and AS680 has the most users in WS-DREAM. And 12 independent experiments are performed in each scenario. The first experiment selects service 1# to service 500#, and the next experiment will continue to add another 500 services until all services have been used. The mean absolute error (MAE) metrics are employed to measure the accuracy of approaches. And the results are shown as Fig. 3.

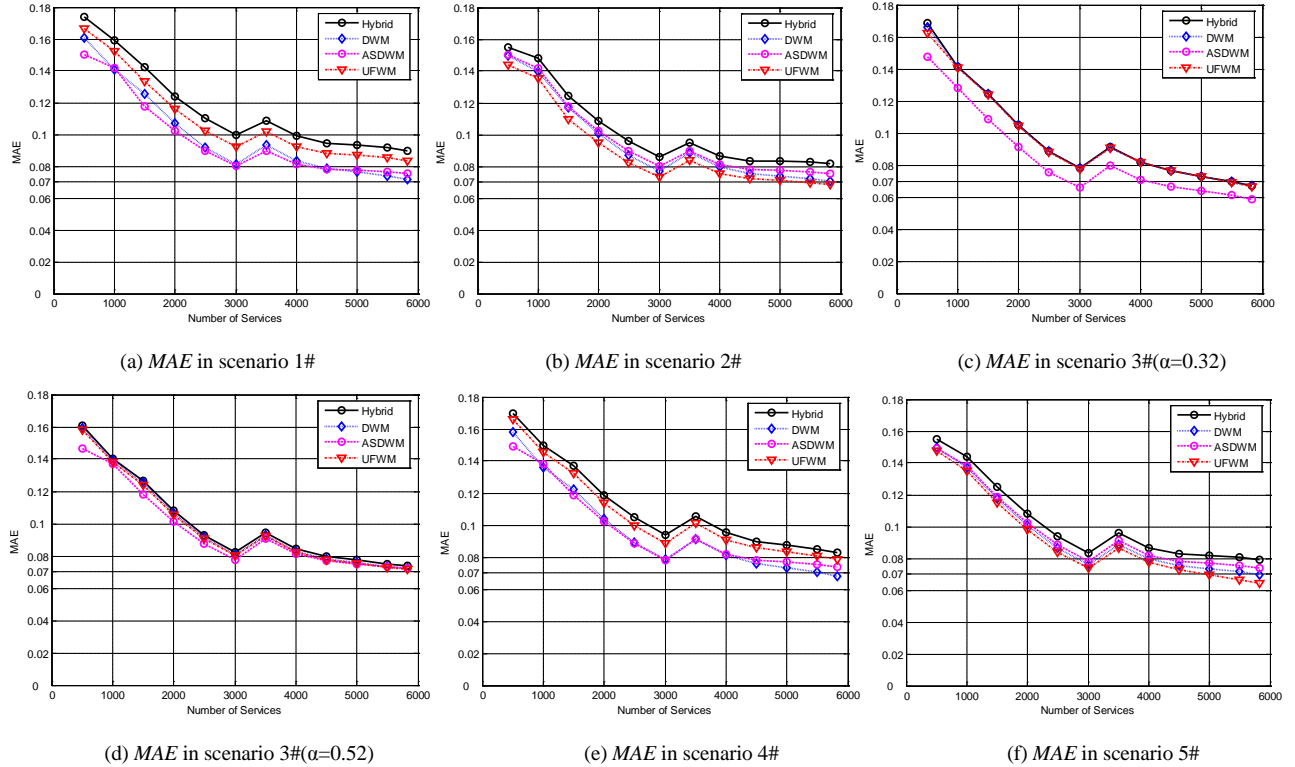


Figure 3. Comparing MAE in different scenarios.

According to Fig. 3, MAEs reduce gradually along with the more and more services used in experiments. However, MAEs show a trend of stable increase after service 3000# to service 3500# are used because the number of service timeout increases sharply. Compared with Fig. 2, Fig. 3 reflects the similar facts. The analysis is as follows: (a) the MAE values of four methods in scenario 1# and scenario 2# are greater than other scenarios due to the influence of false evidences. (b) In scenario 3#, the experiment will obtain different result when

the static distance threshold is given different value. For this dataset,  $\alpha=0.32$  can achieve good performance, while it isn't suitable to other datasets. (c) In scenario 4#, the performance of ASDWM is best because there are exactly enough consumers in AS680 network. In practice, most of AS in WS-DREAM have few users. Thus, ASDWM is unable to provide good performance for all users. (d) All of four methods gained good performance in scenario 5# because user feature community was identified and false evidences were filtered. Especially, for

the consideration of user feature weights, UFWM obtained the best result.

## VI. CONCLUSIONS AND DISCUSSION

In cloud computing diagram, the long tail effect can help cloud service providers to reap huge profits. So it is important to measure the trust degree and find the trustworthy services for potential users. This paper proposed a user features-aware service trust measurement approach via evidence synthesis. The main contributions of this paper are the following:

- The user features consisting of objective factors and subjective factors are analyzed systematically, and the quantitative similarity measurement methods are proposed. In existing researches, the impacts of objective factors, such as network feature and location feature, are studied based on lab experiments or field trial experiments, and the subjective factors related to cloud users themselves aren't considered. Meanwhile, the location feature only focuses on country level. This paper designed a binary feature coding method to support multiple-level location feature, which can more accurately measure the differences between users.
- For ensuring the accuracy of trust measurement, the user feature community is identified based on FAHP method, and the false evidences are iteratively filtered by utilizing dynamic mean distance threshold, and the service trust is measured via evidence synthesis method based on user feature weights. Those users who have the similar user features probably experience uniform *QoS*, so consumers from the user feature community can provide more valuable *QoE* evaluations for potential user. In addition, the evidence distance within a limited range is rational, but a few exceptions, namely false evidences caused by occasional anomaly, should be ignored.

The experiments analysis showed that our approach is effective to solve cold start problem and to synthesize trust evaluation with false evidences, which can provide high-quality measurement result of service trust. How to achieve personalized recommendation and establish incentive mechanism of trust evaluation will be studied in the future.

## ACKNOWLEDGMENT

This research is supported by National Natural Science Foundation of China (No.61272148, No.61301136) and Science and Technology Project of Hunan Province (No.2014FJ3122).

## REFERENCES

- [1] M. Jarschel, D. Schlosser, S. Scheuring, et al, Gaming in the clouds: QoE and the users perspective, *Mathematical and Computer Modelling*, 57 (2013), pp. 2883-2894.
- [2] P. Casas, R. Schatz, Quality of experience in cloud services: survey and measurements. *Computer Networks*, 68 (2014), pp. 149-165.
- [3] D. Rosaci, G. M. L. Sarne, Recommending multimedia web services in a multi-device environment, *Information Systems*, 38 (2013), pp. 198-212.
- [4] D. Jannach, M. Zanker, A. Felfernig, et al., *Recommender systems: an introduction*, Cambridge University Press, United Kingdom, 2011.
- [5] LY Lüa, M. Medob, CH Yeung. Recommender systems, *Physics Reports*, 519(1) (2012), pp. 1-49.
- [6] YM Lin, XL Wang, TZ Zhou. Survey on quality evaluation and control of online reviews. *Journal of Software*, 25(3) (2014), pp.506-527,
- [7] X. Chen, X. Liu, Z. Huang, et al., RegionKNN: a scalable hybrid collaborative filtering algorithm for personalized web service recommendation, in: *Proceedings of the IEEE Int'l Conf. on Web Services*, 2010: 9-16.
- [8] ZB Zheng, Michael R. Lyu, Collaborative reliability prediction for service-oriented systems, in: *Proceedings of the ACM/IEEE 32<sup>nd</sup> Int'l Conf. on Software Engineering*, 2010, pp. 35-44.
- [9] MD Tang, YC Jiang, JX Liu, User location-aware web services QoS prediction, *Journal of Chinese Computer Systems*, 33(12) (2012), pp. 2664-2668.
- [10] Michael Armbrust, Armando Fox, Rean Griffith, et al., A view of cloud computing, *Communications of the ACM*, 53(4) (2010), pp.50-58.
- [11] P. Wang, SY Zhang, XJ Chen, A novel reputation reporting mechanism based on cloud model and gray system theory, *International Journal of Advancements in Computing Technology*, 3(10) (2011), pp. 75-84.
- [12] WJ Jiang, LM Zhang, P. Wang, Dynamic scheduling model of computing resource based on MAS cooperation mechanism, *Science in China Series F-Information Sciences*, 52(8) (2009), pp. 1302-1320.
- [13] Seog-Chan Oh, Dongwon Lee, Soundar R.T. Kumara, Effective web service composition in diverse and large-scale service networks, *IEEE Transactions on Services Computing*, 1(1) (2008), pp. 15-32.
- [14] W. Sherchan, S. W. Loke, S. Krishnaswamy, A fuzzy model for reasoning about reputation in web services, in: *Proceedings of the ACM Symposium on Applied Computing*, 2006, pp.1886-1892.
- [15] XH Sun, FS Kong, S. Ye, A comparison of several algorithms of collaborative filtering in startup stage, in: *Proceedings of the IEEE Int'l Conf. on Networking, Sensing and Controlling*, 2005, pp. 25-28.
- [16] K. Yu, A Schwaighofer, V Tresp, et al., Probabilistic memory-based collaborative filtering, *IEEE Transactions on Knowledge and Data Engineering*, 16(1) (2004), pp. 56-69.
- [17] AM Rashid, I. Albert, D Cosley, et al., Getting to know you: learning new user preferences in recommender systems, In: *Proceedings of the Int'l Conf. on Intelligent User Interfaces*, 2002, pp. 127-134.
- [18] N. Nori, D. Bollegala, M. Ishizuka, Exploiting user interest on social media for aggregating diverse data and predicting interest, *Artificial Intelligence* 109(B3) (2011), pp. 241-248.
- [19] C. Lin, J. Hu, XZ Kong, Survey on models and evaluation of quality of experience, *Chinese Journal of Computers*, 35(1) (2012), pp.1-15.
- [20] W3C, QoS for web services: requirements and possible approaches, <http://www.w3c.or.kr/kr-office/TR/2003/ws-qos/>.
- [21] United Nations, International Standard Industrial Classification of All Economic Activities (ISIC), Rev4, <http://laborsta.ilo.org/applv8/data/isc3e.html>.
- [22] L. Mikhailov, P Tsvetnikov, Evaluation of services using a fuzzy analytic hierarchy process, *Applied Soft Computing*, 5(1) (2004), pp. 23-33.
- [23] ZS XU, Two Methods of Maximizing Deviations of Multi-attribute Decision Making, *Journal of Industrial Engineering and Engineering Management*, 15(2) (2001), pp. 21-29.
- [24] S.B. Amor, K. Jabeur, J. Martel M., Multiple criteria aggregation Procedure for mixed evaluations, *European Journal of Operational Research*, 181(3) (2007), pp. 1506-1515.
- [25] W. Jiang, An Zhang, Selecting false evidence in information fusion, *Computer Engineering and Application*, 44(33) (2008), pp. 138-140.
- [26] SL Yang, WD Zhu, ML Ren, Combination theory and method for interrelated evidences based optimal adjustment coefficient, *Journal of Management Sciences in China*, 6(5) (2003), pp. 12-16.
- [27] T. Denoeux, Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence, *Artificial Intelligence*, 172(2-3) (2008), pp. 234-264.
- [28] ZB Zheng, H Ma, MR. Lyu, et al., Collaborative web service QoS prediction via neighborhood integrated matrix factorization, *IEEE Transactions on Services Computing*, 6(3) (2013), pp. 289-299.
- [29] BAJB.Net. QueryIP, <http://www.query-ip.com/>.
- [30] Hurricane Electric. Hurricane Electric Internet Service, <http://bgp.he.net/>.
- [31] Extended WSDream-QoS Dataset, <http://pan.baidu.com/s/1kTwqRx9>.
- [32] databaseskill.com, LoadRunner did not tell you, <http://www.DatabaseSkill.com/471323/>.
- [33] YN Li, XQ Qiao, XF Li, An uncertain context ontology modeling and reasoning approach based on D-S theory, *Journal of Electronics & Information Technology*, 32(8) (2010), pp. 1806-1811.

# Efficiency of Hybrid Index Structures - Theoretical Analysis and a Practical Application

Richard Göbel, Carsten Kropf, Sven Müller

Institute of Information Systems  
University of Applied Sciences Hof  
Hof, Germany

{richard.goebel, carsten.kropf, sven.mueller}@iisys.de

**Abstract**— Hybrid index structures support access to heterogeneous data types in multiple columns. Several experiments confirm the improved efficiency of these hybrid access structures. Yet, very little is known about the worst case time and space complexity of them. This paper aims to close this gap by introducing a theoretical framework supporting the analysis of hybrid index structures. This framework then is used to derive the constraints for an access structure which is both time and space efficient. An access structure based on a B+-Tree augmented with bit lists representing sets of terms from texts is the outcome of the analysis which is then validated experimentally together with a hybrid R-Tree variant to show a logarithmic search time complexity.

**Keywords**—hybrid index structures, theoretical analysis, experimental validation

## I. INTRODUCTION

Modern database systems often manage data of multimedia types. Texts, images or video data are stored inside those database systems. Some approaches with specialized database systems which allow storing and retrieving those data fast exist. Relational database management systems are still the most used technique, especially as data stores in enterprises, although NoSQL databases are also present. Mixing up different storage systems does not help in retrieving the data fast, because of having to search multiple systems and generating a finally intersected result set at the end. This implies, on the one hand, a large overhead of temporarily allocated memory and, on the other, a large overhead of time as the distinct search results must be combined to a final result set.

Most existing hybrid access structures focus on the efficient storage and retrieval of data composed by textual and geographical data. In this paper, we focus on a probably more common scenario of data consisting of texts and conventional relational (single-valued) data sets. For this purpose the access structure is based on a conventional B+-Tree augmented by bit lists for indicating the presence of terms below a node. Besides this structure, also an R-Tree based one is evaluated.

Although several of these hybrid approaches with the ability to index data of this mixed type are present, there is no evidence about the temporal and spatial worst case complexity.

The major contributions of this work are a theoretical basis to analyse hybrid access structures, an in-depth analysis of index structures leading to the theoretical construction of a hy-

brid index structure and an analysis of asymptotic time and space complexity (see section III).

Finally, a practical construction and evaluation of the previously analysed hybrid index structure with focus towards the theoretical analysis (see section IV) is carried out. Based on a lack of space, related work is only discussed shortly.

## II. RELATED WORK

New hybrid indexing strategies, enhancements, variations and compositions of existing concepts, like the B-Tree [1] or the R-Tree [2] have been proposed to address performance issues on heterogeneous data. Approaches are present treating terms differently according to the occurrence frequency like [3]. Also a couple of different hybrid indexing methods or methods for management of data in hybrid data spaces like [4] exists. Approaches like [5] (KR\*-tree), [6] ((M)IR<sup>2</sup>-Tree) or [7] (bR\*-tree) investigate, among others, the use of hybrid index structures combining textual and spatial retrieval utilizing the R-Tree [2] or its variants (e.g. R\*-Tree [8]) which augment the R-Tree with certain secondary structures (bitlists or inverted lists) to enable set annotations at R\*-Tree elements. Approaches like [9], [10] or [11] represent hybrid index structures for textual and spatial types which differentiate the treatment of textual entries based on the relative or absolute term frequency.

## III. ANALYSIS OF ACCESS STRUCTURES

This section analyses the worst case time and space complexity of hybrid access structures. For this purpose we will introduce a formal notation as a general basis for analysing non-trivial access structures.

The general idea to formalise an index structure is similar to the work of Hellerstein et al. (e.g. [12]). The differences in our approach are motivated by the fact that this paper deduces upper bounds for search time complexity instead of lower bounds as in [12].

We will also show that a hybrid tree providing information about both single-valued and multi-valued columns in the upper nodes of the primary tree structure ensures a time complexity of  $\mathcal{O}(\log(n) \cdot m)$  and a space complexity of  $\mathcal{O}(n)$ . The upper nodes of the tree only have to be sorted according to the single-valued column.



### A. Basic Definitions

For analysing the access structure we will consider a database table with “normalised” columns containing single values and “non-normalised” columns with multiple values. Although most of this analysis is more generic we will assume that a multi-valued column contains a set of terms. We will denote the set of entries for such a table by a capital  $E$  and individual entries by  $e$ :

$$E = \{e_1, \dots, e_n\} \quad (1)$$

For reasons of simplicity we assume a single set of values  $V$  for all columns. A set of  $k$  projection functions  $p_i$  retrieves the values of the  $k$  individual columns:

$$p_i : E \rightarrow 2^V \text{ with } i = 1, \dots, k \quad (2)$$

Projection functions may also be applied to sets of entries:

$$p_i(\{e_1, \dots, e_j\}) = p_i(e_1) \cup \dots \cup p_i(e_j) \quad (3)$$

Single-valued (normalised) columns contain not more than one value per entry:

$$\forall e \in E : |p_i(e)| \leq 1 \quad (4)$$

With  $q_i$  we denote the intersections between the sets of values in the related column:

$$q_i(\{e_1, \dots, e_j\}) = p_i(e_1) \cap \dots \cap p_i(e_j) \quad (5)$$

The key idea of many access structures  $\mathcal{A}$  is the assignment of entries to groups which are not necessarily disjoint:

$$\mathcal{A} = \{N_1, \dots, N_p\} \text{ with } N_1, \dots, N_p \subseteq E \quad (6)$$

With this approach not all of these groups need to be searched. For this purpose, each group  $N$  usually corresponds to a value  $v$  which occurs in all entries of  $N$  at the related column  $i$ :

$$\forall N \in \mathcal{A} \exists v \in V : v \in q_i(N) \quad (7)$$

It is also important that each group  $N$  provides every entry  $e$  which contains the corresponding value  $v \in q_i(N)$  in the related column:

$$\begin{aligned} \forall v \in V, e \in E, N \in \mathcal{A} : \\ v \in p_i(e) \wedge v \in q_i(N) \Rightarrow e \in N \end{aligned} \quad (8)$$

An index structure with this definition is usually called an inverted index. This definition, however, is sufficiently generic to represent the group of entries referenced from the (leaf) nodes of a tree structure (e.g. a B-Tree) for a normalised col-

umn as well ( $\forall e \in E : |p_i(e)| \leq 1$ ). With condition (7) and condition (8) the groups of the access structure for a single-valued column are disjoint.

### B. Complexity of Queries Addressing Single Columns

With the definitions from the previous section we are ready to introduce complexity measures for time and space required for processing queries. A simple search condition for a column  $i$  is a set of alternative values  $C_i \subseteq V$ . With this approach we cannot only model conditions specifying a single value for a column but also other types of conditions like search ranges. All entries  $e \in E$  which contain at least one value of this set ( $p_i(e) \cap C_i \neq \emptyset$ ) are part of the result set.

Complex search conditions consist of multiple search conditions which may not only refer to different columns but also to the same column. An example is a set of words which all need to be included in a text column. For the lack of space we do not provide a formal definition of complex search conditions in this paper.

We need to visit a group of an access structure if at least one entry in the group satisfies the search condition. Accordingly we define the function *visit* returning exactly these groups:

$$visit(C_i, \mathcal{A}) = \{N | N \in \mathcal{A} \wedge (p_i(N) \cap C_i) \neq \emptyset\} \quad (9)$$

The result set for a search condition is the union of all groups of the access structure which need to be visited:

**Lemma 1.** *Let  $C_i = \{v_1, \dots, v_p\}$  be a simple search condition and  $\mathcal{A}$  be an access structure. Then the following function “result” provides all entries which satisfy  $C_i$ :*

$$result(C_i, \mathcal{A}) = \bigcup_{N \in visit(C_i, \mathcal{A})} N$$

This lemma follows directly from conditions (7) and (8). With these definitions we are ready to specify functions as complexity measures for space and time. The first function *time* summarises the number of entries for all groups in the access structure  $\mathcal{A}$  which need to be visited for a given search condition  $C_i$ .

$$time(C_i, \mathcal{A}) = \sum_{N \in visit(C_i, \mathcal{A})} |N| \quad (10)$$

The spatial complexity is given by the function *space* summarising the sizes of all groups of the access structure:

$$space(\mathcal{A}) = \sum_{N \in \mathcal{A}} |N| \quad (11)$$

With these definitions and the previous lemma the search time of an access structure is limited by the size of the result set:

**Lemma 2.** Let  $C_i = \{v_1, \dots, v_p\}$  be a simple search condition and  $A$  be an access structure. Then the search time is limited as follows:

$$\text{time}(C_i, A) \leq p \cdot |\text{result}(C_i, A)| \quad (12)$$

*Proof.* With the definition of the function *time* we get

$$\begin{aligned} \text{time}(C_i, A) &= \sum_{N \in \text{visit}(C_i, A)} |N| \\ &= |N_1| + \dots + |N_p| \\ &\quad \text{with } \{N_1, \dots, N_p\} = \text{visit}(C_i, A) \\ &\quad \text{and } N_i \subseteq N_1 \cup \dots \cup N_p \\ &\quad \text{for } i = 1, \dots, p \\ &\leq p \cdot |\text{result}(C_i, A)| \quad \text{with Lemma 1} \end{aligned}$$

A direct conclusion is that the search time is not dependent on the number of entries  $|E|$  from the considered table (constant time complexity).  $\square$

The next lemma shows that the space required for an access structure is limited by the number of entries and the average number of values in the considered column  $i$ .

**Lemma 3.** Let  $E = \{e_1, \dots, e_n\}$  be a set of entries,  $A_i$  an access structure for column  $i$  and  $\text{avg}_i(A_i)$  the average number of values in this column  $i$ :

$$\text{avg}_i(A_i) = \frac{|p_i(e_1)| + \dots + |p_i(e_n)|}{n} \quad (13)$$

Then the space required by the access structure for column  $i$  is limited by the following expression:

$$\text{space}(A_i) \leq n \cdot \text{avg}_i(A_i) \quad (14)$$

*Proof.* With conditions (7) and (8) every entry  $e$  occurs in not more than  $|p_i(e)|$  groups. As a consequence the summarised number of entries in the access structure  $A_i$  is limited by the summarised number of values in column  $i$  of all entries:

$$\leq |p_i(e_1)| + \dots + |p_i(e_n)|$$

This expression can be rewritten as follows:

$$\leq n \cdot \frac{|p_i(e_1)| + \dots + |p_i(e_n)|}{n} \leq n \cdot \text{avg}_i(A_i) \quad \square$$

This indicates that the space of the access structure grows linearly with the number of entries in the table, if we assume that the average number of values in the considered column  $i$  can be limited by a constant. This seems to be a reasonable assumption for most applications.

### C. Complexity of Queries Addressing Multiple Columns

Queries addressing multiple columns can be already supported by separate access structures generated for each column. A standard approach is the selection of the most selective con-

dition for one of these columns returning fewer entries than the conditions for the other columns. This approach tries to linearly filter items retrieved after having searched for the more selective condition. The efficiency of this approach strongly depends on the selectivity of the individual conditions and the size of the result set. In some cases this might lead to a linear time complexity.

Many existing approaches supporting search conditions addressing multiple columns define sets for all possible combinations of values from these columns. We will model this idea by considering the cross product between the previously defined access structures, like hybrid index structures or extended B-Trees including combinations of values (e.g. concatenation, bit interleaving, etc.). Without loss of generality we will consider only the combination of pairs of access structures. The subsequent analysis can be easily extended by repeatedly combining the relevant access structures.

The following definition introduces the concept of a combined access structure by computing the cross product between two existing access structures.

**Definition 1.** Let  $A_i$  and  $A_j$  be two access structures for the columns  $i$  and  $j$ . Then the combined access structure  $A_{ij}$  is defined as follows:

$$A_{ij} = \{N_i \cap N_j \mid N_i \in A_i \wedge N_j \in A_j\} \quad (15)$$

The function for the set of visited nodes as well as the function for the time complexity can be easily extended for two search criteria:

$$\begin{aligned} \text{visit}(C_i, C_j, A_{ij}) &= \\ &= \{N \mid N \in A_{ij} \wedge (p_i(N) \cap C_i) \neq \emptyset \wedge \\ &\quad \wedge (p_j(N) \cap C_j) \neq \emptyset\} \\ \text{time}(C_i, C_j, A_{ij}) &= \sum_{N \in \text{visit}(C_i, C_j, A_{ij})} |N| \end{aligned} \quad (16)$$

The function *visit* applied to a combined access structure  $A_{ij}$  which was derived from two access structures  $A_i$  and  $A_j$  provides exactly the cross product of the groups which need to be visited for the access structures  $A_i$  and  $A_j$ :

**Lemma 4.** Let  $A_i$  and  $A_j$  be two access structures and  $A_{ij}$  be the combination of these access structures. Then for every pair of search conditions  $C_i$  and  $C_j$  the following property holds:

$$\text{visit}(C_i, C_j, A_{ij}) = \text{visit}(C_i, A_i) \times \text{visit}(C_j, A_j) \quad (17)$$

The proof of this lemma is straight forward. We can show that every element of one set is also an element of the other set using the above definitions. Also lemma 1 can be extended for combined access structures:

**Lemma 5.** Let  $C_i$  and  $C_j$  be two simple search conditions and  $A_{ij}$  be a combined access structure. Then the following function "result" provides all entries which satisfy both search conditions:

$$\text{result}(C_i, C_j, A_{ij}) = \bigcup_{N \in \text{visit}(C_i, C_j, A_{ij})} N \quad (18)$$

*Proof.* The result for a query with the search conditions  $C_i$  and  $C_j$  is the intersection of the result sets for the individual conditions:

$$\text{result}(C_i, C_j, \mathcal{A}_{ij}) = \text{result}(C_i, \mathcal{A}_i) \cap \text{result}(C_j, \mathcal{A}_j)$$

With lemma 1 we get:

$$\text{result}(C_i, C_j, \mathcal{A}_{ij}) = \left( \bigcup_{N \in \text{visit}(C_i, \mathcal{A}_i)} N \right) \cap \left( \bigcup_{N \in \text{visit}(C_j, \mathcal{A}_j)} N \right)$$

With lemma 4 we get:

$$\text{result}(C_i, C_j, \mathcal{A}_{ij}) = \bigcup_{N \in \text{visit}(C_i, C_j, \mathcal{A}_{ij})} N \quad \square$$

With these results we can deduce an upper bound for the search time complexity of a combined access structure.

**Lemma 6.** Let  $\mathcal{A}_{ij}$  be an access structure derived from the two access structures  $\mathcal{A}_i$  and  $\mathcal{A}_j$  and  $C_i = \{v_{i1}, \dots, v_{ip}\}$  and  $C_j = \{v_{j1}, \dots, v_{jq}\}$  be two search conditions. Then the search time is limited by the following expression:

$$\text{time}(C_i, C_j, \mathcal{A}_{ij}) \leq p \cdot q \cdot |\text{result}(C_i, C_j, \mathcal{A}_{ij})| \quad (19)$$

*Proof.* With the definition of the function *time* we get:

$$\text{time}(C_i, C_j, \mathcal{A}_{ij}) = \sum_{N \in \text{visit}(C_i, \mathcal{A}_i) \times \text{visit}(C_j, \mathcal{A}_j)} |N|$$

With  $\{N_{i1}, \dots, N_{ip}\} = \text{visit}(C_i, \mathcal{A}_i)$  and  $\{N_{j1}, \dots, N_{jq}\} = \text{visit}(C_j, \mathcal{A}_j)$  we can rewrite this expression as follows:

$$= \frac{|N_{i1} \cap N_{j1}|}{|N_{i1} \cap N_{jq}|} + \dots + \frac{|N_{ip} \cap N_{j1}|}{|N_{ip} \cap N_{jq}|} + \dots$$

Since each expression of the form  $N_{ix} \cap N_{jy}$  is a subset of  $\text{result}(C_i, C_j, \mathcal{A}_{ij})$  we get the following upper bound:

$$\leq |\text{result}(C_i, C_j, \mathcal{A}_{ij})| + \dots + |\text{result}(C_i, C_j, \mathcal{A}_{ij})| + \dots + |\text{result}(C_i, C_j, \mathcal{A}_{ij})| + \dots + |\text{result}(C_i, C_j, \mathcal{A}_{ij})|$$

This formula with  $p$  columns and  $q$  rows can be simplified as:

$$\text{time}(C_i, C_j, \mathcal{A}_{ij}) \leq p \cdot q \cdot |\text{result}(C_i, C_j, \mathcal{A}_{ij})| \quad \square$$

We can conclude from this lemma that also for a combined access structure the search time is not dependent on the number of entries in the table but only on the size of the result set and on the number of values specified by the search condition.

Unfortunately the space required for a combined access structure is not always acceptable. The following lemma provides an upper bound for the space complexity of such an access structure.

**Lemma 7.** Let  $\mathcal{A}_{ij}$  be an access structure derived from the two access structures  $\mathcal{A}_i$  and  $\mathcal{A}_j$ . Further let  $\text{avg}_i(\mathcal{A}_i)$  be the average number of values for column  $i$  and  $\max_j(\mathcal{A}_j)$  be the maximum number of values in column  $j$ :

$$\max_j(\mathcal{A}_j) = \max(\{|p_j(e)| \mid e \in E\})$$

Then the space of the access structure  $\mathcal{A}_{ij}$  is limited by the following expression:

$$\text{space}(\mathcal{A}_{ij}) \leq n \cdot \text{avg}_i(\mathcal{A}_i) \cdot \max_j(\mathcal{A}_j) \quad (20)$$

*Proof.*

$$\text{space}(\mathcal{A}_{ij}) = \sum_{N \in \mathcal{A}_{ij}} |N|$$

With conditions (7) and (8) every entry  $e$  occurs in not more than  $|p_i(e)|$  sets from  $\mathcal{A}_i$  and  $|p_j(e)|$  sets from  $\mathcal{A}_j$ . Since every group in  $\mathcal{A}_{ij}$  is an intersection of a group from  $\mathcal{A}_i$  and a group from  $\mathcal{A}_j$ , the entry  $e$  will not occur in more than  $|p_i(e)| \cdot |p_j(e)|$  groups. As a consequence the summarised number of entries in the access structure  $\mathcal{A}_{ij}$  is limited by the following expression:

$$\text{space}(\mathcal{A}_{ij}) \leq |p_i(e_1)| \cdot |p_j(e_1)| + \dots + |p_i(e_n)| \cdot |p_j(e_n)|$$

We can replace every size of a set of values in column  $j$  by the maximum number of values in this column:

$$\text{space}(\mathcal{A}_{ij}) \leq (|p_i(e_1)| + \dots + |p_i(e_n)|) \cdot \max_j(\mathcal{A}_j)$$

With the definition of  $\text{avg}_i(\mathcal{A}_i)$  we get:

$$\text{space}(\mathcal{A}_{ij}) \leq n \cdot \text{avg}_i(\mathcal{A}_i) \cdot \max_j(\mathcal{A}_j) \quad \square$$

Note that we get a tighter bound with this lemma if we choose the column with the lower maximum number of values as the column  $j$ . But even in this case a combined access structure may have unacceptably high space requirements. On the other hand this lemma ensures an acceptable space complexity if one column has a low maximum number of values. This is in particular true for combinations of single-valued columns with arbitrary other columns.

#### D. Tree Structure for Navigating to Sets of Access Structures

In this section we consider the tree structure required for the navigation to the sets of entries from an access structure. Nodes of these trees relate to sets of values representing sets of entries below the node. We only take into account structures for secondary storage. As we want to prove a logarithmic worst case time complexity for our structure, the B(+)-Tree is the choice

as the R-Tree cannot provide this. It is a well-known fact that a B+-Tree has the following time complexity for a range search retrieving  $m$  entries from a table with  $n$  entries:  $\mathcal{O}(\log(n) + m)$ . Similar to many other tree structures a B+-Tree has linear space complexity:  $\mathcal{O}(n)$ .

There are three different possibilities of navigation to the sets of an access structure: concatenation of structures for different columns, alternating the nodes on a path (we will not take into account this option, here) and augmenting nodes to provide information about both columns.

We will start with the concatenation of tree structures and analyse the time and space complexity depending on the order of these trees (multi-valued first or single-valued first). We will show that only the approach of "single-valued first" has reasonable space requirements but unfortunately a linear time complexity. Finally we will prove that the augmentation of the nodes in the primary tree with information about values from the multi-valued column of entries below this node is sufficient to ensure a logarithmic search time complexity.

For our analysis we consider the access structures  $\mathcal{A}_i = \{N_{i1}, \dots, N_{ip}\}$  and  $\mathcal{A}_j = \{N_{j1}, \dots, N_{jq}\}$  for the two columns  $i$  and  $j$ . The primary tree for this access structure refers to column  $i$  and supports the navigation to the  $p$  groups of the access structure  $\mathcal{A}_i$ . Every group  $N$  of this primary access structure is the starting point of a secondary tree. A secondary tree supports the navigation to groups which were generated by intersections between the group  $N$  and the groups from  $\mathcal{A}_j$ . The space required by this structure depends on the size of the primary tree and the number and sizes of the secondary trees. The size of a secondary tree depends not only on the number of entries from the group  $N$  which is the starting point for the tree but also the number of values in column  $j$  of these entries.

The average number of entries of a group  $N \in \mathcal{A}_i$  can be derived by dividing the space ( $space(\mathcal{A}_i)$ ) required for the access structure (summarised number of entries) divided by the number  $p$  of groups in the access structure  $\mathcal{A}_i$ :

$$\frac{space(\mathcal{A}_i)}{p} = n \cdot \frac{avg_i}{p} \quad (21)$$

A common function estimating the number of separate values versus the number of entries (size of a data set) is Heaps' law [13]. We will use the index  $j$  in this formula to indicate, that the parameters refer to column  $j$  in our case. Heaps' law may be applied to the average size of groups in access structure  $\mathcal{A}_i$ , here:

$$p_j(N) \approx c_j \cdot \left( n \cdot \frac{avg_i}{p} \right)^{\beta_j} \quad (22)$$

The parameters  $c_j$  and  $\beta_j$  are application specific parameters.

The number of values is an upper bound for the number of groups in the secondary access structure. Since the space required for the tree grows linearly with this number of groups

and we have  $p$  trees as secondary access structures, the total space required by these secondary trees can be estimated by:

$$\begin{aligned} & c_j \cdot \left( n \cdot \frac{avg_i}{p} \right)^{\beta_j} \cdot p \\ &= c_j \cdot \left( n^{\beta_j} \cdot avg_i^{\beta_j} \cdot p^{-\beta_j} \right) \cdot p \\ &= c_j \cdot n^{\beta_j} \cdot avg_i^{\beta_j} \cdot p^{1-\beta_j} \end{aligned} \quad (23)$$

Now we may consider two cases depending on whether column  $i$  or column  $j$  is the multi-valued column. As an immediate observation, the expression  $avg_i^{\beta_j}$  has the value 1 if  $i$  is the single-valued column. Otherwise this expression may have a high value, if the average number of values for column  $i$  is also high. Also the number of groups  $p$  in access structure  $space(\mathcal{A}_i)$  is usually much greater for a multi-valued column than for a single-valued column. This may result in an unacceptably high space requirement for a solution with the primary access structure for the multi-valued column.

In the following we will assume, that the primary access structure relates to the single-valued column. In this case the average number of values in entries for this column is less or equal than one. This means that we can rewrite the previous expression as follows:  $c_j \cdot n^{\beta_j} \cdot p^{1-\beta_j}$

We may safely assume that the number of values for the single-valued column  $i$  may not grow faster than linear with the number of entries  $n$ . Therefore,  $n$  is an upper bound for  $p$  and we get the following expression indicating a linear space complexity for this tree:  $c_j \cdot n^{\beta_j} \cdot n^{1-\beta_j} = c_j \cdot n$

The time complexity for searching the concatenated B+-Trees can be directly deduced from the time complexity of an ordinary B+-Tree. The first B+-Tree for column  $i$  needs to manage all  $n$  entries from the table and supports the retrieval of an intermediate result set with  $m_i$  entries satisfying only the related condition for column  $i$ . The entries of the intermediate result set are included in not more than  $m_i$  groups of the primary access structure  $\mathcal{A}_i$ . Each of these groups is the starting point for another tree structure managing only the  $n_k$  entries from the group. These secondary tree structures will deliver  $m_k$  entries satisfying both search conditions. With this approach the summarised search time complexity is given by the following expression:

$$\log(n) + \sum_{k=1}^{m_i} (\log(n_k) + m_k) \quad (24)$$

Since the primary access structure refers to a single-valued column, their groups are disjoint and the appended secondary access structures manage also disjoint groups of entries. As a consequence we may rewrite the previous expression as follows, assuming that  $m$  is the total number of entries satisfying both search conditions:

$$\log(n) + \sum_{k=1}^{m_i} \log(n_k) + \sum_{k=1}^{m_i} m_k = \log(n) + \sum_{k=1}^{m_i} \log(n_k) + m \quad (25)$$

In comparison to the time complexity of a conventional B+-Tree this time complexity has the additional addend  $\sum_{k=1}^{m_i} \log(n_k)$ . In the worst case this addend may result in a linear time complexity. We can show that by a simple example. We consider for this example a wide range search for column  $i$  which is satisfied by all entries. Further we consider a condition for column  $j$  defining a value which occurs in no entry. In this case the intermediate result set contains all entries ( $m_i = n$ ) and all secondary tree structures need to be searched without retrieving a result. If we assume that every secondary tree structure has at least one node this expression would imply at least a linear search time complexity:  $\sum_{k=1}^{m_i} \log(n_k) = \sum_{k=1}^n \log(n_k) \geq n$ .

This follows from the problem that secondary trees need to be searched even if they do not return any results. It can be avoided if only those secondary trees need to be searched which guarantee the retrieval of a certain number  $\varepsilon$  of entries which is not dependent on the total number  $n$  of entries. We can guarantee this property by adding information about all values of column  $j$  which occur at entries below a node or below the groups from the primary access structure. In the next section we will introduce such a structure augmenting every reference to nodes or secondary trees with bit lists indicating which values are still available in entries following a reference.

We summarise the result of the analysis in the following theorem.

**Theorem 1.** *Let  $\mathcal{A}_{ij}$  be a combined access structure for the single-valued column  $i$  and the multi-valued column  $j$  using concatenated B+-Trees for the navigation to the groups from*

*$\mathcal{A}_{ij}$ . The primary tree refers to column  $i$  and the secondary tree refers to column  $j$ . The references of the primary tree are augmented in such a way that a search continues beyond a reference if at least  $\varepsilon$  entries from the referenced structure satisfy the conditions for column  $j$ . Then the search time is limited by the following expression where  $n$  is the number of entries and  $m$  is the size of the result set:*

$$\mathcal{O}(\log(n) \cdot \frac{m}{\varepsilon} + m)$$

*Proof.* In the previous section we could already prove that the time complexity of an access structure does not depend on the number of entries. Therefore, we may focus on the time required to navigate through the tree structure. Based on equation (25) and our assumption that the search continues only beyond a reference of the primary structure if at least  $\varepsilon$  entries satisfy

the search condition for column  $j$ , means that not more than  $\frac{m}{\varepsilon}$  secondary tree structures are searched. Additionally, the number of entries in a secondary access structure is limited by  $n$ :

$$\leq \log(n) + \sum_{k=1}^{\frac{m}{\varepsilon}} \log(n) + m$$

Now we may rewrite this expression as follows:

$$\leq \log(n) + \frac{m \cdot \log(n)}{\varepsilon} + m$$

Since we need to consider only the fastest growing term for the  $\mathcal{O}$ -notation we get the following expression as the upper bound for the search time complexity (see lemma 2):

$$\mathcal{O}(\log(n) \cdot \frac{m}{\varepsilon} + m) \quad \square$$

In the next section we use bit lists to indicate the presence of a value below a node. This means that none of these frequencies is zero and therefore their product is also greater than zero. This expression also indicates that the parameter  $\varepsilon$  grows with the number of entries managed by a secondary access structure. As a probably surprising consequence, the efficiency of the hybrid access structure increases with the number of entries stored in the secondary access structures!

#### IV. VALIDATION OF THE ANALYSIS

The theory proved that a hybrid indexing approach might be used for retrieval of normalised and non-normalised values combined inside a hybrid data space using a logarithmic complexity. Yet, it did not propose a concrete implementation structure with the ability to do so. One key challenge for such a structure is the combined representation of the two value types. A second one is the basic storage structure to be selected. The first one is solved by the augmentation of the elements of a base structure with a bitlist. The second one is strongly application dependent. For the validation of the analysis, we implemented two different storage structures. One uses a B+-Tree as the primary structure and the other an adopted R-Tree variant ensuring a logarithmic retrieval performance under certain circumstances (see [14]). Hence, two hybrid index structures supporting the combined storage of normalised and non-normalised values using a B-Tree variant or an R-Tree variant (proposed in [10]) are validated regarding their retrieval behaviour to evaluate the analysis given in the previous section regarding a real-world application. Based on the limited space available in realistic environments, an assumption must be made based on Zipf's Law [15] to limit the amount to a reasonable number. A parameter called *HLimit* is introduced which separates the set of non-normalised values into two subsets of high and low frequently occurring values. Only the high frequent values are stored inside the hybrid index.

The tests are carried out by using a specially preprocessed document set from a Wikipedia dump<sup>1</sup>. Textual analysis comprises stop word removal, character normalization and stemming. Geographical analysis consists of assignment of coordinates to textual occurrences. Texts represent the non-normalised and the coordinates the normalised value parts. The measurements consist of a set of 699 queries constructed from the AOL Query Log<sup>2</sup>. The number of documents at each measurement point is doubled which leads to the possibility of showing logarithmic query behavior. The B-Tree applies only the latitude values of the respective documents and the R-Tree uses both dimensions.

<sup>1</sup> <http://dumps.wikimedia.org/enwiki/20111007/enwiki-20111007-pages-articles.xml.bz2>, accessed 2012-10-17

<sup>2</sup> <http://www.gregsadetsky.com/aol-data/>, from 2012-10-17

The goal of this evaluation is solely to experimentally validate the logarithmic behaviour of the retrieval complexity for this kind of index structures.

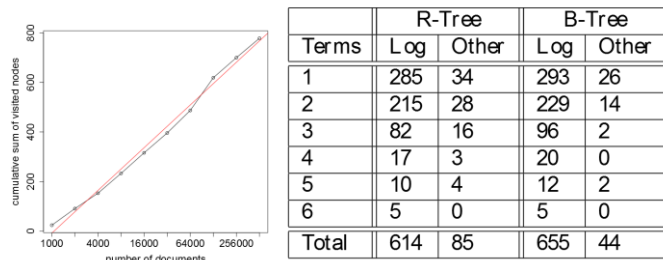


Figure 1. Logarithmic Example and Table of Results

A typical example of the retrieval behaviour can be seen in the plot of figure 1 (left side). The abscissa is scaled logarithmic which means that the given linear increase of cumulative sum of loaded pages leads to the fact that the total retrieval complexity behaves logarithmic. The table in figure 1 shows the combined results of the respective experiments for B-Tree and R-Tree. It can be seen that most cases show a logarithmic behaviour, even for more than one search term in the query. Approximately 87.84% of the queries could be answered with logarithmic complexity for the R-Tree and 93.7% for the B-Tree case. Hence, a clear tendency towards the logarithmic behaviour can be stated from these data. The remaining cases are explainable based on corpus specific properties. This includes, e.g., a change in the vocabulary. Hence, either constant query cases for a very small number or a sudden rise in the number of references leads to these non-logarithmic query behaviours.

Thus, summarizing, the evaluation using the given document set and the adopted hybrid access methods experimentally validates the theoretical proves from section III.

## V. SUMMARY AND FUTURE PROSPECTS

In this paper we showed how to combine multiple index structures caring about single-valued and multi-valued attributes simultaneously. We formed a theoretical basis which proved the efficiency of one possibility of combinations to manage the given types of data. The practical evaluation carried out in a real world relational database management system confirmed the assumptions asserted in the theoretical analysis part.

Therefore, we could show in the theoretical analysis as well as in the practical evaluation that an augmented hybrid B-Tree indexing concept can achieve a logarithmic time complexity regarding typical queries. This also holds for a specialized hybrid R-Tree variant with logarithmic base complexity.

There are still objects of study regarding the particular storage parts in the hybrid access structure. Especially, the proper

set up of the arbitrary user defined upper bound  $HLimit$  for frequent terms will have to be investigated.

## REFERENCES

- [1] R. Bayer and E. McCreight, "Organization and Maintenance of Large Ordered Indices," in Proceedings of the 1970 ACM SIGFIDET (Now SIGMOD) Workshop on Data Description, Access and Control, New York, NY, USA, 1970, pp. 107–141.
- [2] A. Guttman, "R-trees: A Dynamic Index Structure for Spatial Searching," in Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data, New York, NY, USA, 1984, pp. 47–57.
- [3] C. Faloutsos and H. V. Jagadish, "Hybrid index organizations for text databases," in Advances in Database Technology — EDBT '92, A. Pirotte, C. Delobel, and G. Gottlob, Eds. Springer Berlin Heidelberg, 1992, pp. 310–327.
- [4] C. Chen, S. Pramanik, Q. Zhu, W. Alok, and G. Qian, "The C-ND Tree: A Multidimensional Index for Hybrid Continuous and Non-ordered Discrete Data Spaces," in Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, New York, NY, USA, 2009, pp. 462–471.
- [5] R. Hariharan, B. Hore, C. Li, and S. Mehrotra, "Processing Spatial-Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems," in 19th International Conference on Scientific and Statistical Database Management, 2007. SSBDM '07, 2007, pp. 16–16.
- [6] I. D. Felipe, V. Hristidis, and N. Rishe, "Keyword Search on Spatial Databases," in 2013 IEEE 29th International Conference on Data Engineering (ICDE), Los Alamitos, CA, USA, 2008, vol. 0, pp. 656–665.
- [7] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsuregawa, "Keyword Search in Spatial Databases: Towards Searching by Document," in 2013 IEEE 29th International Conference on Data Engineering (ICDE), Los Alamitos, CA, USA, 2009, vol. 0, pp. 688–699.
- [8] N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger, "The R\*-tree: An Efficient and Robust Access Method for Points and Rectangles," in Proceedings of the 1990 ACM SIGMOD International Conference on Management of Data, New York, NY, USA, 1990, pp. 322–331.
- [9] J. B. Rocha-Junior, O. Gkorgkas, S. Jonassen, and K. Nørkvåg, "Efficient Processing of Top-k Spatial Keyword Queries," in Proceedings of the 12th International Conference on Advances in Spatial and Temporal Databases, Berlin, Heidelberg, 2011, pp. 205–222.
- [10] R. Göbel, A. Henrich, R. Niemann, and D. Blank, "A Hybrid Index Structure for Geo-textual Searches," in Proceedings of the 18th ACM Conference on Information and Knowledge Management, New York, NY, USA, 2009, pp. 1625–1628.
- [11] R. Göbel and C. Kropf, "Towards Hybrid Index Structures for Multi-Media Search Criteria," in Proceedings of the 16th International Conference on Distributed Multimedia Systems, DMS 2010, October 14–16, 2010, Hyatt Lodge at McDonald's Campus, Oak Brook, Illinois, USA, 2010, pp. 143–148.
- [12] J. M. Hellerstein, E. Koutsoupias, D. P. Miranker, C. H. Papadimitriou, and V. Samoladas, "On a Model of Indexability and Its Bounds for Range Queries," J ACM, vol. 49, no. 1, pp. 35–55, Jan. 2002.
- [13] H. S. Heaps, Information Retrieval: Computational and Theoretical Aspects. Orlando, FL, USA: Academic Press, Inc., 1978.
- [14] R. Göbel, "Towards Logarithmic Search Time Complexity for R-Trees," in Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, T. Sobh, Ed. Springer Netherlands, 2007, pp. 201–206.
- [15] G. K. Zipf, Human Behaviour and the Principle of Least Effort: an Introduction to Human Ecology. Addison-Wesley, 1949.



# Creating Web3D Educational Stories from Crowdsourced Annotations

Ivano Gatto and Fabio Pittarello  
Università Ca' Foscari Venezia  
Via Torino 155 – 30172 Mestre (VE), Italy  
{igatto, pitt}@dsi.unive.it

**Abstract**— 3D representation and storytelling are two powerful means for educating students while engaging them. This paper describes a novel software architecture that couples them for creating engaging linear narrations that can be shared on the web. The architecture takes advantage of a previous work focused on the semantic annotation of 3D worlds that allows the users to go beyond the simple navigation of 3D objects, permitting to retrieve them with different search tools. The novelty of our architecture is that authors don't have to build stories from scratch, but can take advantage of the crowdsourced effort of all the users accessing the platform, which can contribute providing assets or annotating objects. At our best knowledge no existing workflow includes the collaborative annotation of 3D worlds and the possibility to create stories on the top of it. Another feature of our design is the possibility for users to switch from and to any of the available activities during the same session. This integration offers the possibility to define a complex user experience, even starting from a simple linear narration. The visual interfaces of the system will be described in relation to a case study focused on culture heritage.

*annotation; tag; folksonomy; ontology; storytelling; web3d*

## I. INTRODUCTION

3D representation for the web has been available since the advent of the first VRML specification [1]. From 1995 to present different specifications have been proposed. Millions of users have accessed 3D worlds using both open-source and proprietary browsers and platforms, in some cases sharing the experience with remote users, as it happens for 3D multi-users environments such as Second Life. Usually the access to these environments is realized using browsing paradigms that allow the users to navigate the 3D worlds walking, flying or rotating around viewpoints chosen by the content authors. Unfortunately this access model enables often only a partial exploitation of the 3D world that results from a complex and time-consuming modeling work. That is the reason why in recent years a number of researchers have proposed the use of annotations for exploiting further the potential of 3D worlds. The association of high-level descriptions to geometric entities has enabled the access to content through searching paradigms, retrieving specific components of the 3D environments after a keyword-based search or more sophisticated queries involving also spatial operators. Besides, the introduction of folksonomic annotation styles has enabled common users to describe the 3D components with their own words.

In this work the term annotation will be referred both to the use of keywords belonging to an ontology/taxonomy, to free

tags belonging to a folksonomy and even to extended description associated to 3D entities. The novel approach presented in this work builds on a previous research work related to 3D annotation [2] but goes a step further, introducing the opportunity to use the annotation work done by the users of the ToBoA-3D platform for creating 3D narrations that can be shared on the web. The shift is paramount: from the free navigation and search paradigm to the possibility to design and share linear narrations through a 3D environment. In particular, in our proposal, stories don't start from scratch, but from the annotation work done by the users of a shared platform for annotating 3D environments. The overall workflow is novel as well. While some proposals described in literature take advantage of tools for enabling single content experts to annotate 3D objects [3], no workflow includes the crowdsourced annotation and the possibility to create a story on the top of it. This approach opens a number of interesting opportunities for the educational domain, starting from the creations of engaging lessons delivered as narrations. The software architecture described in this work can be seen as a form of sentient multimedia system, because it allows people distributed over the network to gather, annotate, process and retrieve distributed Web3D resources, taking advantage of the crowdsourced effort done by many users that play different parts (e.g., content provider, content annotator, content browser, story creator and listener) even in the same session. The rest of the work is organized as follows: Section II will consider the related work; Section III will describe the features of the annotation architecture where we started from; Section IV will describe the requirements of our storytelling system, focusing in particular on the crowdsourced approach and the integration of the different user activities; Section V will present the user interfaces for the creation and the navigation of stories; Section VI will draw the conclusions.

## II. RELATED WORK

Among the different proposals for describing 3D objects and worlds on the net, a special role for their longevity is reserved to VRML and to its heir X3D [4]. VRML and X3D plugins associated to web browsers have allowed millions of web users to navigate Web3D worlds. The latest implementations permit even to play X3D worlds using only standard web browsers compliant with WebGL [5], enlarging further the number of platforms and users that can access 3D representations on the web. While the representation of 3D geometrical entities on the web is an interesting opportunity for communicating knowledge, their annotation permits to exploit

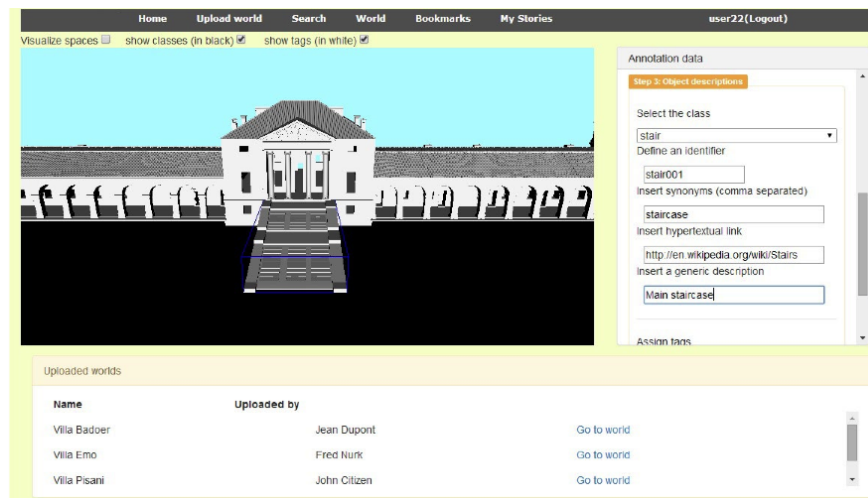


Figure 1. The interface for annotating the 3D objects.

the potential of virtual representation, allowing users to go beyond the simple navigation of 3D worlds. For this reason in the last few years a number of proposals for adding semantics to the components of 3D worlds have been made, based on different specifications such as MPEG-7 [6], Collada [7] or X3D [2]. While most of these proposals focus on annotations referred to predefined ontologies, some researchers have focused also on the use of tags as a complementary means for annotating 3D entities with an informal approach more suited to common people [2]. The benefits of using social tagging in education and, in particular, in cultural heritage contexts have already been described in [8].

As noted by Scopigno et al. [9], the greater challenge for digital technologies is the creation of tools that use 3D models for supporting cultural heritage research. In this respect, the annotation of 3D models is only a first step for supporting more appropriately cultural heritage studies. We claim that the application of storytelling techniques to annotated 3D worlds belonging to the cultural heritage domain can bring great advantages for researchers and pupils. The benefits of storytelling for educational experiences have been demonstrated by several studies [10] [11]. As far as cultural heritage is concerned, in literature there are different examples showing how storytelling techniques can be profitably used for engaging students while learning. The techniques used can be different, relating the different fragments of the narrations to photos and videos [12], real environments [13] and virtual representations [14]. As far as the latter ones are concerned, most examples are focused on the delivery of linear and non-linear stories rather than on authoring tools. The proposals for the creation of stories can be split in two different categories: the “autonomous agents approach”, where a set of software agents influences the evolution of the story [15], and the “drama manager approach”, where a software architecture controls the narration on the basis of the story model and of the narrative choices of the author [16]. The approach described by Kriegel et al. [17] belongs to the first category. The authors present an authoring tool for the emergent narrative agent architecture FATiMA, which powers the virtual bullying drama FearNot!. The system allows the creation of a story starting from the behavior of the characters. The authors specify the

actions that the character can perform, the goal that he can reach and the way in which he interacts with the story events. The process starts with the decision on the story setting, and the placement of the characters and objects in the story. After the set-up of the scene, the author specifies the behavior of all the participating characters. Both the delivery and the authoring of the story happen inside a 3D environment. The approaches described in [18], [19] and [20] belong to the category of drama managers, with a stronger emphasis on the story structure rather than on the characters. Mehm et al. [18] describe a system named StoryTec which allows the creation of story-based serious games. This system is conceived for people with low skills in computer science. Teachers can use this tool for creating small educational games that can be played during courses; game programmers and content producers can take advantage of StoryTec as a prototyping environment for developing and testing ideas. The story is divided in units, which the author can link together in order to create a path through the story. Once the story path is established, the author defines the details by adding objects to the scene and specifies the events and the actions for each step of the story. The whole system is built on a hybrid 2D/3D framework based on the Windows Presentation Foundation libraries and provides different components for managing the creation of the narration: stage editor, story editor and resource center. The first one is a WYSIWYG editor and is required to define the details and objects of the single stage of the story. The author can take any object available on the resource center (for example a 2D or 3D asset, a sound file) and insert it into the stage using a drag and drop interface. The second component, the story editor, allows creating the story structure. This interface allows the author to specify the path through the story, connecting with arrows the visual objects that represent the stages. The author takes advantage of the story editor also for specifying complementary information for the playing phase, such as annotations or the expected time the user will remain in a given stage.

Robertson et al. [19] introduce the authoring tool named Adventure Author. The tool is based on a 3D game engine which gives to the story a graphical aspect similar to the graphics of commercial videogames. This tool has been

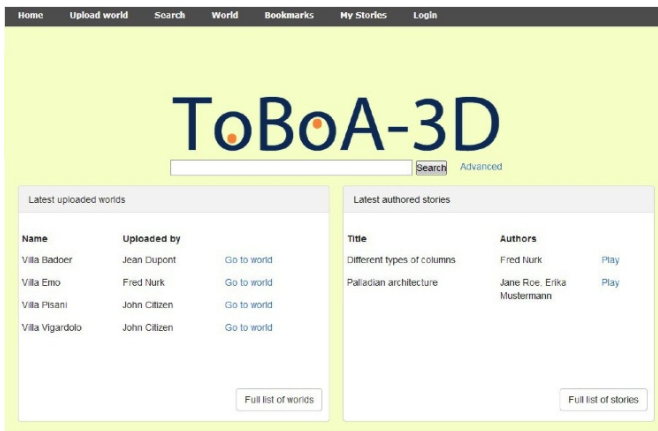


Figure 2. The main interface of the annotation platform.

conceived for young students, who can experiment the creation of non linear interactive stories. The educational aspect of the tool relies on the fact that users are encouraged to use their imagination for structuring the possible evolution of the story. The system is made of two main components: the authoring interface for the definition of the interactive story and the game engine for rendering the story. The JGraph library [21] has been used for giving the author the possibility to define the non linear structure of stories. The nodes of the acyclic graph represent the scenes of the story. Each scene contains different information, such as textual descriptions which will be displayed to the player entering the scene and interactive conversations between the characters of the story. The author can also take advantage of a set of wizards for creating new characters, locations and story scenes. At the end of the story creation process, the user will take the role of one of the characters of the story and will interact in a 3D world.

Barrenho et al. [20] describe an interesting authoring tool for the InStory platform, designed for the development of exploratory geo-referenced activities. InStory has been developed for mobile devices and allows the user to interact with the story by moving inside a physical space of cultural, historical or natural interest. The system divides the creation work in two different phases. In the first phase the InAuthor component allows the author to specify the nodes of the story, corresponding to different activities, and to define the connections among them. The interface is articulated in different modules. The first one allows users to create the new nodes and to browse those ones that have been previously created. The second one allows users to position the nodes on a map and to link them for creating the story path. In the second phase the InContent component is used to assign the content to the nodes. The interface is composed by a workspace for the visual creation of content, with a list of the elements that can be used. Authors have just to drag and drop the desired media or interface elements on the workspace representing the client device screen and place them in the desired position. After the drag and drop operation, all the properties of the elements can be edited for reaching the final result.

The approach proposed in this work is compliant with the drama manager approach described above, because we were more interested in the development of a storytelling system suitable to educational purposes, more controllable for what

concerned the number of choices and endings. As far as the delivery of narrations is concerned, we must point out that in a number of proposals available in literature stories can be published on the web, for maximizing the possibility of access by the users, but the potential of the net is not always exploited in terms of collaboration and content sharing. Our proposal differs from the others that we have analyzed, because it is strongly characterized by a web-based collaborative approach that allows the story authors to take advantage, for the story creation process, of the work done by the other users of the platform in terms of creation of assets and annotations. The authors of stories themselves can add assets and annotations, contributing to the growth of the repositories. At our best knowledge no workflow includes the collaborative annotation of 3D worlds and the possibility to create stories on the top of it. Finally, in the current implementation our approach is less complex for what concerns the structure of the story proposal, focusing on the design of linear stories that are easier to create for content experts not particularly skilled and easier to follow for pupils. In spite of this simplicity, the possibility to switch from and to the different user activities gives the possibility to create a rewarding user experience, as it will be explained in the following sections.

### III. THE ANNOTATION PLATFORM

The tool for creating stories presented in this work is conceived as a significant step in the evolution of the ToBoA-3D platform for annotating Web3D environments, whose first design and implementation has been described in [2]. The client-server platform is entirely based on the use of web technologies, including X3D, XHTML, CSS and Javascript on the client side; Apache, MYSQL and PHP on the server side. The annotation platform can be accessed using a web browser and the BS Contact Player [22] for navigating the X3D worlds. 3D annotations are applied to a repository of 3D worlds compliant with the X3D standard, uploaded by registered users. The annotations result from the crowdsourced effort done by many users, even those ones that didn't upload the worlds, as it happens for the free tagging in the hypertextual web 2.0. Each user can add one or more tags to any simple or grouped 3D entity and she can even create new logical groups to tag, in the case the 3D modeler didn't define a given group of entities. While all the registered users are enabled to add tags, the owner of a given world can adopt also a more formal annotation style, referencing to a closed set of keywords belonging to a formal ontology. Because the space is an important component of the human experience, the system allows users to add annotations not only to 3D objects but even to spaces defined by them. Fig. 1 shows the selection and annotation of a 3D object, an external staircase annotated as an occurrence of the class "stair", belonging to an ontology that describes architectural artifacts. In the example the user has inserted also a synonym, "staircase", and an hyper textual reference to an external description. An additional field, partially visible in the figure, allows users to add one or more comma separated tags. At the end of the process the visual interface allows to record the annotated object and the reference to the identity of the user that has annotated it. The annotated components of 3D worlds can be searched using both a simple - Google like - keyword search (see Fig. 2) or a more sophisticated - yet simple to use -

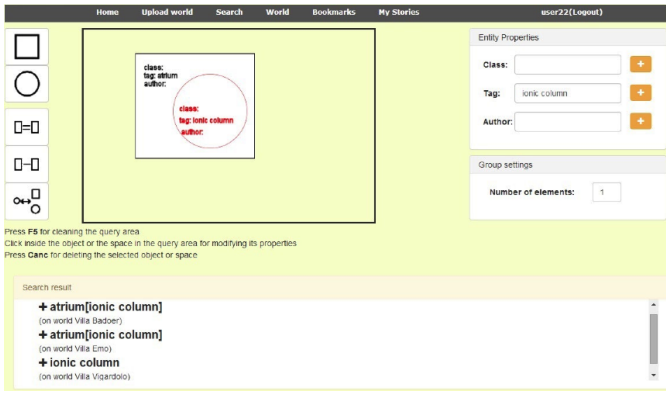


Figure 3. The interface for composing complex visual queries.

visual interface that permits to specify also spatial relations (e.g., search all the objects, annotated as “ionic column”, embedded in spaces annotated as “atrium”, see Fig. 3). In both cases the system presents the query results as a clickable lists of items that can be selected for teleporting the user in front of the retrieved entities. Further details can be found in [2] and [23]. During the navigational phase the user can access directly any 3D world uploaded to the system and visualize the textual labels related to the annotated objects, as it will be explained in detail in Section V.

#### IV. DESIGN AND IMPLEMENTATION OF CROWDSOURCED STORYTELLING

The decision of building the storytelling module on the top of the annotation platform was one of the early choices of our design. The main reason was the will to break the traditional production workflow of interactive 3D worlds, where often the decisions about the user experience inside the 3D worlds are taken by the 3D modelers and by the interaction designers. This often leads to poor results from a narrative point of view. What we wanted instead was to design an architecture supporting the effort of a community of users, targeted at the creation of interesting educational stories. While some workflows described in literature take advantage of tools for enabling single content experts to annotate 3D objects [3], there is no workflow that includes the collaborative annotation and the possibility to create a story on the top of it. In our proposal the story creators don't need to start from scratch, but can take advantage of the 3D assets and the annotations provided by a community of registered users that include of course content experts. In our vision the story creator is an active contributor to the assets as well, uploading 3D worlds and defining annotations that are functional to the story she's going to create, but that contribute also to the overall growth of the repository of annotated 3D environments.

We define a “story” as a sequence of stages that have a specific location and that are associated to different multimedia content, such as images, audio and video. While in literature there are more sophisticated story models (e.g., non linear models), we started from the linear structure because of its suitability to educational applications. Linear stories are easier to create and easier to follow for novices. Besides, they offer a constrained path that fits to educational needs. Even with these limitations, our storytelling application offers a significant

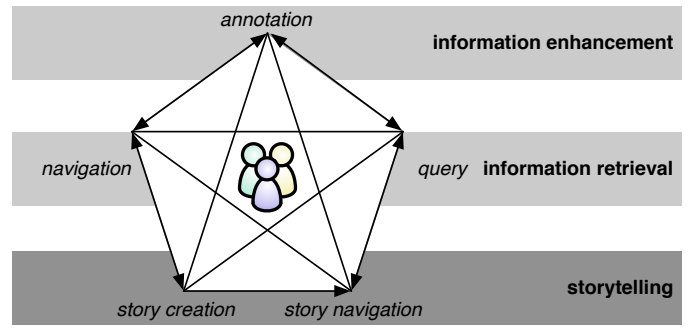


Figure 4. Interconnected user activities.

potential, giving for example to a teacher the possibility to create a narration for her pupils and then to ask them to play and comment it, adding contextual annotations related to objects, spaces and story stages. As a matter of fact, as displayed by Fig. 4, a peculiar feature of our design is the integration of the different activities that can be made by the users of the platform (i.e., annotation, navigation, query, story creation and navigation) and the possibility to switch from and to any of them during the same session. The figure highlights also three categories of activities: information retrieval, information enhancement and storytelling. While the integration of the navigation and query activities (information retrieval) is recommended since the advent of the first web [24] for enhancing the “findability” of information objects, the advent of the hyper textual web 2.0 has brought the possibility for the web user to play a more active role, enhancing the information available on the web sites through different annotation techniques (information enhancement). In a typical web 2.0 session the user plays the part of the content reader and author, switching seamlessly between the two different roles. In the prior version of our annotation architecture [2] we brought to the Web 3D the same possibility, previously available only for the hyper textual web, to switch from and to the activities related to the information retrieval and enhancement.

In this work we introduce, with the novel architecture, the activities related to the creation and navigation of stories (storytelling), going a step further with the integration of the different activities in a rewarding experience. If the paths from the information retrieval and enhancement activities to storytelling seem interesting and even obvious, because stories are created from the annotated Web3D objects, the design choice of allowing the user to select the opposite path brings additional benefits. In particular the possibility to shift from the storytelling to the other activities can be interesting for the ordinary user wanting to explore a world he's hearing a story about or to add information he's aware of about one of the objects that are visualized. This chance is even more interesting for educational purposes: for example the user may be asked by the story teller (the teacher or another content expert) to explore the world starting from the locations of the story or to add annotations to the objects for proving their cognitive involvement and knowledge acquisition (see Section V). In this context, an interesting feature of our approach is that all the assets, annotations and stories are social but not anonymous. Resuming, the integration of the different activities gives the possibility, starting from a simple linear narration, to define a complex educational experience. The stories will be created



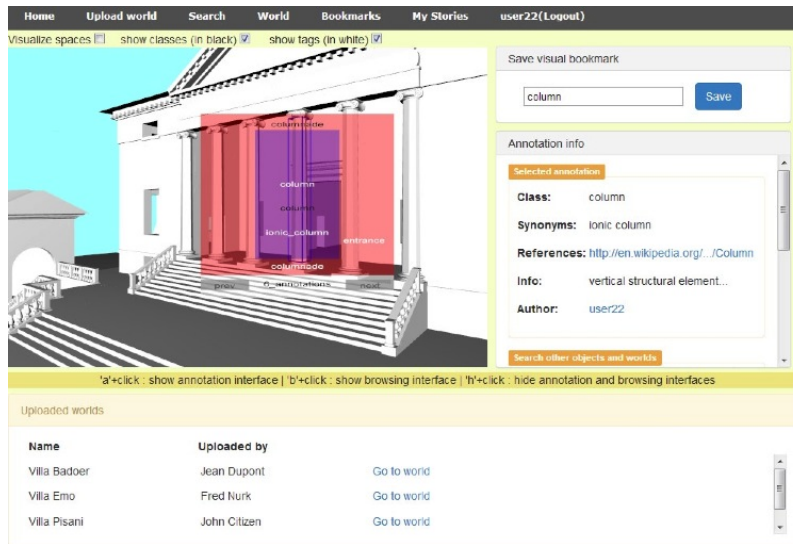


Figure 5. The navigational interface.

and played through a set of interfaces that will be described in the following section. At any time the users of the platform will be enabled to switch seamlessly among the different activities, for playing different roles even in the same session.

## V. THE VISUAL INTERFACES FOR STORYTELLING

After the definition of the requirements for the new storytelling modules, we designed the set of interfaces supporting the sequence of tasks functional to the creation and listening of the narrations: the gathering of interesting annotated objects and spaces, their selection and organization in an ordered sequence of story stages, the association of story stages to multimedia content, the preview and the final play. The design process gave as a result a new set of interfaces (Fig. 6 and 7). Also the set of existing interfaces (Fig. 1, 2, 3 and 5) underwent a consistent process of redesign as well, for introducing novel functionalities and for supporting the integration of the different activities. The technical implementation was completely revised, taking advantage of the Angular.js [25] and Bootstrap [26] frameworks.

As already underlined in the previous Section the architecture supports a number of user activities that make sense for different scenarios and personas: the 3D modeler wanting to contribute to the social platform uploading a new model, the casual user navigating and tagging the available 3D content, the art historian using the advanced visual search module for retrieving and examining specific architectural patterns. In this Section we will describe the visual interfaces for storytelling in relation to two different personas and use scenarios: the art expert creating an educational story for his class of students and the art student listening to the story.

### 1) The art expert creating an educational narration.

The art expert wants to take advantage of the ToBoA-3D platform for creating a narration for his class of students. His goal is to provide students with information about the features of the architectures designed by Andrea Palladio, one of the most renewed architects of the Italian Renaissance. He wants also to check what the students have learned listening to his narration. In the initial phase the art expert explores the 3D

content of the platform, accessing directly the worlds available from the repository list (Fig. 2, left panel) or querying the system (Fig. 3). In both cases, as a final result, the art expert enters the navigational interface displayed in Fig. 5. The main part of the interface is occupied by the 3D representation. The expert explores the available 3D scenes for finding interesting locations and objects that will be part of the narration. When he selects a given object (e.g., the ionic column evidenced by the blue bounding box), a semi-transparent rectangle highlights all the annotations associated to it. The right part of the screen evidences all the information associated to the annotation currently selected, including the author and the reference to an extended hyper textual description. The interface features a button placed on the top right corner of the screen, for saving the current viewpoint as a visual bookmark and associating a label to it (i.e., by default the name of the annotation). If the annotation includes also an extended description of the object (in Fig. 5 the link to a Wikipedia page), this reference is saved for future uses in the story. The art expert defines also other visual bookmarks, associated to viewpoints derived from the simple navigation of the 3D worlds or from visual queries (Fig. 3). The visual bookmarks are gathered across all the set of 3D worlds that belong to the repository, enabling the expert to define narrations whose stages are localized in different 3D environments. A simple interface, accessible from the main menu of the web application, allows the author to manage the collection of visual bookmarks progressively gathered, allowing him to modify their order, change their labels or delete them. The expert accesses the initial interface for creating the structure of the story from the item “MyStories” of the web main application menu. This interface allows the story creator to define the general data associated to the story: the title, the name of the author and the summary. After filling in the initial data, the story author proceeds to the main interface for creating stories, displayed in (Fig. 6-1). The interface allows the story creators to define - through simple drag-and-drop operations - the sequence of the story stages. The stages, represented in the central panel of the interface, derive from a selection of the visual bookmarks that the author gathered before and that are represented on the left panel. The icons

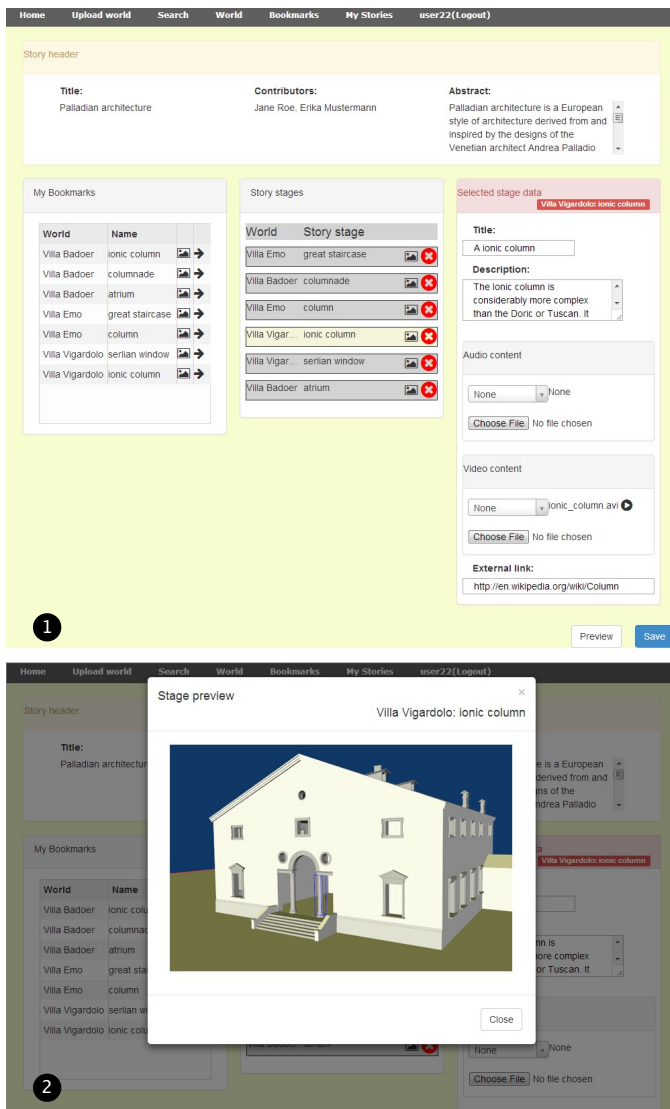


Figure 6. (1) visualizes the interface for managing the creation of the stages. The sequence of stages is created from the list of bookmarks visualized on the left panel. The central panel allows the users to manage the sequence of stages, changing their order, deleting the unwanted ones or showing the associated 3D view. The panel on the right enables the users to associate a title, a short description, an audio-visual track and an extended description to each stage. At the end of the process the stories can be previewed and saved using the buttons on the bottom. (2) shows the 3D view associated to one of the stages.

placed on each stage of the central panel enables the author to have a preview of the associated viewpoints, as displayed by Fig. 6-2. The author then prepares, with a simple webcam, the set of videos that define the narrative content and uploads them using the right panel of the interface, deciding which video to associate to each stage. While some videos contain only narrative content, others contain also instructions for the students. For example, in some of the videos the art expert asks his pupils to navigate the 3D world around the location of the current stage and to tag certain architectural components, recalling the explanations given in the previous stages of the story. He will check later the work done by the students for verifying the acquisition of skills and knowledge. The story

author can decide also to include in the stage the extended descriptions associated to the story annotations (e.g., the link to the Wikipedia page discussed before), for giving the pupils the possibility to read additional information after having listened to the narration. At any time the content author can visualize the result of his authoring effort, clicking the Preview button. The final result is a linear story spanning through a set of stages belonging to different 3D worlds, each one containing the model of a famous building by Andrea Palladio.

## 2) The student listening to a story and doing homework.

The art student uses the ToBoA-3D platform as a tool for remote learning that enables him to learn new concepts through storytelling and to send feedback to his teacher. The student receives an assignment from his teacher, as homework. The student is asked to listen to a linear story focused on the architecture of Andrea Palladio and to complete the assignments specified during the narration. The student accesses the visual interface for listening to stories from the main page of the web application, displayed in Fig. 2. The main page displays on the right panel a list of the last authored stories and of their authors. A full list can be accessed from the button placed at the bottom of the panel. The student, after having selected the story created by his teacher, enters the interface for playing the stories, displayed in Fig. 7. While the left area of the interface represents the 3D environment, the right area is reserved to the presentation of the audio-visual content. The lower part of the screen is reserved to the visualization of the story summary and to the controls for moving through the different stages of the narration. The student presses the “play” icon for starting the narration. The engine plays the stages of the story, following the sequence defined by the teacher during the creation of the narration. The system takes care of downloading the required 3D environments for giving to the user a seamless experience. For each stage the application engine shows the related 3D location or object and plays automatically the associated video content. In Fig. 7 the teacher explains the features of the ionic column, relating it to a well-known building by Andrea Palladio, Villa Vigardolo. At any time the student can stop the automatic play of the story, for moving through the different steps at his own pace or navigating freely the 3D environment. In the example displayed, at the end of the narration of the stage the student is asked to navigate the 3D world around the current location and tag the architectural components around him, recalling the terms used by the teacher during the explanation.

After the completion the teacher will have the opportunity to verify the knowledge acquired by his pupils. For example the teacher might navigate the 3D world and select a given 3D object for visualizing the annotations, which are always associated to a specific author. Alternatively the teacher might search all the objects tagged by a given student, filling in the field “author” on the right panel of the search interface (see Fig. 3).

## VI. CONCLUSION

The work done so far has led to the design and to the implementation of a storytelling system for creating narrations based on a repository of annotated 3D worlds. As far as the usability of the system is concerned, we plan to start an





Figure 7. The interface for browsing stories.

experimentation with end users chosen among different categories, such as content experts, teachers and students, exploring also new educational scenarios characterized by the use of more complex narrative structures. The experimentation will be also an opportunity to have a feedback about the novel workflow for content production and delivery, measuring the users' appreciation and the impact on their activities.

## REFERENCES

- [1] VRML97, "Virtual reality modeling language (VRML97) ISO/IEC IS 14772:1997," <http://www.web3d.org/x3d/specifications/vrml/ISO-IEC-14772-VRML97/>, 2003.
- [2] F. Pittarello and I. Gatto, "ToBoA-3D: an architecture for managing top-down and bottom-up annotated 3d objects and spaces on the web," in *Proc. of Web3D '11*, 2011, pp. 57–65.
- [3] R. Berndt, G. Buchgraber, S. Havemann, V. Settgast, and D. W. Fellner, "A publishing workflow for cultural heritage artifacts from 3d-reconstruction to internet presentation," in *Proc. of EuroMed'10*, 2010, pp. 166–178.
- [4] X3D, "X3D Architecture and Base Components," <http://www.web3d.org/x3d/specifications/ISO-IEC-19775-1.2-X3D-AbstractSpecification/>, 2008.
- [5] J. Behr, P. Eschler, Y. Jung, and M. Zöllner, "X3DOM: a DOM-based HTML5/X3D integration model," in *Proc. of Web3D '09*, 2009, pp. 127–135.
- [6] I. M. Bilasco, J. Gensel, M. Villanova-Oliver, and H. Martin, "On indexing of 3D scenes using MPEG-7," in *Proc. of the 13th Annual ACM International Conference on Multimedia*. ACM Press, 2005, pp. 471–474.
- [7] S. Havemann, V. Settgast, R. Berndt, O. Eide, and D. Fellner, "The Arrigo Showcase reloaded - towards a sustainable link between 3D and semantics," *J. Comput. Cult. Herit.*, vol. 2, no. 1, pp. 4:1–4:13, 2009.
- [8] J. Trant, "Exploring the potential for social tagging and folksonomy in art museums: Proof of concept," *New Review of Hypermedia and Multimedia*, vol. 12, no. 1, pp. 83–105, 2006.
- [9] R. Scopigno, M. Callieri, P. Cignoni, M. Corsini, M. Dellepiane, F. Ponchio, and G. Ranzuglia, "3d models for cultural heritage: Beyond plain visualization," *Computer*, vol. 44, no. 7, pp. 48–55, Jul. 2011.
- [10] F. Garzotto and M. Forfori, "Fate2: Storytelling edutainment experiences in 2d and 3d collaborative spaces," in *Proc. of IDC '06*, 2006, pp. 113–116.
- [11] F. A. Hansen, K. J. Kortbek, and K. Grønbaek, "Mobile urban drama for multimedia-based out-of-school learning," in *Proc. of MUM '10*, 2010, pp. 17:1–17:10.
- [12] K. Kwiatek and M. Woolner, "Let me understand the poetry: Embedding interactive storytelling within panoramic virtual environments," in *Proc. of EVA '10*, 2010, pp. 199–205.
- [13] J. Halloran, E. Hornecker, G. Fitzpatrick, M. Weal, D. Millard, D. Michaelides, D. Cruickshank, and D. De Roure, "The literacy fieldtrip: using ubicomp to support children's creative writing," in *Proc. of IDC '06*, 2006, pp. 17–24.
- [14] D. Pletinckx, N. A. Silberman, and D. Callebaut, "Heritage presentation through interactive storytelling: a new multimedia database approach," *J. of Visualization and Computer Animation*, vol. 14, no. 4, pp. 225–231, 2003.
- [15] S. Sanchez, O. Balet, H. Luga, and Y. Duthen, "Autonomous virtual actors," in *Proc. of TIDSE 2004*. Heidelberg: Springer, 2004, pp. 68–78.
- [16] A. Lamstein and M. Mateas, "Search-based drama management," in *Proc. of the AAAI-04 Workshop on Challenges in Game AI*. AAAI Press, 2004.
- [17] M. Kriegel, R. Aylett, J. Dias, and A. Paiva, "An authoring tool for an emergent narrative storytelling system," in *AAAI Fall Symposium on Intelligent Narrative Technologies*, Technical Report FS-07-05, 2007, pp. 55–62.
- [18] F. Mehm, S. Göbel, S. Radke, and R. Steinmetz, "Authoring environment for story-based digital educational games," in *Proceedings of the 1st International Open Workshop on Intelligent Personalization and Adaptation in Digital Educational Games*, vol. 1, 2009, pp. 113–124.
- [19] J. Robertson and J. Good, "Adventure author: an authoring tool for 3d virtual reality story construction," in *Proceedings of the AIED-05 Workshop on Narrative Learning Environments*, 2005, pp. 63–69.
- [20] F. Barrenho, T. Romão, T. Martins, and N. Correia, "Inauthoring environment: Interfaces for creating spatial stories and gaming activities," in *Proc. of ACE '06*, 2006.
- [21] JGraph, <http://www.jgraph.com/>.
- [22] BS Contact, Bitmanagement Interactive Web3D Graphics, <http://www.bitmanagement.com/>.
- [23] F. Pittarello and I. Gatto, "A visual interface for querying ontologically and socially annotated 3d worlds for the web," in *Proc. of AVI '12*, 2012, pp. 377–381.
- [24] P. Morville and L. Rosenfeld, *Information Architecture for the World Wide Web*. O'Reilly, 2006.
- [25] AngularJS Javascript MVW Framework, <http://angularjs.org/>.
- [26] Bootstrap Frontend Framework, <http://getbootstrap.com>.

# Terminological Ontology Learning and Population using Latent Dirichlet Allocation

Francesco Colace and Massimo De Santo  
Department of Information Technology  
and Electronics Engineering  
University of Salerno, ITALY  
Email: {fcolace,desanto}@unisa.it

Luca Greco  
Department of  
Industrial Engineering  
University of Salerno, ITALY  
Email: lgreco@unisa.it

Vincenzo Moscato and Antonio Picariello  
Department of  
Informatica e Sistemistica  
University of Naples "Federico II", ITALY  
Email: {vmoscato,picus}@unisa.it

**Abstract**—The success of Semantic Web will heavily rely on the availability of formal ontologies to structure machine understanding data. However, there is still a lack of general methodologies for ontology automatic learning and population, i.e. the generation of domain ontologies from various kinds of resources by applying natural language processing and machine learning techniques. In this paper, the authors present an ontology learning and population system that combines both statistical and semantic methodologies. Several experiments have been carried out, demonstrating the effectiveness of the proposed system.

**Keywords**—Ontologies, Ontology Learning, Ontology Population, Latent Dirichlet Allocation

## I. INTRODUCTION

In the last decade many researchers have been involved in the development of methodologies for ontology definition, building, learning and population, due to the fact that ontologies are considered as an effective answer to the need of semantic interoperability among modern information systems: it is well known, in fact, that ontologies are the backbone of the Semantic Web and important means for sharing, reusing and reasoning about domain knowledge. Several theories have been developed, in different application domains and especially in the semantic web framework: however how to learn and populate ontologies is generally a not trivial and time consuming task and still remains an open research challenge.

The term “ontology learning” was introduced in [21] and can be described as the acquisition of a domain model from data. This process is historically connected to the introduction of the semantic web and needs input data from which to learn the concepts relevant for a given domain, their definitions as well as the relations holding between them. Ontologies can be learnt from various sources, be it databases, structured and unstructured documents or even existing preliminaries like dictionaries, taxonomies and directories.

With the explosion of information due to the Read/Write Web, ontology learning from text is becoming the most investigated in literature: ontology learning from text is the process of identifying terms, concepts, relations, and axioms from textual information and of using them in

order to construct and maintain ontology [28]. In other words ontology learning from text is the process of deriving high level *concepts* and *relations* as well as *axioms* from information to form ontology.

Ontology Learning from text is generally composed by five phases that aim at returning five main outputs: *terms*, *concepts*, *taxonomic relations*, *non-taxonomic relations* and *axioms* [5].

To obtain each output, some tasks have to be accomplished and the techniques employed for each task may change among systems. In this sense the ontology learning process is really modular: in [28] the corresponding tasks and the plethora of employed techniques for each output are described.

The extraction of terms from text usually needs a *preprocessing phase* that arranges the text in the correct format for an ontology learning system and, generally, includes noisy text analytics. The extraction of terms begins with the tokenization or part of speech tagging to break texts into smaller constituents. In this phase, statistical or probabilistic measures are adopted for determining the “unithood”, the collocational stability of a noun sequence, and the termhood, the relevance or the specificity of a term with respect to a domain. Starting from the terms is possible to derive the concepts that can be formed by grouping similar terms and labeling them. The grouping phase involves discovering the variants of a term and grouping them together, while the concept’s label can be inferred by the use of existing background knowledge, such as WordNet, that may be used to find the name of the nearest common ancestor.

The relations model the interactions among the concepts in ontology: in general, two types of relations can be recognized in ontology: taxonomic and non-taxonomic relations. Taxonomic relations, that are hypernym, build hierarchies and can be labeled as “is-a” relations [8]. This kind of relations can be performed in various ways such as using predefined relations from existing background knowledge, using statistical subsumption models, relying on semantic similarity between concepts and utilizing linguistic and logical rules or patterns. The non-taxonomic relations are the interactions among the concepts other than hypernymy and their extraction is a challenging task. In this context

verbs play a significant role such as the support of domain experts.

Axioms are propositions or sentences that are always taken as true and are the starting point for deducing other truth, verifying the correctness of the ontological elements and defining constraints. The process of learning axioms is still complex and there are few examples in the literature.

Ontology Learning has adopted well known techniques coming from different fields such as information retrieval, machine learning, and natural language processing [9]. This techniques can be generally classified into *statistics-based*, *linguistic-based*, *logic-based* or *hybrid* [28].

The statistics-based techniques are derived from information retrieval, machine learning, data mining and work at a syntactical level. In particular, these approaches are effective in the early stages of ontology learning, such as term extraction and hierarchy construction [4]. Some of the common techniques include clustering [27], Latent Semantic Analysis [25], term subsumption [17] and contrastive analysis [26]. The main idea behind these techniques is that the co-occurrence of words provides a reliable estimate about their semantic identity. In this way a concepts can be inferred.

The linguistics-based techniques can support all tasks in ontology learning and are based on natural language processing tools. In general some of the techniques include Part of Speech (POS) tagging, such as [2], syntactic structure analysis [19] and dependency analysis [7]. Other adopted techniques are related to semantic lexicon [22], lexico-syntactic patterns [5][24], subcategorization frames [18] and seed words [30].

The logic-based techniques and resources are the least common in ontology learning and are mainly adopted for more complex tasks involving relations and axioms. The two main techniques employed are inductive logic programming [20] and logical inference [23]. In the inductive logic programming, rules are derived from existing collection of concepts and relations which are divided into positive and negative examples. In logical inference, implicit relations are derived from existing ones using rules (transitivity and inheritance). In general it is difficult to say what of these techniques is the better one and, maybe, none of them is the only solution for the ontology learning.

As previously said, each phase of the ontology learning process can adopt one of these approaches in order to maximize the process effectiveness. In particular the terms and concepts extraction can be performed by the use of the statistics-based techniques while the inference of relations can be obtained by the use of linguistic and logic based techniques. In reality, the hybrid approach is mainly used in the existing studies and furnishes the best results [28][9].

Differently from the other described papers in ontology learning and population, that usually produces concept hierarchies by means of statistical and/or probabilistic methods (LSA, LDA, pLSA and so on), we enrich our terminological ontologies with the semantic features presented in general purpose lexical ontologies, such as WordNet. The use of

both statistical and semantic techniques allows to have suitable and effective domain ontologies particularly suitable for a number of applications such as topic detection and tracking, opinion and sentiment analysis, text mining and classification [11], [15], [12].

The paper is organized as follows. Section 2 describes at glance our Ontology Learning and Population system architecture. Section 3 is devoted to a general description of the adopted methods and algorithms. Experiments and conclusions are reported in section IV and V respectively.

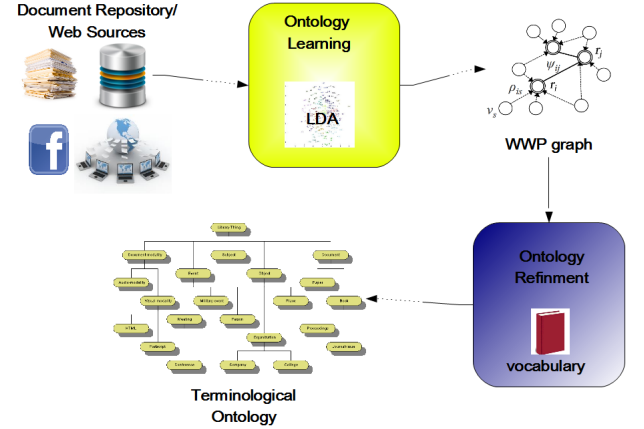


Fig. 1. The System Architecture.

## II. SYSTEM ARCHITECTURE

Figure 1 describes the proposed system architecture. The system analyzes a number of documents coming from different web sources or from document collections related to a given domain of interests classified into a set of semantically disjoint topics. The system is formed by two main components:

- *Ontology Learning* Component, that uses *Latent Dirichlet Analysis* (LDA) on the input documents and produces a *Weighted Word Pairs* (WWP) representation containing the most relevant domain concepts and their co-occurrence values (relations) in the analyzed set.
- *Ontology Refinement* Component, that using general purpose or domain-specific lexical databases, refines the previous discovered concepts, exploiting their lexical relationships (e.g. *is\_a* taxonomic relations), adding hidden concepts, and producing the final ontology schema and population.

In the following we will discuss into details the basic components of the proposed architecture.

## III. TERMINOLOGICAL ONTOLOGY BUILDING

### A. Ontology Learning: concepts and relation extraction

In this section we explain how a WWP structure (Weighted Word Pairs) can be extracted from a corpus of

documents.

The Feature Extraction module (FE) is represented in Fig.2. The input of the system is the set of documents:

$$\Omega_r = (\mathbf{d}_1, \dots, \mathbf{d}_M)$$

After the pre-processing phase, which involves tokenization, stopwords filtering and stemming, a Term-Document Matrix is built to feed the Latent Dirichlet Allocation (LDA) [3], [14]. The LDA algorithm, assuming that each document is a mixture of a small number of latent topics and each word's creation is attributable to one of the document's topics, provides as output two matrices -  $\Theta$  and  $\Phi$  - which express probabilistic relations between topic-document and word-topic respectively. Under particular assumptions [13], [10], [14], LDA module's results can be used to determine: the probability for each word  $v_i$  to occur in the corpus  $W_A = \{P(v_i)\}$ ; the conditional probability between word pairs  $W_C = \{P(v_i|v_s)\}$ ; the joint probability between word pairs  $W_J = \{P(v_i, v_j)\}$ . Details on LDA and probability computation are discussed in [3], [16], [13].

Defining *Aggregate roots* (AR) as the words whose occurrence is most implied by the occurrence of other words of the corpus, a set of  $H$  aggregate roots  $\mathbf{r} = (r_1, \dots, r_H)$  can be determined from  $W_C$ :

$$r_i = \operatorname{argmax}_{v_i} \prod_{j \neq i} P(v_i|v_j) \quad (1)$$

This phase is referred as Root Selection (RS) in Fig.2. A weight  $\psi_{ij}$  can be defined as a degree of probabilistic correlation between AR pairs:  $\psi_{ij} = P(r_i, r_j)$ . We define an *aggregate* as a word  $v_s$  having a high probabilistic dependency with an aggregate root  $r_i$ . Such a dependency can be expressed through the probabilistic weight  $\rho_{is} = P(r_i|v_s)$ . Therefore, for each aggregate root, a set of aggregates can be selected according to higher  $\rho_{is}$  values. As a result of the Root-Word level selection (RWL), an initial WWP structure, composed by  $H$  aggregate roots ( $R_i$ ) linked to all possible aggregates ( $W_i$ ), is obtained. An optimization phase allows to neglect weakly related pairs according to a fitness function discussed in [13]. Our algorithm, given the number of aggregate roots  $H$  and the desired max number of pairs as constraints, chooses the best parameter settings  $\tau$  and  $\mu = (\mu_1, \dots, \mu_H)$  defined as follows:

- 1)  $\tau$ : the threshold that establishes the number of *aggregate root/aggregate root* pairs. A relationship between the aggregate root  $v_i$  and aggregate root  $r_j$  is relevant if  $\psi_{ij} \geq \tau$ .
- 2)  $\mu_i$ : the threshold that establishes, for each aggregate root  $i$ , the number of *aggregate root/word* pairs. A relationship between the word  $v_s$  and the aggregate root  $r_i$  is relevant if  $\rho_{is} \geq \mu_i$ .

Note that a WWP structure can be suitably represented as a *graph*  $\mathbf{g}$  of terms (Fig. 3). Such a graph is made of several clusters, each containing a set of words  $v_s$  (*aggregates*) related to an *aggregate root* ( $r_i$ ), the centroid of the cluster.

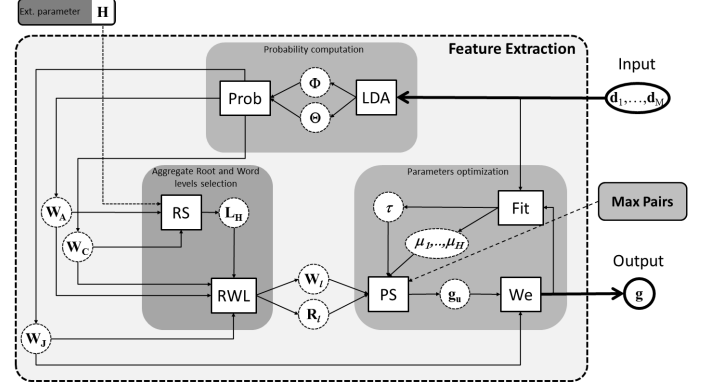


Fig. 2. Proposed feature extraction method. A WWP  $\mathbf{g}$  structure is extracted from a corpus of training documents.

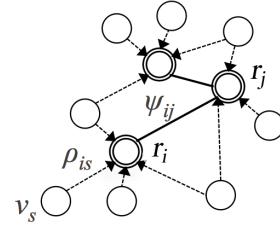


Fig. 3. Graphical representation of a WWP structure.

*Aggregate roots* can be also linked together building a centroids subgraph.

### B. Ontology Refinement

The main goal of such a Component is to transform, for each topic, the WWP graphs into a terminological ontology, in order to represent and manage the knowledge coming from the document corpus in a more effective way. In the following, we will introduce some preliminary definitions and describe an algorithm for automatically building the described ontologies.

We first introduce the concept of a *Semantic Node* ( $w$ ), as a triple  $w = \langle s, REL, t \rangle$ ,  $s$  being the code of a given vocabulary synset,  $REL$  is the set of references to the other nodes and  $t$  the related concept label. A *Local Terminological Ontology* is a particular graph data structure  $T = \langle w^*, W \rangle$ ,  $w^*$  being the aggregate root node and  $W$  the set of the other semantic nodes. A *Domain Terminological Ontology* is a particular graph data structure  $T = \langle W^*, W \rangle$ ,  $W^*$  being set of the aggregate root nodes for a given semantic domain/topic and  $W$  the set of the other semantic nodes.

If we consider only *is\_a* relationships among concepts, each semantic node is characterized by a single IS\_A reference to the ancestor node and vice-versa, while a terminological ontology corresponds to a concepts taxonomy.

Algorithm 1 allows to build the local terminological ontology constituted by the taxonomy of discovered concepts for a single topic and for a given aggregate root node

and by a set of generic relationships among the root node and other semantic nodes, using *WordNet* as general lexical vocabulary. The algorithm has in input the WWP graph and in particular considers the aggregate root node and a set of aggregated words (we also consider words containing the aggregated words) for a given topic/domain. In a first phase, the common hypernyms between the aggregate root node and aggregated words are determined and eventually added to the ontology as semantic nodes if they are semantically similar to the root node. In a second phase, ontology is updated by computing the correct IS\_A relationships among the concepts, corresponding to the ancestor and leave nodes.

The following functions are exploited by the algorithm.

- **find\_synset**( $t, WN$ ) - returns all possible WordNet synsets for a generic word  $t$ .
- **add\_vector**( $v_t^1, v_t^2$ ) - adds a list of words to a vector of words.
- **find\_composite\_words**( $t, WN$ ) - returns the set of the words that contain a given word  $t$  in the WordNet database.
- **find\_minimum\_common\_hypernym**( $s, S, WN$ ) - returns the minimum common ancestor with the related synset between a synset  $s$  and a set of synset  $S$  in the WordNet hierarchy.
- **add\_ontology**( $w, T$ ) - adds a new node  $w$  to the taxonomy  $T$  and links the node with the related ancestor.
- **exists**( $w, T$ ) - returns true if a node with the same synset is already contained in the ontology  $T$ .
- **collapse**( $v_t, WN$ ) - modifies a vector of words collapsing the synonymous words in a unique term.
- **leaves\_ontology**( $T$ ) - returns the leaf words in the ontology  $T$ .
- **ancestors\_ontology**( $T$ ) - returns the ancestor words in the ontology  $T$ .
- **is\_parent**( $s_i, s_j, WN$ ) - returns true if the synset  $s_j$  is an ancestor of synset  $s_i$  in the WordNet hierarchy.
- **update\_ontology**( $w_{old}, w_{new}, T$ ) - updates a node in the ontology  $T$ .
- **ordering**( $v, WN$ ) - performs an ordering of a vector of words on the base of the depth in the WordNet hierarchy.
- **semantic\_distance**( $s_i, s_j, D$ ) - computes a semantic distance between two synsets using the Wu & Palmer metric [29] based on a domain specific dictionary  $D$ .

Eventually, generic relationships, whose semantics cannot be retrieved as IS\_A relation in the WordNet vocabulary, are instantiated between root and aggregated nodes that appear in the concepts' taxonomy.

The previous algorithm is then iteratively repeated for each aggregate word in the considered domain and the obtained local ontologies are opportunely aligned and merged in a single domain terminological ontology, exploiting ontology-mapping techniques [6].

Figure 4 reports an example of terminological ontology (i.e., taxonomy of concepts with the related WordNet Synsets)

---

**Algorithm 1** Local Terminological Ontology Building

---

**Input:**  $v_t = [t_1, t_2, \dots, t_n]$ , a vector of aggregated words;  $\langle \hat{t}, \hat{s} \rangle$ , the considered aggregate root node with the related synset  $\hat{s}$ ,  $\gamma$  a given threshold.

**Output:**  $T$ , a Terminological Ontology.

*Computing of the composite words that contain the input words*

**for**  $k = 1 \rightarrow n$  **do**

$\tilde{v}_t = \text{find\_composite\_words}(v_t[k], WN)$ ;

**add\_vector**( $v_t, \tilde{v}_t$ );

**end for**

*Computing of the common ancestors*

**for**  $i = 1 \rightarrow m$  **do**

$v_s = \text{find\_synset}(v_t[i])$ ;

$\langle t^*, s^* \rangle = \text{find\_minimum\_common\_hypernym}(v_s[i], \hat{s}, WN)$ ;

**if** ( $t^* \neq \text{Entity} \wedge \text{semantic\_distance}(s^*, \hat{s}) \leq \gamma$ ) **then**

$w^* = \langle s^*, \text{NULL}, t^* \rangle$ ;

**if** ( $\neg \text{exists}(w^*, T)$ ) **then**

**add\_ontology**( $w^*, T$ );

**end if**

$\hat{w} = \langle \hat{s}, \text{rel}_{ISA}(w^*), \hat{t} \rangle$ ;

**if** ( $\neg \text{exists}(\hat{w}, T)$ ) **then**

**add\_ontology**( $\hat{w}, T$ );

**end if**

$w_t = \langle v_s[i], \text{rel}_{ISA}(w^*), v_t[i] \rangle$ ;

**if** ( $\neg \text{exists}(w_t, T)$ ) **then**

**add\_ontology**( $w_t, T$ );

**end if**

**end if**

**end for**

*Updating of the ontology*

$v_w^a = \text{ancestors\_ontology}(T)$ ;

**ordering**( $v_w^a, WN$ );

**for**  $i = 1 \rightarrow \text{length}(v_w^a) - 1$  **do**

**for**  $j = i + 1 \rightarrow \text{length}(v_w^a)$  **do**

**if** ( $\text{is\_parent}(v_w^a[i].s, v_w^a[j].s, WN)$ ) **then**

**update\_ontology**( $v_w^a[i]$ ,

$\langle v_w^a[i].s, \text{rel}_{ISA}(v_w^a[j]), v_w^a[i].t \rangle, T$ );

**end if**

**end for**

**end for**

$v_w^l = \text{leaves\_ontology}(T)$ ;

**ordering**( $v_w^l, WN$ );

**for**  $i = 1 \rightarrow \text{length}(v_w^l) - 1$  **do**

**for**  $j = i + 1 \rightarrow \text{length}(v_w^l)$  **do**

**if** ( $\text{is\_parent}(v_w^l[i].s, v_w^l[j].s, WN)$ ) **then**

**update\_ontology**( $v_w^l[i]$ ,

$\langle v_w^l[i].s, \text{rel}_{ISA}(v_w^l[j]), v_w^l[i].t \rangle, T$ );

**end if**

**end for**

**end for**

---



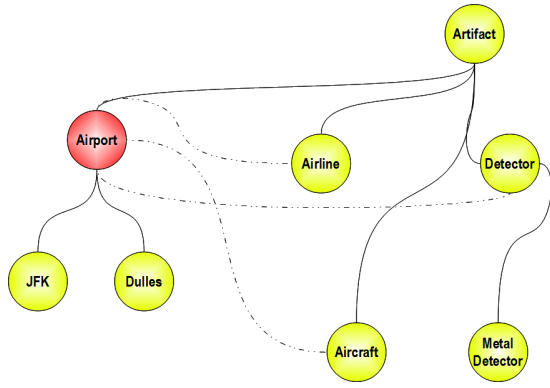


Fig. 4. An example of terminological ontology.

obtained by the algorithm application considering as root concept *airport* and as aggregated concepts *metal detector*, *airline*, *JFK*, *Dulles*. The relationships marked with broken line are generic relations, the other ones correspond to IS\_A relationships.

#### IV. PRELIMINARY EXPERIMENTAL RESULTS

Evaluating the “quality” of an ontology is an important issue both for ontology developers and for final users: this task allows to compare different ontologies describing the same domain in order to choose the more suitable one for a given application.

The dataset from TREC-8<sup>1</sup> collections (minus the Congressional Record) was used for performance evaluation. It contains about 520,000 news documents on 50 topics (no.401-450) and relevance judgements for the topics. Word stopping and word stemming with single keyword indexing were performed before building WWP.

In particular, we have selected for the experiments a subset of TREC documents related to the following topics:

- **osteoporosis** (21 documents),
- **cosmic events** (18 documents),
- **tropical storms** (118 documents),
- **airport security** (123 documents),
- **heroic acts** (94 documents),
- **robotic technology** (130 documents),
- **UV damage, eyes** (50 documents),
- **creativity** (75 documents),
- **counterfeiting money** (162 documents),
- **drugs, Golden Triangle** (136 documents).

For each topic, our system produced a terminological ontology: we have thus asked a set of users’ groups to generate ontologies containing the concepts returned by the *Concept Extraction* component. The users have been grouped on the base of their expertise in the domain topics (namely high, low and medium experts). Each group was formed by about 10 students, for a total of about 200 users involved in the experimental process.

In particular, we have evaluated the *effectiveness* of our generated terminological ontologies with respect to humans on the base of the following several criteria: *Class Match Measure* (CMM), *DEnsity measure* (DEM), *Semantic Similarity Measure* (SSM), *BETweenness Measure* (BEM) as in [1].

The *Class Match Measure* is meant to evaluate the “coverage” of an ontology for the given search terms. This measure evaluates class instances (high-level nodes in our case) in each ontology having labels matching a search term either exactly (node label identical to search term) or partially (node label contains the search term). An ontology that contains all search terms will obviously score higher than others, and exact matches are considered better than partial matches. The CMM can be obtained by the following equation:

$$CMM(\mathcal{O}, T) = \alpha \cdot \sum_{c \in C(\mathcal{O})} \sum_{t \in T} I(c, t) + \beta \cdot \sum_{c \in C(\mathcal{O})} \sum_{t \in T} J(c, t) \quad (2)$$

where:  $\mathcal{O}$  is the assigned ontology,  $C(\mathcal{O})$  is the set of the high level nodes,  $T$  is the set of search terms, and  $I(c, t)$  and  $J(c, t)$  are two binary functions that return 1 in the case of a generic concept  $c$  of the ontology matching or containing a search term  $t$  respectively, 0 otherwise. For what the CMM metric computation concerns, for each topic we used as search terms the query keywords provided by TREC.

The *DEnsity Measure* is a metric that tries to measure the “representational density” or “informative content” of classes and consequently the level of knowledge detail. Density calculations are currently limited to the numbers of relations, subclasses, superclasses, and siblings for the different high-level nodes. We dropped the number of instances from this measure as this might skew the results unfairly towards populated ontologies which may not necessarily reflect the quality of the schema. The DEM can be obtained by the following equation:

$$DEM(\mathcal{O}) = \frac{1}{n} \cdot \sum_{i=1}^n \frac{rel(c_i) + sup(c_i) + sub(c_i) + sibl(c_i)}{M} \quad (3)$$

where:  $\mathcal{O}$  is the assigned ontology,  $n$  is the number of matched high-level nodes respect to the search terms in the ontology,  $M$  is a normalization factor,  $rel(c)$ ,  $sup(c)$ ,  $sub(c)$ ,  $sibl(c)$  are apposite functions returning the number of relations, subclasses, superclasses, and siblings, respectively, of a generic concepts  $c_i$ .

The *Semantic Similarity Measure* calculates how close the classes that matches the search terms are in an ontology. The motivation for this is that ontologies whose position concepts further away from each other are less likely to represent the knowledge in a coherent and compact manner. The SSM formula used here is based on the *Rada Shortest Path* measure. SSM is measured from the minimum number

<sup>1</sup><http://trec.nist.gov/>



of links that connects a pair of concepts. The SMM can be obtained by the following equation:

$$SSM(\mathcal{O}) = \frac{1}{m} \cdot \sum_{c_i, c_j \in C(\mathcal{O})} sim(c_i, c_j) \quad (4)$$

where:  $\mathcal{O}$  is the assigned ontology,  $m$  is the number of matchings for matched classes respect to the search terms in the ontology,  $sim(c_i, c_j)$  is an apposite function returning the similarity (computing by the Rada measure) between two connected concepts  $c_i, c_j$ .

The *BETweenness Measure* calculates the number of the shortest paths that pass through each couple of matched high-level nodes (*betweenness*) in the ontology. The nodes occurring on many shortest paths among other nodes have higher betweenness value than others. The assumption is that if a class instance has a high betweenness value in an ontology then this node is central to that ontology. Ontologies where those classes are more central will receive a higher score. The BEM can be obtained by the following equation:

$$BEM(\mathcal{O}) = \frac{1}{n} \cdot \sum_{k=1}^n \sum_{c_i \neq c_j} \frac{\sigma_{c_i c_j}(c_k)}{\sigma_{c_i c_j}} \quad (5)$$

where:  $\mathcal{O}$  is the assigned ontology,  $n$  the number of matched high-level nodes respect to the search terms in the ontology,  $\sigma_{c_i c_j}$  and  $\sigma_{c_i c_j}(c_k)$  are apposite functions returning the shortest path from  $c_i$  to  $c_j$  and the number of shortest paths from  $c_i$  to  $c_j$  that passes through a generic concept  $c_k$  respectively.

As we can see from Figure 5, our ontology has a quality index very close to that of an ontology generated by experts on the considered domain.

Finally, we measured the times of building a terminological ontology depending on the number of input documents for each topic<sup>2</sup>. We observed that the *drugs*, *Golden Triangle* ontology, built from 136 documents, requires less than 30 seconds for its complete building, thus ensuring enough scalability on more large data set.

## V. CONCLUSIONS AND FUTURE WORK

In this paper, we presented an ontology learning and population technique that exploit both statistical and semantic methodologies for generating terminological ontologies for a given semantic domain. The reported experimental results demonstrated the effectiveness of the proposed system in terms of goodness and quality of produced ontologies with respect to the ones manually generated by experts or less humans on the considered domain. Thus, the automatic generated ontologies can be suitably used topic for different application such as semantic-based retrieval, topic detection and tracking, sentiment analysis and so on.

<sup>2</sup>we use a Linux Ubuntu platform running on a 8GB RAM single CPU

## REFERENCES

- [1] Harith Alani and Christopher Brewster. Metrics for ranking ontologies. In *4th Int. EON Workshop, 15th Int. World Wide Web Conf.*, 2006.
- [2] Steven Bird, Ewan Klein, Edward Loper, and Jason Baldridge. Multidisciplinary instruction with the natural language toolkit. In *Proceedings of the Third Workshop on Issues in Teaching Computational Linguistics*, TeachCL '08, pages 62–70, 2008.
- [3] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.
- [4] Christopher Brewster, Simon Jupp, Joanne Luciano, David Shotton, Robert Stevens, and Ziqi Zhang. Issues in learning an ontology from text. *BMC Bioinformatics*, 10(Suppl 5):S1, 2009.
- [5] Paul Buitelaar and Bernardo Magnini. Ontology learning from text: An overview. In *In Paul Buitelaar, P., Cimiano, P., Magnini B. (Eds.), Ontology Learning from Text: Methods, Applications and Evaluation*, pages 3–12. IOS Press, 2005.
- [6] Namyoun Choi, Il-Yeol Song, and Hyeon Han. A survey on ontology mapping. *SIGMOD Rec.*, 35(3):34–41, September 2006.
- [7] Massimiliano Ciarmita, Aldo Gangemi, Esther Ratsch, Jasmin Šarić, and Isabel Rojas. Unsupervised learning of semantic relations between concepts of a molecular biology ontology. In *Proceedings of the 19th international joint conference on Artificial intelligence, IJCAI'05*, pages 659–664, 2005.
- [8] Philipp Cimiano, Aleksander Pivk, Lars Schmidt-Thieme, and Steffen Staab. Learning taxonomic relations from heterogeneous evidence.
- [9] Philipp Cimiano, Johanna Viker, and Rudi Studer. Ontologies on demand a description of the state-of-the-art, applications, challenges and trends for ontology learning from text, 2006.
- [10] F. Clarizia, F. Colace, M. De Santo, L. Greco, and P. Napoletano. Mixed graph of terms for query expansion. In *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*, pages 581–586, 2011.
- [11] F. Clarizia, F. Colace, M. De Santo, L. Greco, and P. Napoletano. A new text classification technique using small training sets. In *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*, pages 1038–1043, Nov 2011.
- [12] Francesco Colace, Massimo De Santo, and Luca Greco. A probabilistic approach to tweets' sentiment classification. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 37–42, 2013.
- [13] Francesco Colace, Massimo De Santo, and Luca Greco. Weighted word pairs for text retrieval. In *Proceedings of the 3rd Italian Information Retrieval (IIR), volume 964 of CEUR Workshop Proceedings*, 2013.
- [14] Francesco Colace, Massimo Santo, Luca Greco, and Paolo Napoletano. Improving text retrieval accuracy by using a minimal relevance feedback. In Ana Fred, JanL.G. Dietz, Kecheng Liu, and Joaquim Filipe, editors, *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, volume 348 of *Communications in Computer and Information Science*, pages 126–140. Springer Berlin Heidelberg, 2013.
- [15] Francesco Colace, Massimo De Santo, and Luca Greco. An adaptive product configurator based on slow intelligence approach. *International Journal of Metadata, Semantics and Ontologies*, 9(2):128–137, 01 2014.
- [16] Francesco Colace, Massimo De Santo, Luca Greco, and Paolo Napoletano. Text classification using a few labeled examples. *Computers in Human Behavior*, 30:689–697, 2014.
- [17] Hermine Njike Fotzo and Patrick Gallinari. Learning generalization/specialization relations between concepts application for automatically building thematic document hierarchies.
- [18] Pablo Gamallo, Alexandre Agustini, and Gabriel P. Lopes. Learning subcategorisation information to model a grammar with "co-restrictions", 2003.
- [19] Andrew Hippiusley, David Cheng, and Khurshid Ahmad. The head-modifier principle and multilingual term extraction. *Nat. Lang. Eng.*, 11(2):129–157, June 2005.
- [20] Nada Lavrac and Saso Dzeroski. New York.
- [21] Alexander Maedche and Steffen Staab. Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2):72–79, March 2001.

- [22] Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. Wordnet::similarity: measuring the relatedness of concepts. In *Demonstration Papers at HLT-NAACL 2004*, HLT-NAACL-Demonstrations '04, pages 38–41, 2004.
- [23] Mehrnoush Shamsfard and Ahmad Abdollahzadeh Barforoush. The state of the art in ontology learning: a framework for comparison. *Knowl. Eng. Rev.*, 18(4):293–316, December 2003.
- [24] Rion Snow. Semantic taxonomy induction from heterogenous evidence. In *Proceedings of COLING/ACL 2006*, pages 801–808, 2006.
- [25] Peter D. Turney. Mining the web for synonyms: Pmi-ir versus lsa on toefl. In *Proceedings of the 12th European Conference on Machine Learning*, EMCL '01, pages 491–502, 2001.
- [26] Paola Velardi, Roberto Navigli, Alessandro Cucchiarelli, and Francesca Neri. Evaluation of ontolearn, a methodology for automatic learning of ontologies. In Paul Buitelaar, Philipp Cimiano, and Bernardo Magnini, editors, *Ontology Learning from Text: Methods, Evaluation and Applications*, pages 92–105. IOS Press.
- [27] Wilson Wong, Wei Liu, and Mohammed Bennamoun. Tree-traversing ant algorithm for term clustering based on featureless similarities. *Data Min. Knowl. Discov.*, 15(3):349–381, December 2007.
- [28] Wilson Wong, Wei Liu, and Mohammed Bennamoun. Ontology learning from text: A look back and into the future. *ACM Comput. Surv.*, 44(4):20:1–20:36, September 2012.
- [29] Zhibiao Wu and Martha Palmer. Verb semantics and lexical selection. In *32nd. Annual Meeting of the Association for Computational Linguistics*, pages 133 –138, New Mexico State University, Las Cruces, New Mexico, 1994.
- [30] Roman Yangarber, Ralph Grishman, and Pasi Tapanainen. Automatic acquisition of domain knowledge for information extraction. In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 940–946, 2000.

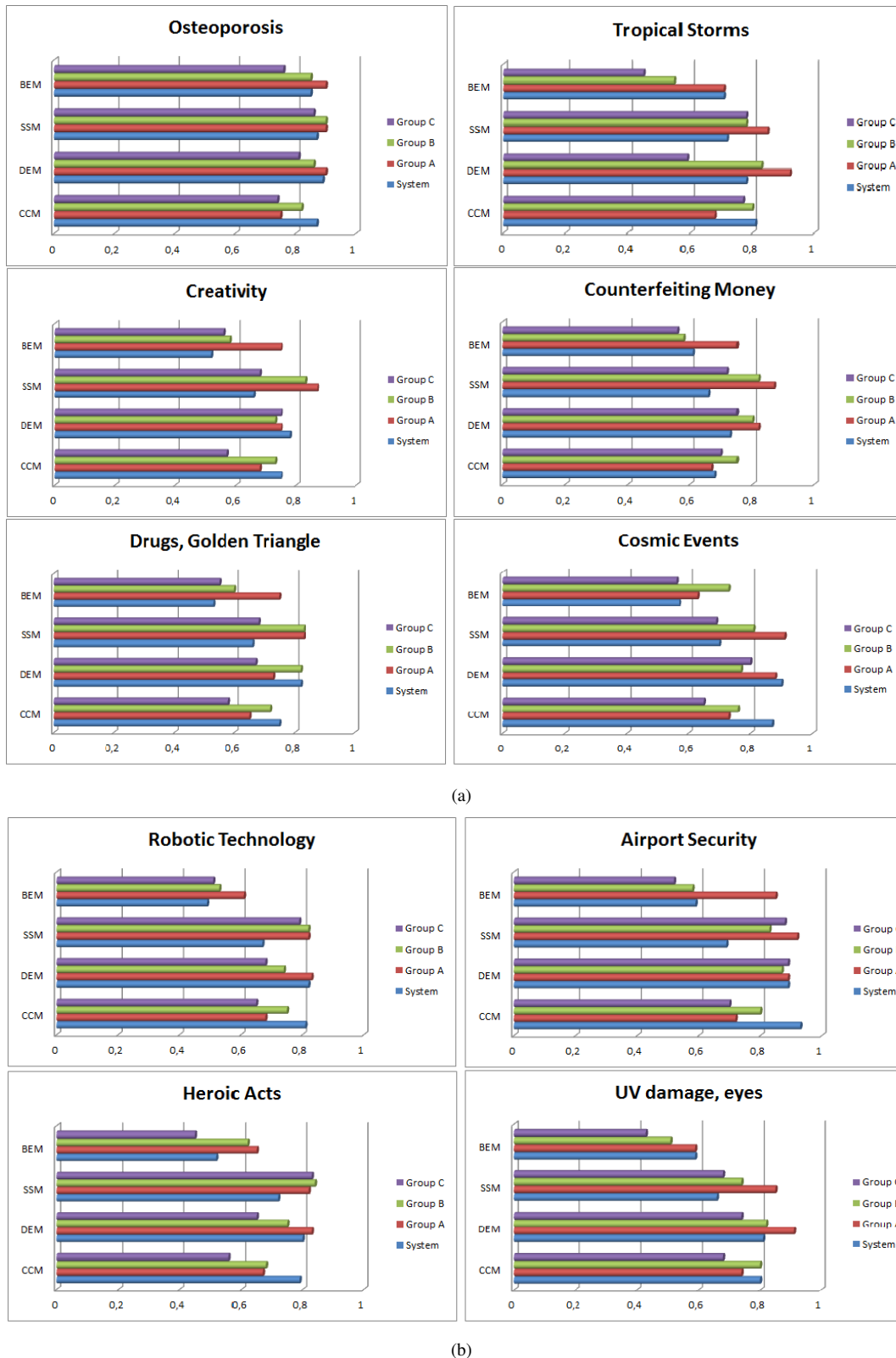


Fig. 5. Experimental Results.

# A Formal Model for Intellectual Relationships among Knowledge Workers and Knowledge Organizations

Mao-Lin Li<sup>1</sup> and Shi-Kuo Chang<sup>1</sup>

<sup>1</sup>Department of Computer Science, University of Pittsburgh, Pittsburgh, PA, USA  
{ moli, chang }@cs.pitt.edu

*Abstract*—Academic learning network is consisted of multiple knowledge organizations and knowledge workers. The degree of relationship can be derived from the interaction within them. In this paper, we propose a formal ownership model to describe the interactions within academic learning network and further provide an evaluation process to quantify the degree of relationship. Proposed approach is also integrated with realistic social platform SMNET.

**Keywords**- Social-Network, Formal Model, Knowledge Agents

## I. INTRODUCTION

The analysis of social-network relationships becomes a critical issue in recent years. People build their own network and connect to others by sharing their life (e.g. Photos, Videos, Articles, etc...) on specific platforms. (e.g. Google+, Facebook, etc...). These platforms analyse the information provided by users and further discover potential relationships among users. Numerous approaches including graph theory, information retrieval or machine learning techniques are used to evaluate the relationships among users and groups in conventional social-networks. Such relationships will also be significant metrics for commercial purpose (e.g. in advertisement and in recommendation systems).

An academic learning network can be considered as another type of social network. An academic learning network consists of knowledge workers and knowledge organizations. Workers/Organizations publish/upload their papers, view/download others' publications. Workers can cooperate with others in an organization for the same research goal or compete with other research organization. However, there are few discussions about how to model and analyse these intellectual relationships among different academic organizations.

In this paper, we propose a formal model to describe the intellectual relationships of completed academic network. Combined with existed social graph platform, the relationships within different people/groups in academic network can be visualized and further provide more information about academic social network.

Our main contributions consist of three parts:

1. Propose a formal ownership model to describe completed academic networks.

2. Combine with social graph to visualize intellectual relationships.
3. Provide significant and flexible characteristic matrices to evaluate relationships without using sophisticated algorithm.

The intellectual relationships would allow us to minimally recognize who are the collaborators and who are the competitors in an academic learning network. Slow Intelligence principles [2] can then be applied so that the academic learning network can achieve its goals.

The paper is organized as follows: In Section 2 we present the ownership model in academic learning network. The social graph will be presented in Section 3.

## II. RELATED WORK

Social network analysis use different metrics (measure) as a reference for evaluating the relationships within various individuals in social network, e.g. Centrality, degree and closeness [6]. Most of these metrics are based on the situation of nodes and link in social network and then using graph theory to analyze the relationships. In proposed model, in addition to above static metrics, we further consider various dynamic actions within different situations in academic learning network.

Some analysis software like UCInet [9] helps research to explore various size of social network with visualization and further investigation. Our ownership model is applied in a realistic academic learning network SMNET [4], SMNET integrated with Intellectual Property Rights model (IPR) model to formalize an institutional aggregator for metadata and content. In addition to traditional social graph, SMNET can represent hierarchical level in academic learning network, from individual worker to academic organization Users can describe their network with our formal descriptions, and upload their contents to the academic learning network, the content could be academic publications, technical reports the completed social graph can be generated automatically.

Relationships evaluation in academic learning network. Many researches [7,8] use bibliometric approach like citation and co-citation analysis to estimate the relationships in the network, the result can be used for finding strategic alliance or identifying the structure of invisible colleges. Our formal model also can include citation and co-citation metrics, since the source of above two metrics can be obtained by "references" part in a publication. Furthermore, our model can

evaluate the degree of cooperative and competitive in academic learning network.

### III. OWNERSHIP MODEL

In this section, we first introduce the concept of academic learning network. Then we present the definition of proposed ownership model and demonstrate a completed academic network with proposed ownership model.

#### A. Academic Learning Network

Fig. 1 shows an example of academic learning network. Knowledge workers ( $KW_1$  and  $KW_2$ ) publish their academic papers ( $P_1$  to  $P_3$ ) or view/download academic paper from others' publications ( $KW_3$ ). A knowledge organization ( $KO_1$ ) can have multiple knowledge workers ( $KW_1$  and  $KW_2$ ). Workers have the ownership of their publications and workers can configure specific permission for their publications. There will be some actions and relationships within different objects. The academic learning network will be established through multiple objects and their interaction.

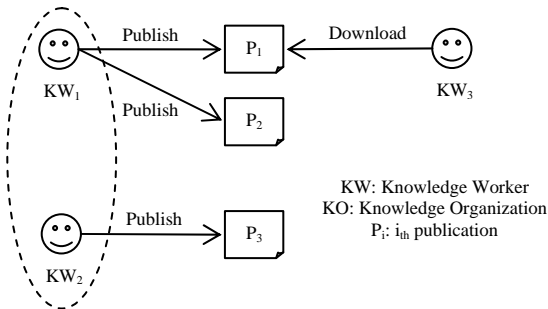


Figure 1 A sketch of academic learning network

#### B. Formal Descriptions

Here we introduce our ownership model. DISIT[1] defines a set of formal definition for Intellectual Property Right (IPR) model. We construct our model based upon their definition and further provide a light-weighted but sufficient model for academic learning network. It is a flexible model because all the parameters in ownership model can be configured according to the requirement. The sets for the ownership models are defined as follows:

##### i. Basic Attributes

- $WID := \{WID_1 \dots WID_n\}$  is the set of worker IDs, each knowledge worker has its unique ID as its identification.
- $OID := \{OID_1 \dots OID_n\}$  is the set of organization IDs, each organization has its unique ID as its identification.
- $PID := \{PID_1 \dots PID_n\}$  is the set of publications IDs, each publication has its unique ID as its identification.
- $Description := \{\text{keywords, abstract } \dots\}$  is the set of publication descriptions.
- $ContentKind := \{\text{journal, conference, workshop } \dots\}$  is the set of publication categories.
- $Right := \{\text{publish, viewAll, viewPartial, download, } \dots\}$  is the set of actions of workers/organizations.
- $UserKind := \{\text{author, lab member, } \dots\}$  is the set of user kinds.

- $Permission := \{ContentKind \times Right \times UserKind\}$  is the set of permission depend on  $ContentKind$  and  $UserKind$  to map corresponding action rights.
- $AccessHistory := \{\text{Who, When, What, Frequency}\}$  is the set of access history of each publication.
- $Relationship := \{\text{author, member, leader } \dots\}$  is set of relationships within workers, organizations and publications.

##### ii. Filter Functions

To represent the ownership in academic learning network precisely, we further propose filter functions to achieve permission purpose. Knowledge workers will configure the permission constraints of each publication according to  $ContentKind$  and  $UserKind$  of publication. Every non-author must access publication through corresponding filter. The representation of filter function as follows:

$$WID_i.permission(PID_j), i, j \in \text{int}$$

, which means the permission for  $worker_i$ 's  $publication_j$  and filter function will return corresponding  $Right$  to requesters.

### IV. SOCIAL GRAPH

The formal definitions described above helps us to model our knowledge workers, organizations publications and filter. Now we further analyze the relationships within these participants in academic learning network. We consider workers/organizations/publications/filter in academic learning network as individual objects (Nodes) and the relationships within these participants can be represented by the links connecting the nodes. With this visualization, we can quantify the features in graph, e.g. the number of node, the number of link to evaluate the relationship within participants in an academic learning network.

#### A. Object Representation

Each object is comprised of various attributes. We visualize these objects with social graph concept introduced in [1]. Table 1 lists all visual icons we use in academic learning networks, the detailed explanations are in the following sections.

Table 1 Visual icons for Social Graph

Name	Graph	Attributes
Knowledge Worker		{WID, OID, PID, Permission, Filters}
Knowledge Organization		{WID, OID, Leader}
Publication		{WID, PID, ContentKind, Description, Permission}
Action		{Right}
Filter		{WID, PID, Permission, ContentKind, UserKind}

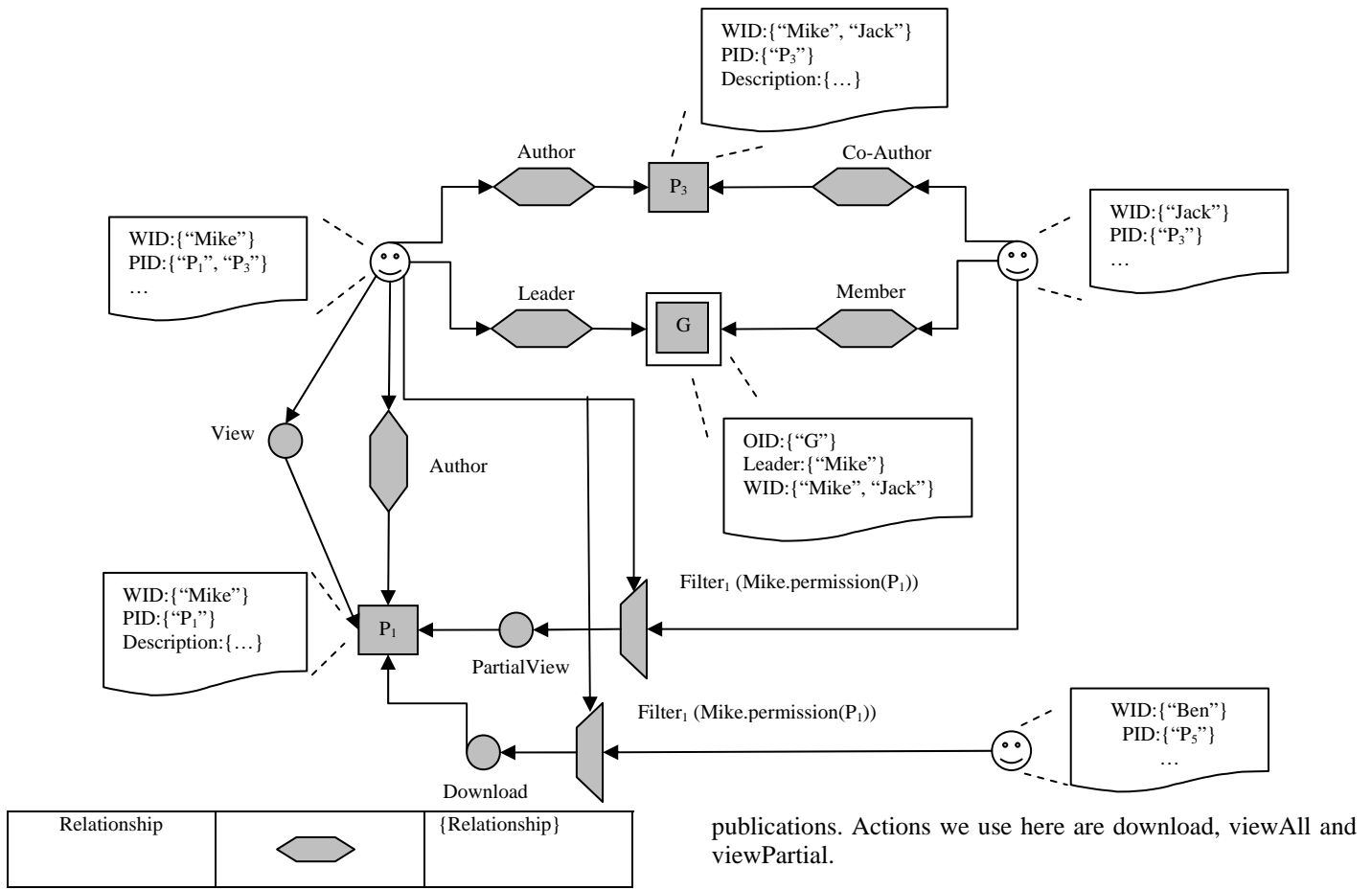


Figure 2 Academic Learning Network with Social Graph

publications. Actions we use here are download, viewAll and viewPartial.

#### v. Filter

A filter controls the access rights for a publication, the permission constraints are configured by author. In academic learning network, every publication should be access through its filter except its author.

#### vi. Relationship

There are at least three kinds of relationships in academic learning network:

1. *Workers/Organizations V.S. Workers/Organizations*  
Workers and organizations in networks have some relationships with others. E.g. Colleague, members in same organization.
2. *Workers V.S. Publications*  
Workers have relationship with publications. E.g. Author, Co-Author.
3. *Publication V.S. Publications*  
There exist some relationships within publications. E.g. Reference.

#### B. Academic Learning Network with Social Graph

With above social graph, now we explain how to combine our ownership model with social graph to establish an academic learning network. Fig. 2 depicts a simple academic learning

#### i. Knowledge Worker

A knowledge worker object includes its unique ID (WID) and their publication ID (PID) in academic learning network. Workers configure filters for their publications with specific permission to control the access right of a publication.

#### ii. Knowledge Organization

An organization comprises multiple members (e.g. Professors and Students in a University). Knowledge organization object includes its unique ID (OID) and knowledge workers (WID) in this organization. Each organization will be hosted by a leader. Leader will manage the relationships within members in the same organization.

#### iii. Publication

A publication object consists of its unique ID (PID), authors' ID (WID), type (ContentKind) and the description about this publication.

#### iv. Action

Actions in academic learning network are defined in Right in our ownership model, which means the actions to access



network with our social graph. We will go through this graph to explain our idea in detail.

#### i. Sub-Graph

Each knowledge worker can be represented by a sub-graph. In Fig. 2, *Mike* is a knowledge worker and he has a publication  $P_1$ . Hence, the object *Mike* has attributes:  $WID:\{Mike\}$  and  $PID:\{“P1”, “P3”\}$ . Each publication object also has their attributes, e.g.  $WID$ ,  $PID$  and  $Descriptions$ , the relationship between authors and their publication are represented with *Author* relationship graphs. According to defaulted rule in our ownership model, authors can directly access their publications without filter, so *Mike* can perform an action *View* to his publication  $P_1$ . *Mike* also sets a  $filter_1$  for  $P_1$ . Any non-author have to access  $P_1$  through  $filter_1$ , and the permission of  $filter_1$  depends on the *ContentKind* of  $P_1$  and the *UserKind* of accessors.

#### ii. The connection within sub-graphs

In addition to the representation of individual knowledge worker, there will be some connections within different knowledge workers (sub-graphs), the connections can be categorized into three types:

1. *Relationship link*:  
Workers cooperative with each other.
2. *Filter-action link*:  
One want to access the publication of others.
3. *Group link*:  
Workers in the same knowledge organization.

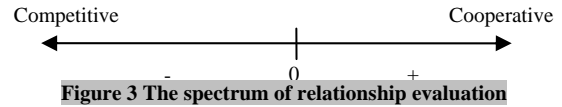
In Fig. 2, *Mike* and *Jack* are in the same knowledge organization  $G$ , and *Mike* is the leader in  $G$ , so there are two relationship graphs *Leader* and *Member* links  $\langle Leader, Member \rangle$  within *Mike*, *Jack* and  $G$ , this connection is called *Group link*. *Mike* has one cooperative publication  $P_3$  with *Jack*, so both *Mike* and *Jack* have *Author* and *Co-Author* relationships with  $P_3$ , this connection is called *Relationship link* and we use  $\langle Author, Co-Author \rangle$  to represent it. Now if *Jack* and *Ben* want to access  $P_1$ , since both of them are not author of  $P_1$ , the request will be checked by the filter  $Filter_1$  configured by *Mike*. The permission here lets non-author knowledge worker can partially view  $P_1$ . This connection is called *Filter-action link* and denoted with  $\langle Filter_1, PartialView \rangle$ .

This social graph is flexible and can be expanded with arbitrary connections, in next section, we will introduce how to utilize the features in social graph to evaluate the relationships in academic learning network.

### V. RELATIONSHIP EVALUATION

After establishing academic learning network with proposed social graph. Now we explain how to utilize our social graph to evaluate the relationship within academic learning networks. In this paper, we focus on the degree of cooperative and competitive relationship as our evaluation reference. Fig. 3

shows a spectrum to represent our evaluation reference. Here we consider positive weight represent the degree of cooperative relationship and negative weight the degree of competitive relationship.



#### A. Evaluation Metrics

The matrices we use have two categories, baseline weight and link weight. The evaluation process will accumulate the weights from these two parts.

##### i. Baseline Weight

The baseline weight represents a base relationship within different knowledge workers in the academic learning network. Here we use the similarity in their publications' description. Since we can obtain abstract and keyword information in their publications' description. This information can give us an initial evaluation for knowledge workers in academic learning network. We choose the number of the same words in their descriptions as a reference.

$$\begin{aligned} \text{Baseline weight } (BW_{\langle a,b \rangle}) \\ = \# \text{ of the same words in publications' descriptions of } \\ KW_a \text{ and } KW_b. \end{aligned}$$

##### ii. Link Weight ( $LW_{\langle a,b \rangle}$ )

As we mentioned in previous section, the connections within knowledge workers are established with relationship link, filter-action link and group link. Hence, we can assign a specific weight to each link. The value can be configured according to different situations. Here we just propose some examples in our evaluation.

##### ● Weights for relationship link ( $Weight_{RL_{\langle a,b \rangle}}$ )

The weight of relationship link depends on the type of relationship. In our ownership, the relationship within knowledge worker can be established through objects (cooperative publication) or knowledge organization (members, colleagues). We can assign different weights according to various relationship links. E.g. we can set the weight of  $\langle Author, CoAuthor \rangle$  relationship link will larger than the weight of  $\langle Author, Advisor \rangle$  relationship link.

$$\begin{aligned} &Weight_{RL_{\langle a,b \rangle}} \\ &= \sum_i^n \text{weight\_relationship}_i \text{ within } KW_a \text{ and } KW_b, \\ &i: \text{Relationship in Type}_i \end{aligned}$$

##### ● Weights for filter-action link ( $Weight_{FL_{\langle a,b \rangle}}$ )

The weight of filter-action link depends on the information in access history of publications. Since filter-action links will be established when non-author knowledge workers trying to access other knowledge workers' publications, we can obtain these information from publications' access history, e.g. number of download, view and accessed by who. Hence, we can use these information as a reference for evaluation.

$$Weight_{RL_{\langle a,b \rangle}} = \sum_i^n \# of Action_i * weight\_Action_i$$

- Weights for group link ( $Weight_{GL_{\langle a,b \rangle}}$ )

The weight of group link depends on the connection in a knowledge organization, since in a knowledge organization, there are various type of connections, e.g.  $\langle Leader, Member \rangle$  or  $\langle Colleague, Colleague \rangle$ , we can assign different weights to these connections.

$$Weight_{GL_{\langle a,b \rangle}} = \sum_i^n \begin{matrix} connection_i \text{ within } KW_a \text{ and } KW_b \\ \text{in organization} \end{matrix}$$

### iii. Total Evaluation

After defining baseline weight and link weight, we can evaluate the relationship within any knowledge worker.

$$Weight\_relationship(KW_a, KW_b) = BW_{\langle a,b \rangle} + LW_{\langle a,b \rangle}$$

Note that all weights in our evaluation approach can be configured, the type of relationship link and the connection in an organization can also be extended for different academic learning network.

### B. An Example

Now we demonstrate our evaluation approach with an example. Table 2 lists exemplary weights for different links. We use these weights to evaluate the relationship among knowledge workers as shown in Fig. 2.

Table 2 Metrics for Links

Category of Link	Kind	Weight
Relationship Link	$\langle Author, Co-Author \rangle$	+20
Filter-Action Link	$\langle ViewPartial \rangle$	-5
	$\langle Download \rangle$	-10
Group Link	$\langle Leader, Member \rangle$	+5

First, we evaluate the relationship between knowledge workers *Mike* and *Jack*. We assume their baseline weight ( $BW_{\langle Mike, Jack \rangle}$ ) is a constant  $C$ , and there are three links connecting them together, relationship link  $\langle Author, Co-Author \rangle$ , group link  $\langle Leader, Member \rangle$  and filter-action link  $\langle Filter_1, PartialView \rangle$ . According to the formulae given in 4.1, the numeric relationship between *Mike* and *Jack* is:

$$\begin{aligned} &Weight\_Relationship_{\langle Mike, Jack \rangle} : \\ &= BW + (Weight_{RL} + Weight_{FL} + Weight_{GL}) \\ &= C + (W_{\langle Author, Co-Author \rangle} + W_{\langle Filter_1, PartialView \rangle} + W_{\langle Leader, Member \rangle}) \\ &= C + (20 + (-5) + 5) \end{aligned}$$

$$= C + 20$$

And then we evaluate the relationship between *Mike* and *Ben*. We also assume their baseline weight ( $BW_{\langle Mike, Ben \rangle}$ ) is a constant  $C$  and there is a filter-action link  $\langle Filter_1, Download \rangle$ . According to the formula, the numeric relationship between *Mike* and *Ben* is:

$$\begin{aligned} &Weight\_Relationship_{\langle Mike, Ben \rangle} : \\ &= BW + (Weight_{FL}) \\ &= C + (W_{\langle Filter_1, Download \rangle}) \\ &= C + 10 \end{aligned}$$

### C. Analysis

From above example, we can observe the relationship evaluation depends on the type of links and their user-defined parameters. In general case, we can consider the group link (static) and filter-action link (dynamic) will occupy the major part in relationship evaluation. Since the knowledge workers in the same knowledge organization, the probability of their cooperation will be increased. For an example, the relationship link within a group will be more positive (e.g.  $\langle Leader, Member \rangle$ ) than the relation link within different knowledge organization. Furthermore, if there are number of filter-action link between two knowledge workers, we should evaluate their relationship according to the dynamic actions.

## VI. CASE STUDY

We apply our formal model into a realistic academic learning network – SMNET [4], developed by Dr. Paolo Nesi and DISIT Lab in Florence, Italy, as the test bed. In the following, we first explain the scenario of SMNET with a realistic case as an example 5.

### A. The SMNET Scenario

Here we describe how to map our formal description to SMNET platform. Table 3 lists the mapping between SMNET metadata configuration and proposed formal description, once users provide their information, the corresponding social graph will be generated automatically. SMNET has versatile configuration to generate intent social network. Here we only list part of attributes that describing our academic learning network.

Table 3 The mapping between SMNET and Formal Description

SMNET	Formal Description
Title	PID
Creators	{WID <sub>1</sub> , WID <sub>2</sub> , ...}
Descriptions	{ Keywords, Abstract ...}
Taxonomy Classification ->ContentKind	ContentKind
Group Section	Permission

Fig. 4 shows the scenario of using our ownership model with SMNET platform, when knowledge workers want to upload their publications, the information should be provided according to the format of Table 3. After filling publications' information, SMNET generates corresponding social graph, then the relationships within academic learning can be calculated. Fig. 5 shows the screen of the initial SMNET.

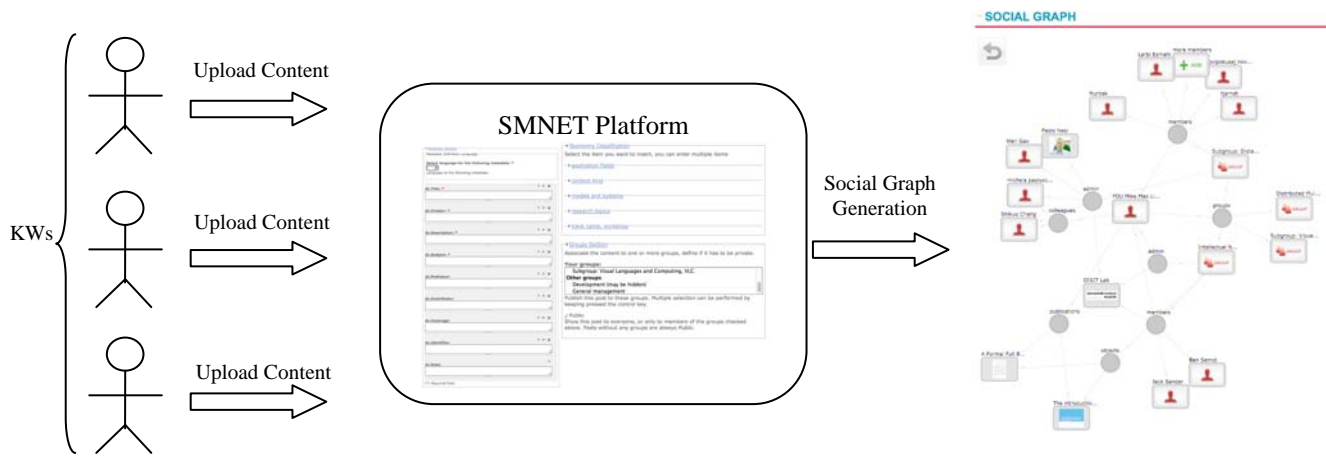


Figure 4 Scenario of SMNET Platform

Note that the filters and dynamic actions (View, Download,...) are invisible in SMNET social graph, so we use dotted line to represent it and all dynamic actions should be recorded in out access history. In Fig. 5, Mike, Ben and Jack (Knowledge Worker) are the members in *Intellectual Network* group (Knowledge Organization). Mike is the admin (Leader) and the creator (Author) of *Publication A* in Intellectual Network. Jack and Ben are cited names of *Publication A* (Co-authors). As an author of *Publication A*, Mike set the permission of filter that

only the member in *Intellectual Network* group can view *Publication A*. So John (Non-Intellectual network group member) cannot view publication A but Ben and Jack (Intellectual network group member) can view it. All dynamic actions will be recorded in access history of each publication for evaluating relationships. After obtaining social graph, the weights can then be computed. Currently the weights are computed off-line. In the next version, the weights computation will be included in SMNET.

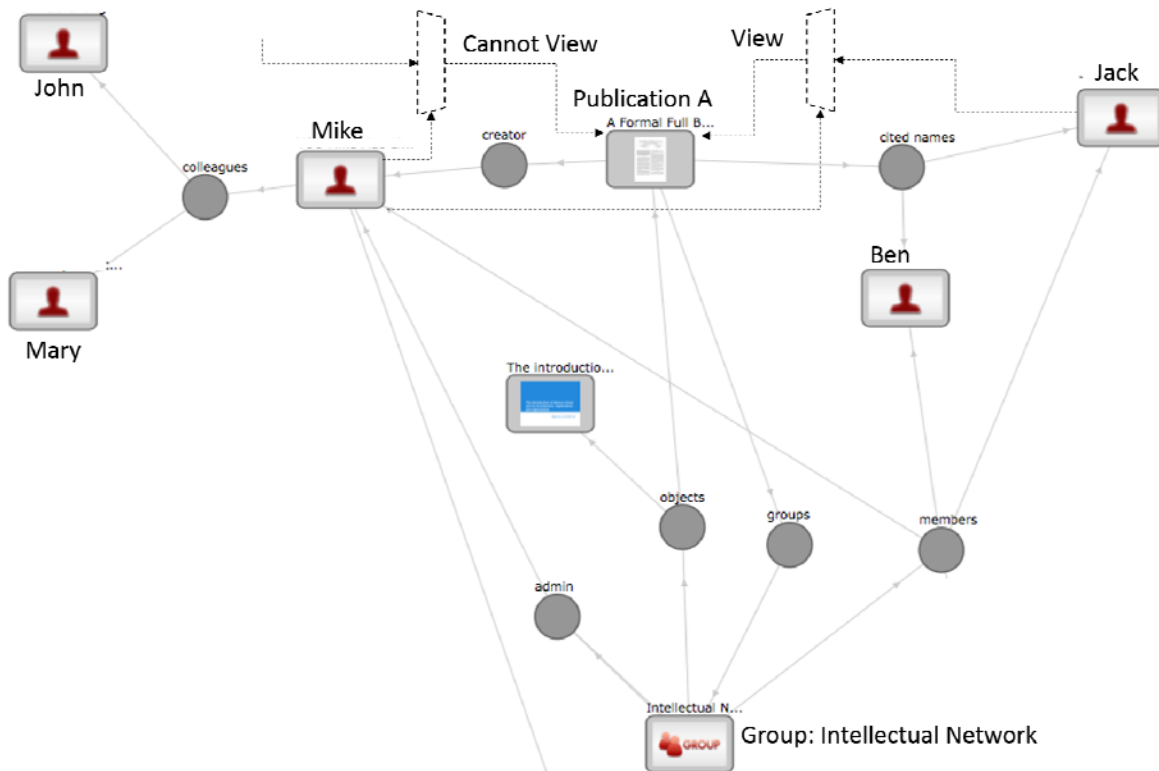


Figure 5 Social Graph in SMNET

## VII. CONCLUSION & FUTURE WORK

Academic learning network can be seen as a subset of social network, but there are few specific approaches to clearly define it and analysis the relationships in it with low-cost algorithm.

In this paper, we propose a formal ownership model that capable to describe academic learning network and can be easily integrated with existed social network platform, the intelligent relationships within Knowledge Worker/Organization can be represented by social graph.

With quantified metrics, the degree of relationship can be easily evaluated and further help us to analyze academic learning network. In addition to static factors in social network, we further consider various dynamic actions as our evaluation reference.

More sophisticate relationship and evaluation metrics will be extended to measure more complex social network in the future.

## ACKNOWLEDGMENT

Thanks for Dr. Paolo Nesi in University of Florence and his student Michela Paolucci for providing SMNET as our experimental platform and helpful advice to complete this work.

## REFERENCES

- [1] Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi, Michela Paolucci, "Institutional Services and Tools for Content, Metadata and IPR Management". DMS, 2013.
- [2] Shi-Kuo Chang, "A General Framework for Slow Intelligence Systems", International Journal of Software Engineering and Knowledge Engineering, Volume 20, Number 1, February 2010, 1-16.
- [3] Shi-Kuo Chang, Yingze Wang and Yao Sun, "Visual Specification of Component-based Slow Intelligence Systems", Proceedings of 2011 International Conference on Software Engineering and Knowledge Engineering, Miami, USA, July 7-9, 2011, 1-8.
- [4] Sentient Multimedia Network, <http://smnet.disit.org>
- [5] European Library of Artistic Performance, ECLAP, <http://www.eclap.eu/>
- [6] Wiki book, "Social Network Analysis: Theory and Applications", 2011.
- [7] Hyunjung Kim and George A. Barnett, "Social Network Analysis Using Author Co-Citation Data", Proceedings of Americas Conference on Information Systems (AMCIS), 2008.
- [8] Tsai-Yuan Lin and Yun-Yao Cheng. "Exploring the Knowledge Network of Strategic Alliance Research: Co-Citation Analysis", International Journal of Electronic Business Management, Vol. 8, No. 2, pp. 152~160, 2010.
- [9] Borgatti, S.P., Everett, M.G. and Freeman, L.C., "Ucinet for Windows: Software for Social Network Analysis", Harvard, MA: Analytic Technologie, 2002.

# Ontology Bulding vs Data Harvesting and Cleaning for Smart-city Services

Pierfrancesco Bellini, Monica Benigni, Riccardo Billero, Paolo Nesi, Nadia Rauch

DISIT Lab, Dep. of Information Engineering, University of Florence, Italy

<http://www.disit.dinfo.unifi.it> , {pierfrancesco.bellini, riccardo.billero, paolo.nesi, nadia.rauch}@unifi.it

**Abstract**— Presently, a very large number of public and private data sets are available around the local governments. In most cases, they are not semantically interoperable and a huge human effort is needed to create integrated ontologies and knowledge base for smart city. Smart City ontology is not yet standardized, and a lot of research work is needed to identify models that can easily support the data reconciliation, the management of the complexity and reasoning. In this paper, a system for data ingestion and reconciliation of smart cities related aspects as road graph, services available on the roads, traffic sensors etc., is proposed. The system allows managing a big volume of data coming from a variety of sources considering both static and dynamic data. These data are mapped to smart-city ontology and stored into an RDF-Store where they are available for applications via SPARQL queries to provide new services to the users. The paper presents the process adopted to produce the ontology and the knowledge base and the mechanisms adopted for the verification, reconciliation and validation. Some examples about the possible usage of the coherent knowledge base produced are also offered and are accessible from the RDF-Store and related services. The article also presented the work performed about reconciliation algorithms and their comparative assessment and selection.

**Keywords**— *Smart city, knowledge base construction, reconciliation, validation and verification of knowledge base, smart city ontology, linked open graph.*

## I. INTRODUCTION

Despite the large work performed by Public Administrations (PAs) on producing open data they are not typically semantically interoperable and neither with the many private data. Open data coming from PA contains typically statistic information about the city (such as data on the population, accidents, flooding, votes, administrations, energy consumption, presences on museums, etc.), location of point of interests on the territory (including, museums, tourism attractions, restaurants, shops, hotels, etc.), major GOV services, ambient data, weather status and forecast, changes in traffic rules for maintenance interventions, etc. Moreover, a relevant role is covered in city by private data coming from mobility and transport such as those created by Intelligent Transportation Systems, ITS, for bus management, and solutions for managing and controlling parking areas, car and bike sharing, car flow, good delivering services, accesses on Restricted Traffic Zone (RTZ) etc. Both open and private data may include real time data such as the traffic flow measure, position of vehicles (buses, car/bike sharing, taxi, garbage collectors, delivering services, etc.), railway and train status, park areas status, and Bluetooth tracking systems for

monitoring movements of cellular phones, ambient sensors, and TV cameras streams for security. Both PAs and mobility operators have large difficulties in elaborating and aggregating these data to provide new services, even if they could have a strong relevance in improving the citizens' quality of life. Therefore, our cities are not so smart as they could be by exploiting a semantically interoperable knowledge base exploiting on these data. This condition is also present in highly active cities on open data publication such as Firenze, that is considered one of the top cities on Open Data.

Therefore, the variability, complexity, variety, and size of these data, make the data process of ingestion and exploitation a "Big Data" problem as addressed in [2], [3]. The variety and variability of data can be due to the presence of several different formats, and to scarce (or non-existing) interoperability among semantics of the single fields and of the several data sets. In order to reduce the ingestion and integration cost, by optimizing services and exploiting integrated information at the needed quality level, a better interoperability and integration among systems is required [1], [2]. This problem can be partially solved by using specific reconciliation processes to make these data interoperable with other ingested and harvested data. The velocity of data is related to the frequency of data update, and it allows to distinguish static data from dynamic data: the first one are rarely updated, such as once per month/year, as opposed to the second one that are updated from once a day up to every minute or more. When these data models are analyzed and then processed to become semantically interoperable, they can be used to create a common knowledge base that can be feed by corresponding data instances (with static, quasi-static and real time data). This process may lead to create a large interoperable knowledge base that can be used to make queries for producing suggestions as well as, predictions, deductions, in the navigation or in the service access and usage.

This scenario enables the creation of new services exploiting the accumulated knowledge for: delivering service predictions and tuning, deducing and predicting critical conditions, towards different actors: public administrations, mobility operators, commercials and point of interests and citizens. In this paper, the above mentioned complex process of knowledge base construction is described from: ontology creation to the data ingestion and knowledge base production and validation. The mentioned process also include processes of data analysis for ontology modeling, data mining, formal verification of inconsistencies and incompleteness to perform data reconciliation and integration. Among the several processes, the most critical aspects are related to the ontology

construction that can enable deduction and reasoning, and on the verification and validation of the obtained model and knowledge base.

The paper is organized as follows. In Section II, the overview of the proposed ontology is presented together with the main problems underlined its construction, and the main macro classes. Section III describes the details associated to each macroclass of the proposed smart city ontology and the integration with other vocabulary. Section IV reports the general architecture adopted for processing Open Data and the motivations that constrained its definition. In the same section, two services are presented that allow navigating in the knowledge base and can be used by non-data engineers. Section V presents the verification and validation process adopted for the knowledge base, and the results regarding the reconciliation precision and recall by using different kind of algorithms. Conclusions are drawn in Section VI.

## II. ONTOLOGY MAIN ELEMENTS

In order to create an ontology for Smart City services, a large number of data sets have been analyzed to see in detail each single data elements of each single data set with the aim of modeling and establishing the needed relationships among element, thus making a general data set semantically interoperable (e.g., associating the street names with toponymous coding, resolving ambiguities). The work performed started from the data sets available in the Florence and Tuscany area. In total the whole data sets are more than 800 data sets. At regional level, Tuscany Region provided a set of open data into the MIIC (Mobility Integration Information Center of the Tuscany Region), and provide integrated and detailed geographic information reporting each single street in Tuscany (about 137,745), and the locations of a large part of civic numbers, for a total of 1,432,223 (a wider integration could be performed integrating also Google maps and Yellow/white pages). From the MIIC, it is possible to recover information regarding streets, car parks, traffic flow, bus timeline, etc. While from Florence municipality, real time data such as those from the RTZ about car passages, tram lines on the maps, bus stops, bus tickets, statistics on accidents, ordinances and resolutions, numbers of arrivals in the city, number of vehicles per year, etc., can be obtained. From the other open data, points of interest (POI) can be recovered as position and information related to: museums, monuments, theaters, libraries, banks, express couriers, police, firefighters, restaurants, pubs, bars, pharmacies, airports, schools, universities, sports facilities, hospitals, emergency rooms, government offices, hotels and many other categories, including weather forecast by LAMMA consortium. In addition to these data sets, those coming from the mobility and transport operators have been collected as well.

The analysis of the above mentioned data sets allowed us to create an integrated ontological model presenting 7 main areas of macroclasses as depicted in Figure 1.

**Administration:** includes classes related to the structuring of the general public administrations, namely PA, and its specifications, Municipality, Province and Region; also includes the class Resolution, which represents the ordinance

resolutions issued by each administration that may change the traffic stream.

**Street-guide:** formed by entities as Road, Node, RoadElement, AdministrativeRoad, Milestone, StreetNumber, RoadLink, Junction, Entry, and EntryRule Maneuver, it is used to represent the entire road system of Tuscany, including the permitted maneuvers and the rules of access to the RTZ. The street model is very complex since it may model from single streets to areas, different kinds of crosses and superhighways, etc. In this case, OTN (Ontology for Transport Network) vocabulary has been exploited to model traffic [4] that is more or less a direct encoding of GDF (Geographic Data Files) in OWL.

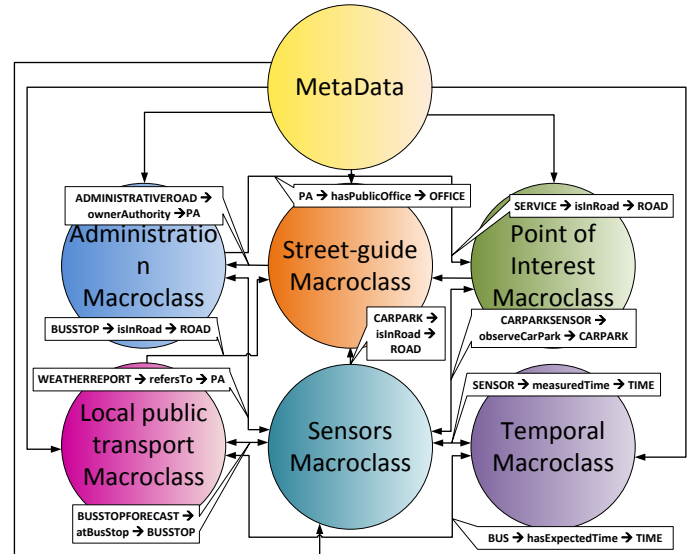


Figure 1 - Ontology Macro-Classes and their connections

**Point of Interest (POI):** includes all services, activities, which may be useful to the citizen and who may have the need to “search-for” and to “arrive-at”. The classification of individual services and activities is based on main and secondary categories planned at regional level. In addition, this macro segment of the ontology may take advantage of reusing Good Relation model of the commercial offers<sup>1</sup>.

**Local public transport:** includes the data related to major LPT (Local Public Transport, in italian: TPL, Transport Public Local) companies scheduled times, the rail graph, and data relating to real time passage at bus stops. Therefore, this macroclass is formed by classes PublicTransportLine, Ride, Route, AVMRecord, RouteSection, BusStopForecast, Lot, BusStop, RouteLink, RouteJunction. (where AVM means Automatic Vehicle Monitoring).

**Sensors:** macroclass concerns data from sensors: ambient, weather, traffic flow, pollution, etc. Currently, data collected by various sensors installed along some streets of Florence and surrounding areas, and those relating to free places in the main car parks of the region, have been integrated in the ontology. Some of the sensors can be located on moving vehicles such as

<sup>1</sup> <http://www.heppnetz.de/projects/goodrelations/>





class *RoadElement* through two inverse ObjectProperties *hasRoadElement* and *formAdminRoad*, while it is connected with only one ObjectProperty, *coincideWith*, to the class *Road*. In order to better clarify the relationship that exists between classes *Road*, *AdministrativeRoad* and *RoadElement*: a *Road*'s instance can be connected to multiple instances of class *AdministrativeRoad* (e.g., if a road crosses the border between two provinces), but the opposite is also true (e.g., when a road crosses a provincial town center and it assumes different names), i.e., there is a N:M relationship between these two classes. On each road element, it is possible to define access restrictions, identified by class *EntryRule*, which is connected to class *RoadElement* through 2 inverse ObjectProperties, i.e., *hasRule* and *accessToElement*. The class *Maneuver* and class *EntryRule* are connected by ObjectProperty *hasManeuver*. Moreover, we verified that only in rare cases maneuvers involving three different road elements, to represent the relationship between classes *Maneuver* and *RoadElement*, three ObjectProperties were defined: *hasFirstElem*, *hasSecondElem* and *hasThirdElem*. In addition to the ObjectProperty that binds a maneuver to the junction that is interested, that is, *concerningNode* (because a maneuver takes place always in proximity of a node). Each instance of *Milestone* class must be associated with a single instance of *AdministrativeRoad*, and it is therefore defined a cardinality restriction equal to 1. Associated with ObjectProperty *isInElement*, also class *Milestone* is defined as subclass of *geo:SpatialThing*, in this case the presence of coordinates is not mandatory, to be capable to model entities that does not present those data. Thanks to the owned data, classes *StreetNumber* and *Entry* were defined: the connection of class *StreetNumber* to class *Road*, is possible respectively through the ObjectProperties *hasStreetNumber* and *belongsToRoad*. The relationship between classes *Entry* and *StreetNumber*, is also defined by the two ObjectProperties, *hasInternalAccess* and *hasExternalAccess*. The class *Entry* is defined as a subclass of *geo:SpatialThing*, and it is possible to associate a maximum of one pair of coordinates *geo:lat* and *geo:long* with each instance. The Street-guide macroclass is connected to the Administration macroclass through two different ObjectProperties -- i.e., *OwnerAuthority* and *managingAuthority*, which represent respectively the public administration which owns an *AdministrativeRoad*, or public administration that manages a *RoadElement*. Thanks to the processing of KMZ files (Keyhole Markup Language file and zero or more supporting files packaged in a ZIP file), is possible to retrieve the set of coordinates that define the broken line of each *RoadElement*. Each of these points is added to the ontology as an instance of class *Junction* (defined as a subclass of *geo:SpatialThing*, with compulsory single pair of coordinates). Each small segment between two instances of *Junction* class is instead an instance of class *RoadLink*, which is defined by a restriction on the ObjectProperties *ending* and *starting*, which connect the two mentioned classes. *RoadLink* and *Junctions* are in total about 20 million of triples.

### C. Point of Interest Macroclass

This macroclass allows to represent services to the citizens, points of interest, businesses activities, tourist attractions, and anything else can be located thanks to a pair of coordinates on a map. Each type of element has been defined starting from the

categories defined by the Tuscany Region taxonomy of categories, including: Accommodation, GovernmentOffice, TourismService, TransferService, CulturalActivity, FinancialService, Shopping, Healthcare, Education, Entertainment, Emergency and WineAndFood.

It is easy to understand that the main class of the Point of Interest Macroclass is a generic class *Service* for which the subclasses above listed have been identified thanks to the value assigned to ObjectProperty *serviceCategory*.

The class *Accommodation* for example, was defined as a restriction of the class *Service* on ObjectProperty *serviceCategory*, which must take one of the following values: *tourist\_resort*, *hotel*, *tourist\_home*, *rest\_home*, *religiuous\_guest\_house*, *bed\_and\_breakfast*, *hostel*, *summer\_residence*, *vacation\_resort*, *farmhouse*, *day\_care\_center*, *camping*, *historic\_residence*, *mountain\_dew*.

We have also defined DataProperty *ATECOcode*, i.e. ATECO is the ISTAT (national institute for statistics in Italy, [www.istat.it](http://www.istat.it)) code for the classification of economic activities, which could be used in future as a filter to define the various services subclasses, in place of the categories proposed by the Tuscany Region, in order to make more precise research of the various types of services. Thanks to the class *Service* the macroclasses *Point of Interest* and *Street guides* can be connected by exploiting ObjectProperty *hasAccess*, with which a service can be connected to only one external access, corresponding to the road and the street number of the service location. If this association is not possible (because of lack of information, missing street number, etc..), the connection between the same two macroclasses listed above, is realized through the ObjectProperty *isInRoad*, that connects an instance of the class *Service* to an instance of the class *Road*. In order to use at least one of these two ObjectProperty to connect macroclasses *Point of Interest* and *Street Guides*, an intense reconciliation phase is necessary, as described in section IV.

### D. Public Transport Macroclass

The TPL (Italian LPT) macroclass (see Figure 3) includes information relating to public transport by road and rail. The public transport by road is organized in public transport lots, each of which is in turn composed of a number of bus and tram lines. Each line includes at least two ride per day (the first in ascendant direction, and the second one in descendant direction), identified through a code provided by the TPL company and each ride is scheduled to drive along a specific path, called route. A route can be seen as a series of road segments delimited by subsequent bus stops, but wishing then to represent to a cartographic point of view the path of a bus, we need to represent the broken line that composes each stretch of road crossed by the means of transport itself, and to do so, the previously used modeling on road elements, has been reused: we can see each path as a set of small segments, each of which delimited by two junctions.

The part relating to rail transport: each railway line, i.e., an infrastructure designed to run trains between two places of service, is composed by a number of railway elements, which can also form a railway direction (a railway line having particular characteristics of importance for volume of traffic

and transport relations linking centers or main nodes of the rail network) and a railway section (section of the line in which you can find only one train at time, and that is usually preceded by a "protective" or "block" signal). In addition, each rail element begins and ends at a railway junction, in correspondence of which there may be train stations or cargo terminals.

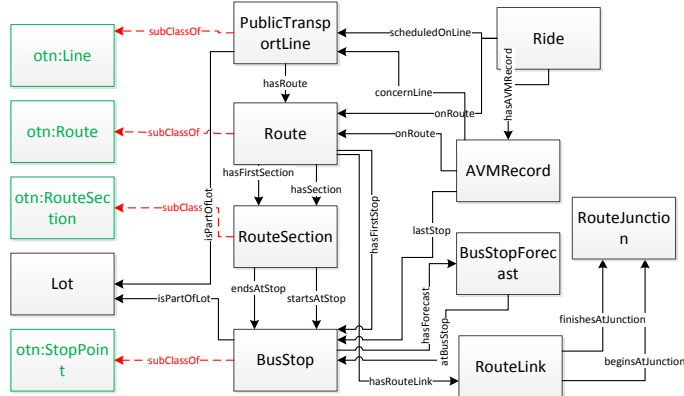


Figure 3 - Public Transport Macroclass (a portion)

Based on the previous description, we have defined class *PublicTransportLine* (that it is also subclass of *OTN:Line*), which is connected to the corresponding instance of class *Lot*, thanks to ObjectProperty *isPartOfLot*. Every instance of class *PublicTransportLine* is connected to class *Ride* through ObjectProperty *scheduledOnLine*, which is also defined as a limitation of cardinality exactly equal to 1, because each stroke may be associated to a single line. To model each path and its sequence of crossed bus stops, classes *Route* and *BusStop* have been defined. We decided to define two ObjectProperties linking classes *Route* and *RouteSection*, i.e. *hasFirstSection* and *hasSection*, since, from a cartographic point of view, wanting to represent the path that a certain bus follows. In details, knowing the first segment and the stop of departure, it is possible to obtain all the other segments that make up the complete path and, starting from the second bus stop (that is identified as the different stop from the first stop, but that it is also contained in the first segment), we are able to reconstruct the exact sequence of the bus stops, and then the segments, which constitute the entire path. For this purpose also ObjectProperty *hasFirstStop* has been defined, which connects classes *Route* and *BusStop* and ObjectProperty *endsAtStop* and *startsAtStop*, which connect instead each instance of *RouteSection* to two instances of class *BusStop* (subclass of *OTN:StopPoint*). Each stop is also connected to class *Lot*, through the ObjectProperty *isPartOfLot*, with a 1:N relation, because there are stops shared by urban and suburban lines so they belong to two different lots. Possessing also the coordinates of each stop, class *BusStop* was defined as a subclass of *geo:SpatialThing*, and was also termed a cardinality equal to 1 for the two DataProperty *geo:lat* and *geo:long*. In order to represent the broken line that composes each route, classes *RouteLink* and *RouteJunction*, and the ObjectProperties *beginsAtJunction* and *finishesAtJunction*, were defined. The

class *Route* is also connected to class *RouteLink* through *hasRouteLink* ObjectProperty.

The Railway Graph is mainly formed by class *RailwayElement*, that can be connected to classes *RailwayDirection* and *RailwaySection*, thanks to two inverse ObjectProperties *isComposedBy* and *composeSection*, and to class *RailwayLine*, through the two inverse ObjectProperties *isPartOfLine* and *hasElement*. Each instance of class *RailwayElement* is connected to two instances of class *RailwayJunction* (defined as a subclass of the OTN:Node), by the ObjectProperties *startAtJunction* and *endAtJunction*. Classes *TrainStation* and *GoodsYard* correspond only to one instance of the *RailwayJunction* class, both through the ObjectProperty *correspondToJunction*.

#### E. Sensors Macroclass

Sensors Macroclass consists of four parts related to car parks sensors, weather sensors, traffic sensors installed along roads/rails and to AVM/kit systems installed on buses, cars and/or bikes. The first part is focused on the real-time data related to parking: for each sensors installed into different car parking areas, a status record is received every 5 minutes. In each status report, there are information about the number of free and occupied parking spaces, for the main car parks in Tuscany Region. The weather sensors produce real-time data concerns the weather forecast, thanks to LAMMA (institute for modeling and monitoring environmental conditions in Tuscany, <http://www.lamma.rete.toscana.it>). This consortium updates the municipality forecast report once or twice per day and every report contains forecast for five days divided into range, which have a greater precision (and a higher number) for the nearest days until you get to a single daily forecast for the 4th and 5th day. The traffic sensors produce real-time data concerning the sensors placed along the roads of the region, which allow making different measures and assessment related to traffic situation. Unfortunately, the location of these sensors is not very precise, it is not possible to place them in a unique point thanks to coordinate, but only to place them within a toponym, which for long-distance roads such as FI-PI-LI road (the highway that connect Florence-Pisa-Livorno), it represents a range of many miles. Each sensor, is part of a group and produces observations which can belong to four types, i.e. they can be related to the average velocity, car flow passing in front of the sensor, traffic concentration, or to the traffic density. On this regards, Bluetooth sensors could be installed to trace the number of people passing by on car and bikes from a given point.

The AVM (Automatic Vehicle Monitoring) systems part concerns the sensors systems installed on most of buses, which, at intervals of few minutes, send a report to the management center. They provide information about: the last stop performed, current GPS coordinates of the vehicle, the identifiers of vehicle and of the line, a list of upcoming stops with the planned passage time.

To model the car parks situation we have defined the class *CarParkSensor* which is linked to instances of the class *SituationRecord*, that represent, as previously stated, the state of a certain parking at a certain instant; this link is performed via the reverse ObjectProperties, *relatedToSensor* and *hasRecord*. This first part of the Sensors Macroclass is also

connected to the Point of Interest Macroclass through two inverse ObjectProperties, *observeCarPark* and *hasCarParkSensor*, which connect the classes *CarParkSensor* and *TransferService*.

The weather situation, instead, is represented by class *WeatherReport* connected to class *WeatherPrediction* via the ObjectProperty *hasPrediction*. Moreover, class *Municipality* is connected to each report by two reverse ObjectProperties: *refersToMunicipality* and *hasWeatherReport*, to realize the connection between the macroclasses *Sensors* and *Administration*.

With regard to traffic sensors, each group of sensors is represented by class *SensorSiteTable* and each instance of class *SensorSite* connects to its group through the ObjectProperty *formsTable* and thanks to ObjectProperty *placeOnRoad* each instance of class *SensorSite* can be connected only to class *Road* (see Figure 2), to create a connection between *Sensors* and *Street-guide* macroclasses. Each sensor produces observations represented by instance of class *Observation* and, as mentioned earlier, there are four possible subclasses: *TrafficConcentration*, *TrafficHeadway*, *TrafficSpeed*, and *TrafficFlow* subclass. Classes *Observation* and *Sensor* are connected via a pair of reverse ObjectProperties, *hasObservation* and *measuredBySensor*.

Finally, the last part of *Sensors* Macroclass is mainly represented by two classes, *AVMRecord* and *BusStopForecast*, and thanks to the ObjectProperty *lastStop*, this first class can be connected to the *BusStop* class. The list of scheduled stops is instead represented as instances of the class *BusStopForecast*, a class that is linked to the class *BusStop* through *atBusStop* ObjectProperty so as to be able to recover the list of possible lines provided on a certain stop (the class *AVMRecord* is in fact also connected to the class *Line* via ObjectProperty *concernLine*).

#### F. Temporal Macroclass

The Temporal Macroclass, is now only "sketchy" within the

ontology, and it is based on the Time ontology [5] as it has been used into OSIM ontology [8]. It requires the integration of the concept of time as it will be of paramount importance to be able to calculate differences between time instants, and the Time ontology comes to help us in this task. We define fictitious URI: *#instantForecast*, *#instantAVM*, *#instantParking*, *#instantWreport*, *#instantObserv* to associate at a resource URI a time parameter -- i.e. respectively *BusStopForecast*, *AVMRecord*, *SituationRecord*, *WeatherReport* and finally *Observation*. It is necessary to create a fictitious URI that links a time instant to each resource, to avoid ambiguity, because identical time instants associated with different resources may be present (although the format in which a time instant is expressed has a fine scale). Time Ontology is used to define precise moments as temporal information, and to use them as extreme for intervals and durations definition, a feature very useful to increase expressiveness.

Pairs of ObjectProperties have also been defined for each class that needs to be connected to the class *Instant*: between classes *Instant* and *SituationRecord* were defined the inverse ObjectProperties *instantParking* and *observationTime*, between classes *WeatherReport* and *Instant*, the ObjectProperties *instantWreport* and *updateTime* have been defined; between classes *Observation* and *Time* there are the reverse ObjectProperties *measuredTime* and *instantObserv*, between *BusStopForecast* and *Time* we can find *hasExpectedTime* and *instantForecast* ObjectProperties, and finally, between *AVMRecord* and *Time*, there are the reverse ObjectProperties *hasLastStopTime* and *instantAVM*.

#### G. Metadata Macroclass

Finally, Metadata macroclass is used to keep track of the status and descriptors associated with the various ingested dataset. Sesame [www.openrdf.org] allows to assign a name (i.e., an identifier) to the various graphs that can be identified within the defined ontology, so defining some Named Graphs. This name, also called "context", allows to expand the triple

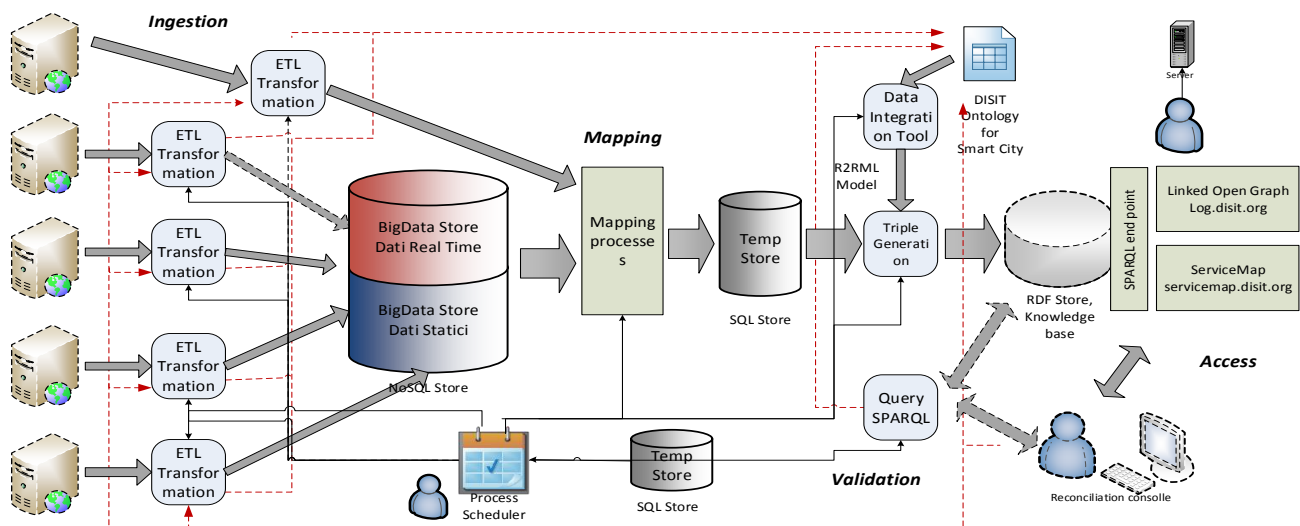


Figure 4 - Architecture Overview



data model to a quad data model, defined as follow: subject-predicate-object-context. Owlrim, allows to assign the context to each triple set, during the data loading phase. Therefore, a description and status context called *dataProperty* is associated with each data set. It allows to store all the useful information related to a certain data set, such as: date of creation, data source, original file format, dataset description, type of license bound to the dataset, kind of ingestion process, and how much automated is the entire ingestion process, type of access to the dataset, overtime, period, associated parameters, date of last update, date of triples creation, status of the ingestion process, etc..

#### IV. DATA ENGINEERING ARCHITECTURE

In this section, the description of the data engineering architecture is proposed (see Figure 4). The whole ingestion and quality improvement process can be regarded as divided into the following phases of: Data Ingestions, knowledge Mapping, knowledge Reconciliation to make the model semantic interoperable, Verification and Validation and Access/exploitation from services. The whole phases of the ingestion processes are managed by a Process Scheduler that allocates processes on a parallel and distributed architecture composed by several servers. To allow the regular update of ingested data the scheduler regularly retrieves data and check for updates. The ingested data are transcoded and then mapped in the Smart City Ontology. After that, they are made available to applications on an RDF Store (OWLIM-SE) using a SPARQL Endpoint. Applications can use the geo-referenced data to provide advanced services to the city citizens, such as the present solution for knowledge base browsing via Linked Open Data (<http://log.disit.org>) and the Service Map (<http://servicemap.disit.org>), described in the following section.

##### A. Data Ingestion

For the data ingestion, the problems are related to the management of the several formats and of the various data sets that may find allocation on different segments and areas of the Smart City Ontology. The solution allows ingesting and harvesting a wide range of public and private data, coming as static, semi-static and real time data as mentioned in the previous sections. For the case of Florence area, we are addressing about 150 different sources of the 564 available. **Static and semi-static data** include points of interests, geo-referenced services, maps, accidents statistic, etc. This information is typically accessible as public files in several formats, such as: SHP, KMZ, CVS, ZIP, XML, etc.

Each Open Data ingestion process retrieves information and produce records in a noSQL Hbase for big data [9], logging all the information acquired to trace back and versioning the data ingestion. Data are then completed; other columns are updated dynamically with other process steps, and finally data obtained are placed on an HBase table.

**Real time data** includes data coming from sensors (e.g., parking, weather conditions, pollution measures, busses, etc.) that are typically acquired from Web Services as well as more static data as road graph description, etc. For example ingestion of data relating to traffic sensors consists of a ETL

transformation (Extract, Transform, and Load). In most cases, the real-time data are directly pushed in the mapping process to feed the temporary SQL store. They are typically streamed into the traditional SQL store and then converted into triples in the RDF final store.

In almost all cases, each single data set is ingested by means of a different ETL process defined by using Pentaho Kettle formalism [10] because, among the several existing solutions, this formalism is quite diffused and easier to understand, and it was already used by Information Systems Directorate of Florence. When the Kettle language presented limitation, external processes in Java have been adopted.

##### B. Data Mapping

The Mapping Phase deals with the transport of information, previously saved into HBase database, into an RDF datastore, in our case managed by Owlrim-SE [11]. The first part of this procedure retrieves information from HBase to put them on a temporary MySQL database (required to use the Data Integration tool chosen), while in the second part data are translated into triples. Transformation is needed to map the traditional structured into RDF triples, based on information contained in a well-defined ontology (DISIT Ontology for Smart City) and all ontologies reused (dcterms, foaf, vCard, etc.). This process may be performed by ad-hoc programs that have to take into account the mapping from linear model to RDF structures. This two steps process allowed us to test and validate several different solutions for mapping traditional information into RDF triples and ontology. The ontological model has been several times updated and thus the full RDF storage has been regenerated from scratch reloading the definition (all the other vocabularies, selecting the testing several different solutions) and the instance triples according to the new model under test. Once the model has been generated, triples can be automatically inserted.

The first essential step is to specify semantic types of the data set, i.e., it is necessary to establish the relationship between the columns of the SQL tables and properties of ontology classes. The second step consists in defining the Object Properties among the classes, or the relationships between the classes of the ontology. When dataset has 2 columns that have the same semantic type but which correspond to different entities, thus multiple instances of the same class have to be defined, associate each column to the correct one.

The process responsible to perform the mapping transformation, passing from Hbase to SQL database has been produced as a corresponding ETL Kettle associated with each specific ingestion procedure for each data set. The second phase, of performing the mapping from SQL to RDF, has been realized by using a mapping model: Karma Data Integration tool [12], which generates a R2RML model, representing the mapping for transport from MySQL to RDF and then it is uploaded in a OWLIM-SE RDF Store instance [11]. Karma initialization phase involves loading the primary reference ontology and connecting dataset containing the data to be mapped. This process allowed the production of the knowledge base that may present a large set of problems due to inconsistencies and incompleteness that may be due to lack of relationships among different data sets, etc. These problems

may lead to the impossibility of making deductions and reasoning on the knowledge base, and thus on reducing the effectiveness of the model constructed. These problems have to be solved by using a reconciliation phase via specific tools and the support of human supervisors.

### C. Exploiting and Exploring Smart City Data

The Smart City Ontology presented is a strong generalization of a large set of data modeling problems. The integration of the several data sets coming from different sources into a semantic interoperable knowledge base is a solution to exploit this information for smart city purpose. To this end, the activities of data quality improvements can be performed in the first phases of the ingestion, and/or after the triple generation and integration to discover problems and to solve them.

The system has been used to ingest the data coming from the Municipality of Florence, the Tuscany Region and MIIC. Considering only files related to the daily weather forecast of all the available municipalities, we have 286 files updated twice a day, each of which, containing also 16 lines of weather prediction for the week, we obtain an increase of approximately 270,000 HBase lines per month that, in terms of triples, corresponds to a monthly increase of about 2.5 million triples.

Moreover, in order to explore the data being ingested and their relationships a tool for data visualization and exploration was used, that allows exploring the semantic graph of the relations among the entities, this Linked Open Graph is available for applications developers to explore and understand better the data available in the ontology.

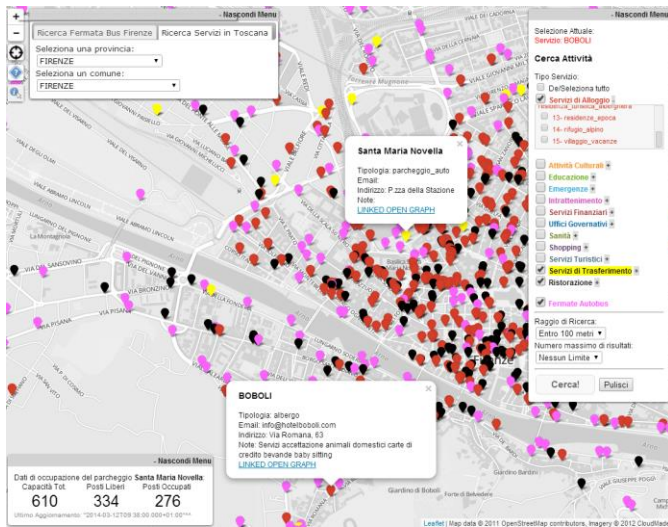


Figure 5 - Service Map (<http://servicemap.disit.org>)

A second tool called ServiceMap to perform geographic queries (for example to get points of interests close to a bus station, to get the street number close to a given point on the map, etc.) has been deployed (see Figure 5).

The service map, for example, allows to (i) get bus stops and from them to access the status line of the bus, providing the waiting time to the next bus, (ii) finding parking and getting in real time the number of empty places, etc. From each “pin” in

the map, it is possible to pass from the entity identified to its model in terms of relationships on the LOG graph.

## V. VERIFICATION AND VALIDATION

To connect services to the Street Guide in the repository a reconciliation phase in more steps, has been required, because the notation used by the Tuscany region in some Open Data within the Street Guide, does not always coincide with those used inside Open Data relating to different points of interest. In substance, different public administration are publishing Open Data that are not semantically interoperable.

Typical problems can be related to: (i) low quality of data, (ii) lack of data that are supposed to arrive in real time, (iii) changes in the data model of the data set, (iv) changes and updates into the data sets (this problem could generate a change into the ontological model and thus the human intervention is activated for model review), etc. To this end, periodic verification and validation processes is needed to be performed by defining a set of SPARQL queries on the knowledge base with the aim of detecting inconsistencies and incompleteness, and verifying the correct status of the model. These periodically executed queries perform a regression testing every time a new update of data process ingestion is performed, and when real time data arrive into the final RDF store. The validation process may lead to identify problems that may be limited to the instances of classes. To this end, the fourth information associated with each triple allows to identify the problems and the data set processes to be revised.

Therefore, an iterative workflow process was defined. During validation there were cases like the Weather forecast where no connection among the data were present due to different encoding of the name of the municipality, for this reason to support the reconciliation process a table containing the ISTAT code of each municipality was created, and each time new weather data are available, they are automatically completed with the correct ISTAT code, thus supporting the search for the instance of the PA class to which connect the weather forecasts.

A relevant process of data improvement for semantic interoperability is related to the application of reconciliations among the entities associated with locations as streets, civic numbers and localities. On this regard, there are different types of inconsistencies within the various integrated dataset, such as:

- typos;
- missing street number, or replacement with "0" or "SNC" (Italian acronym that means without civic number);
- Municipalities with no official name (e.g. Vicchio/Vicchio del Mugello);
- street names with uncommon characters ( -, /, ° ? , Ang., );
- street numbers with strange characters ( -, /, ° ? , Ang. ,(, );
- road name with words in a different order from the usual ( e.g. Via Petrarca Francesco, exchange of name and surname);
- number wrongly written (e.g. 34/AB, 403D, 36INT.1);
- red street numbers (in some cities, street numbers may have a color. So that a street may have 4/Black and 4/Red,



red is the numbering system for shops); Roman numerals in the road name (e.g., via Papa Giovanni XXIII).

As a summary, the whole knowledge base initially created was consisting of more than 81 Million triples, with a growth of 4 million triples per month. A part of them can be discharged when statistical values are estimated and punctual value discharged. For the validation, a total amount of services/points of interest inserted into the repository has been of 30182 instances. Among these, 13185 have been reconciled at street number-level, while the number of elements reconciled at street-level has been 21207. There are also 149 services associated to a coordinate pair, for which reconciliation did not return results, yet for the lack of references into the knowledge base (some streets and civic numbers are still missing or incomplete).

Thanks to the created ontology, is possible to perform services reconciliation at street number level, i.e. connecting an instance of class *Service* to an external access that uniquely identifies a street number on a road, or only at street-level, with less precision (lack that can be compensated thanks to geolocation of the service).

In the collected data sets, an average of about the 15% are automatically connected entities since they refer to perfectly consistent locations (i.e., perfect match in terms of location, street and civic number) in the MIIC with respect to the description reported in the service data set. In the total of location entities ingested, 5,75 % of locations are wrong and not reconcilable due to (i) the presence of wrong values for streets and/or locations, (ii) the lack of a consistent reference location into the MIIC geographical model.

The reconciliation process can be performed with the aim of finding elements that identify the same entity while presenting different URIs. Thus the identified reconciliations are solved creating an *owl:sameAs* triple to the selected location toponym. Reconciliation detection can be performed by using (i) a set of specific SPARQL queries, (ii) program tools for RDF link discovering. To this end, declarative languages for link discovering such as SILK [14] and LIMES [13] have been proposed. As the production of SPARQL queries, the programming of the link discovering algorithms also implies the knowledge of the ontological structure of the RDF stores to be compared/linked.

#### A. SPARQL Reconciliation

The methodology used for SPARQL reconciliation consists of trying to connect each service at street number-level, and then, perform the reconciliation at street-level. The first reconciliation step performed consists of an exact search of the street name associated with each service integrated. For example, to reconcile the service located at "VIA DELLA VIGNA NUOVA 40/R-42/R, FIRENZE", a SPARQL query is necessary, to search for all elements of *Road* class connected to the municipality of "FIRENZE" (via the *ObjectProperty inMunicipalityOf*), which have a name that exactly corresponds to "VIA DELLA VIGNA NUOVA" (checking both fields: official name, alternative name). The query results has to be filtered again, imposing that an instance of *StreetNumber* class

exists and it corresponds to civic number "40" or "42", with the R class code Red. A very frequent problem for exact search, is the existence of multiple ways to express toponym qualifiers, that is dug (e.g. Piazza and P.zza) or parts of the proper name of the street (such as Santa, or S. or S or S.ta): thanks to support tables, inside which the possible change of notation for each individual case identified are inserted, a second reconciliation step was performed, based on exact search of the street name, which has allowed to increase the number of reconciled services at street number-level. The third reconciliation step is based on the research of the last word inside the field *v:Street-Address* of each instance of the *Service* class, because, statistically, for a high percentage of street names, this word is the key to uniquely identify a match.

The above mentioned three steps have been also carried out without taking into account the street number, and so in order to obtain a reconciliation at street-level of each individual service. An additional, phase of *manual correction* has been also performed by manually (i) searching services and incongruent locations via web search service as Google, (ii) cleaning address and street number fields, (iii) accepting and performing association match of non-identified matches, taking into account the list of probable candidates suggested by the query results.

#### B. Link discovering Based Reconciliation and comparison

Link discovering based reconciliation consists in writing specific SILK algorithms for link discovering. They allow to discover links by writing specific algorithms grounded on distances and similarity metrics between patterns and relationships mainly based on string matching and distance measures (*Euclidean, weighted models, tree distances, patterns distance, string match, taxonomical, Jaro, Jaro-Winkler, Leveisthein, Dice, Jaccard, etc.*) [14].

In this case, a number of link discovering algorithms have been developed and assessed. Among them, the better ranked were based on comparing, at the same time, the location and the street. Firstly searching for the perfect match on location name and accepting uncertainty on street number from 0 up to 5 characters, for example. Both criteria have been aggregated considering their weight almost identical.

#### C. Reconciliation Comparison

The obtained results are reported in Table 1. The table reports the results assessed in terms of precision, recall and F1score (the F1 score is also called the F-measure, and it is defined as Harmonic mean of Recall and Precision) [15], in identifying the correct entities to be reconciliated. The first two lines refer to the SPARQL approach with and without manual intervention as described in Section V.A. The manual intervention has strongly improved the recall. On the other hand, the SPARQL approach is very time intensive for the programmers since a set of specific queries have to be produced for each data set to be reconciled. The second part of Table I reported the results related to different implementations of link discovering SILK based solutions, by using different string

distances (i.e., Leveisthein, Dice, and Jaccard), with the above mentioned values for their parameters. Other distance models have been also used without obtaining significant results. The last Link discovering solution has been coded by using an additional knowledge about all the specific strings coding problems reported in Section V.

Table I – Reconciliation Comparison

Method	Precision	Recall	F1
SPARQL –based reconciliation	1,00	0,69	0,820
SPARQL -based reconciliation + manual action	0,985	0,722	0,833
Link discovering - Leveisthein	0,927	0,508	0,656
Link discovering - Dice	0,968	0,674	0,794
Link discovering - Jaccard	1,000	0,472	0,642
Link discovering - Knowledge base + Leveisthein	0,925	0,714	0,806

## VI. CONCLUSIONS

In this paper, a system for the ingestion of public and private data for smart city with related aspects as road graph, services available on the roads, traffic sensors etc., has been proposed. The system includes both open data from public administration and private data coming from transport systems integrated managers, thus addressing and providing real time data of transport system, i.e., the busses, parking, traffic flows, etc. The system allows managing large volumes of data coming from a variety of sources considering both static and dynamic data. This data is then mapped to a Smart City Ontology and stored into an RDF-Store where this data are available for applications via SPARQL queries to provide new services to the users. The derived ontology has been obtained by means of an incremental process performed analyzing, integrating and validating each added data set. Thus the resulting ontology is a strong generalization of a large set of data modeling problems.

In addition, a thorough verification and validation process performed allowed us to identify the set of triples to: (i) improve and enrich the model, and (ii) perform the corrections. Thus improving and enabling the deductive capabilities of the final model. Finally, the proposed system also provides a visualization and exploration tool to explore the data available in the RDF-Store. As a conclusion, the performed assessment and comparison has produced a clear results demonstrating that the best quality of results are obtained by using the approach based on SPARQL queries plus some manual actions. Also the simple usage of SPARQL queries resulted to be better ranked with respect to the SILK based link discovering. On the other hand, the writing of link discovering algorithms resulted to be much simpler and faster than performing a set of specific SPARQL queries.

The next step will be to identify famous names, points of interest, locality names that can be linked to other data set as DBpedia<sup>2</sup> or GeoNames<sup>3</sup> according to a Linked Open Data

model. This process can be performed with a simple NLP algorithm [6], [8]. Furthermore an upcoming integration of the DISIT Ontology for Smart City with the GoodRelations model, is also planned, together with the automation of the reconciliation step, thanks to link discovery and machine learning techniques.

## ACKNOWLEDGMENT

A sincere thanks to the public administrations that provided the huge data collected and to the Ministry to provide the funding for Sii-Mobility Smart City Project, a warm thanks to Lapo Bicchielli, Giovanni Ortolani, Francesco Tuveri.

## REFERENCES

- [1] Caragliu, A., Del Bo, C., Nijkamp, P. (2009), Smart cities in Europe, 3rd Central European Conference in Regional Science – CERS, Kosice (sk), 7-9 ottobre 2009.
- [2] Bellini P., Di Claudio M., Nesi P., Rauch N., "Taxonomy and Review of Big Data Solutions Navigation", Big Data Computing To Be Published 26th July 2013 by Chapman and Hall/CRC
- [3] Vilajosana, I. ; Llosa, J. ; Martinez, B. ; Domingo-Prieto, M. ; Angles, A., "Bootstrapping smart cities through a self-sustainable model based on big data flows", Communications Magazine, IEEE, Vol.51, n.6, 2013
- [4] Ontology of Transportation Networks, Deliverable A1-D4, Project REVERSE, 2005 <http://reverse.net/deliverables/m18/a1-d4.pdf>
- [5] Pan, Feng, and Jerry R. Hobbs. "Temporal Aggregates in OWL-Time." In FLAIRS Conference, vol. 5, pp. 560-565. 2005.
- [6] Embley, David W., Douglas M. Campbell, Yuan S. Jiang, Stephen W. Liddle, Deryle W. Lonsdale, Y-K. Ng, and Randy D. Smith. "Conceptual-model-based data extraction from multiple-record Web pages." Data & Knowledge Engineering 31, no. 3 (1999): 227-251.
- [7] Auer, Sören, Jens Lehmann, and Sebastian Hellmann. "Linkedgeodata: Adding a spatial dimension to the web of data." In The Semantic Web- ISWC 2009, pp. 731-746. Springer Berlin Heidelberg, 2009.
- [8] Andrea Bellandi, Pierfrancesco Bellini, Antonio Cappuccio, Paolo Nesi, Gianni Pantaleo, Nadia Rauch, ASSISTED KNOWLEDGE BASE GENERATION, MANAGEMENT AND COMPETENCE RETRIEVAL, International Journal of Software Engineering and Knowledge Engineering, Vol.22, n.8, 2012
- [9] Apache HBase: A Distributed Database for Large Datasets. The Apache Software Foundation, Los Angeles, CA. URL <http://hbase.apache.org>.
- [10] Pentaho Data Integration, <http://www.pentaho.com/product/data-integration>
- [11] Barry Bishop, Atanas Kiryakov, Damyan Ognyanoff, Ivan Peikov, Zdravko Tashev, Ruslan Velkov, "OWLIM: A family of scalable semantic repositories", Semantic Web Journal, Volume 2, Number 1 / 2011.
- [12] S.Gupta, P.Szekely, C.Knoblock, A.Goel, M.Taheriyan, M.Muslea, "Karma: A System for Mapping Structured Sources into the Semantic Web", 9th Extended Semantic Web Conference (ESWC2012).
- [13] A. Ngomo, S. Auer. "LIMES: a time-efficient approach for large-scale link discovery on the web of data". Proc. of the 22nd int. joint conf. on Artificial Intelligence, Vol.3. AAAI Press, 2011.
- [14] R. Isele, C. Bizer. "Active learning of expressive linkage rules using genetic programming". Web Semantics: Science, Services and Agents on the World Wide Web 23 (2013): pp.2-15.
- [15] Powers, D.M.W. (February 27, 2011). "Evaluation from precision, recall and F-Measure to roc informedness, markedness and correlation". Journal of Machine Learning Technologies 2 (1): 37-63.

<sup>2</sup> <http://dbpedia.org/>

<sup>3</sup> <http://www.geonames.org/>

# Auto-encoder Based Bagging Architecture for Sentiment Analysis

Wenge Rong<sup>1,2</sup>, Yifan Nie<sup>3</sup>, Yuanxin Ouyang<sup>1,2</sup>, Baolin Peng<sup>1,2</sup>, Zhang Xiong<sup>1,2</sup>

<sup>1</sup>School of Computer Science and Engineering, Beihang University, Beijing 100191, China

<sup>2</sup>Research Institute of Beihang University in Shenzhen, Shenzhen 518057, China

<sup>3</sup>Sino-French Engineering School, Beihang University, Beijing 100191, China  
{w.rong@, yifan.nie@ecpk., oyyx@, b.peng@cse., xiongz@}buaa.edu.cn

## Abstract

*Sentiment analysis has long been a hot topic for understanding users statements online. Previously many machine learning approaches for sentiment analysis such as simple feature-oriented SVM or more complicated probabilistic models have been proposed. Though they have demonstrated capability in polarity detection, there exist one challenge called the curse of dimensionality due to the high dimensional nature of text-based documents. In this research, inspired by the dimensionality reduction and feature extraction capability of auto-encoders, an auto-encoder-based bagging prediction architecture (AEBPA) is proposed. The experimental study on commonly used datasets has shown its potential. It is believed that this method can offer the researchers in the community further insight into bagging oriented solution for sentimental analysis.*

**Keywords**—Sentiment Analysis, Bagging, Auto-encoder, Deep Learning

## I. Introduction

Sentiment analysis, which aims at extracting polarity orientation from a statement and identifying if the statement is positive or not, has become a hot topic in both academic and industrial sphere. As a powerful information gathering mechanism [1], sentiment analysis accumulates on-line documents ranging from twitters, blogs to customer reviews and tries to understand customers attitudes, opinion and emotion. It has been proven useful in different domains such as E-commerce [2], social media analysis [3] and political elections [4].

There are a lot of sentimental analysis approaches proposed in the literature and these methods can be roughly divided into two categories, i.e., computational linguistic approach and machine learning approach [5].

Computational linguistic approach is based on linguistic information and employs pre-defined sentiment lexicon where individual words are assigned a polarity score. The overall polarity of a statement is calculated by voting the scores of each words in the statement. Machine learning approach utilises machine learning models to perform sentiment analysis and regards sentiment analysis as a formal binary classification problem [6].

Computational linguistic approach employs pre-defined scores of words in a general context to detect the polarity information [7], while machine learning oriented solutions employs labelled data in the specific domain of the task to conduct sentimental analysis [5]. As a result machine learning approaches normally show better prediction capability as sentiment analysis is indeed a domain-dependent task [8].

Although machine learning approach presents better capability than computational linguistic approach, it still suffers from some challenges. One of them is the curse of dimensionality [9]. This is a typical phenomenon when the dimension of the feature space becomes larger, the volume of the space will increases quickly as such available training data will become sparse, thereby degrading the learning algorithms for sentiment analysis due to the high dimensionality of text dataset. In order to reduce the dimension of the raw data and extract higher order features from them to perform classification, many solutions have been developed and one popular implementation is stacked auto-encoder-based prediction model.

To build a stacked auto-encoder-based prediction model, a common approach is to stack several code layers of auto-encoders and further add one classification layer on the top of the last layer, and then take the learned features as input for the classification layer [10]. Though the auto-encoder-based prediction model has shown its promising potential, it still faces several difficulties among which a notable one is how to reduce generalisation error [11], which measures how well the model generalizes to data not participated in training. To meet this challenge, a lot of approaches have been proposed among which a widely adopted one is

ensemble methods [12].

Bagging (Bootstrap Aggregation) is a popular ensemble method which consists in bootstrapping several copies of the training set and then employing them to train separate models. Afterwards it combines the individual predictions together by a voting scheme for classification applications [13]. As each bootstrapped training set is slightly different from each other, each model trained on these training sets will have different weights and focus, thereby obtaining different generalisation errors. By combining them together, the overall generalisation error is expected to decrease to some extent.

Previous works have shown that bagging works well for unstable predictors [14]. Considering stacked auto-encoder-based prediction models are also unstable predictors [15], it is intuitive to assume that applying bagging methods to auto-encoder-based models could improve classification performance. Inspired by this assumption, in this paper an auto-encoder-based bagging prediction architecture (AEBPA) is proposed to integrate stacked auto-encoder-based prediction models for sentimental analysis. The experimental study on commonly used datasets also shows its promising potential.

The main contribution our work is two-fold. 1) We thoroughly investigate the possibility of integrating feature learning with bagging ensemble method on text-based dataset and empirically evaluated the pros and cons of several integration schemas with different number of bagging sets. 2) We propose an auto-encoder-based bagging prediction architecture (AEBPA) for sentiment analysis and empirical study shows that our approach outperforms traditional methods.

The remainder of this paper is organised as follow. Section II will introduce the background and Section III will illustrates the proposed methodology. In section IV the experimental study will be presented and Section V will conclude the paper and point out possible future work.

## II. Background

### A. Sentiment Analysis

Sentiment analysis is a powerful mechanism to obtain people's opinion and tell if their overall attitudes are favourable or not. It has enjoyed a huge burst of research activity in recent years [1]. For example, sentimental analysis can be used to analyse the users purchasing behaviour as more and more customers take notice of the review of other people before buying a product on-line [16]. It can also be used for predicting political election results [17]. As sentiment analysis becomes an important tool in different domains, many techniques have been proposed and can be roughly categorized into computational linguistic approach and machine learning approach.

Computational linguistic approach employs pre-explored linguistic information where a lexicon is developed [18], either from exterior [19] or through heuristics during sentiment analysis [20]. In this kind of approaches, semantic information is added into the process and every word in the lexicon is attached with a polarity score. The process will further calculate the overall polarity of the statement by voting the polarity score of each word [21]:

$$p = \text{sign}\left(\sum_{i=1}^n q_i + 0.5p_0\right) \quad (1)$$

where  $q_i$  stands for polarity score of each word,  $\text{sign}$  stands for the sign function and  $p_0$  stands for the major polarity in the training data.

Different from computational linguistic approach, machine learning approach utilizes machine learning algorithms and treat sentiment analysis as a formal binary classification problem. Traditional methods convert a statement into a bag-of-word representation, extract other auxiliary features such as part-of-speech information, bi-grams and specific negation words, and then throws these features into SVM, MaxEnt or Naive Bayes based classifiers [22][23][24]. Though these methods are simple and robust, they are insufficient to reveal the underlying relationship between words and the overall sentiment polarity of a statement.

To deal with the shortcomings of simple machine learning methods, many more complex model are proposed among which Socher et al's recursive auto-encoder-based model [25] is a notable one and has shown its effectiveness in sentiment analysis. In this work, word indices are first mapped through a embedding matrix into semantics word vectors. Then the semantic word vectors of the words in the input sentence are lined up in their original sequence. Afterwards, the input word vectors are recursively merged into a fixed size vector representation in the following way.

Given a sequenced list of word vector nodes  $X = (x_1, x_2, x_3, \dots, x_n)$  an auto-encoder first tries to merge all neighbouring couples  $(x_i, x_{i+1})$  from  $X$  by the following formula:

$$p = f(W_1[c_1, c_2] + b_1) \quad (2)$$

Where  $(c_1, c_2) = (x_i, x_j)$  and  $f$  is the activation function,  $p$  is the calculated parent node vector and  $W_1, b_1$  are the weight matrix and bias vector of the auto-encoder, respectively. Then the auto-encoder tries to decode the parent node and calculate the reconstruction error by the following formula:

$$[c'_1, c'_2] = W_2 p + b_2 \quad (3)$$

$$E_{rec}([c_1, c_2]) = \frac{1}{2} \|[c'_1, c'_2] - [c_1, c_2]\|^2 \quad (4)$$

Where  $[c'_1, c'_2]$  are the reconstructed pair and  $W_2, b_2$  are the decoding weight matrix and bias vector respectively.  $E_{rec}$  is the reconstruction error. Then, the pair  $(x_i, x_{i+1})$

with the lowest reconstruction error is chosen to be 2 child nodes and merged to form a parent node  $p$ . Afterwards, the process repeats with new vector node list  $X' = (x_1, x_2, \dots, x_{i-1}, p, x_{i+2}, \dots, x_n)$  until the tree is fully constructed.

Once the tree is recursively constructed, the vectors at each node are used as features to predict the sentiment label of the sentence with the following objective function:

$$J = \frac{1}{N} \sum_{(x,t)} E(x, t; \theta) + \frac{1}{2} \|\theta\|^2 \quad (5)$$

Where  $N$  is number of training instances,  $(x, t)$  is a training (sentence, label) pair, and  $\theta$  represents the parameters of the model.

Although traditional machine learning have many advantages, it still faces several challenges. One of them is the curse of dimensionality [9]. As the dimension of the feature space increases, the volume of space increases exponentially and given a fixed number of training instances, the distribution of the training data in the feature space would become sparse, which may harm the predictive power of the learned model. As datasets for sentiment analysis are text-based, they are of high dimension when expressed in its raw representation [5].

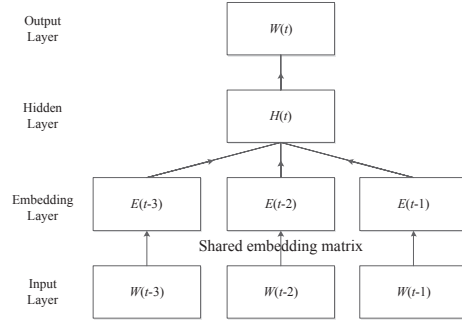
To meet this challenge of reducing the dimension of the training data, one possible implementation is auto-encoder which maps high dimension raw features into a reduced feature space. However as a neural-network-based model, a good initialisation of the connection weights is crucial to the optimisation process [26]. Traditional methods employ random initialisation which does not capture semantic information. To deal with this problem, word embedding is widely employed as it is able to capture semantic information between different words in the training data.

## B. Word Embedding

In neural-network-based models, how to better represent a statement composed by a series of words into a vector has been researched for many years. Traditional bag-of-words-based representations are not optimal because they do not take the sequential information into account [27], while embeddings learned by neural-network-based language model maps the integer vector into a continuous space and takes information from surrounding words.

The first popular neural-network-based language model proposed by Bengio et al. utilised a four layers feed-forward network [28], as shown in Fig. 1 where an input layer composed of the indexes of  $n$  consecutive words in a corpus, a linear projection layer, a non-linear layer and a softmax output layer which outputs the probability of a certain word being next word. The key feature of word embedding is that words with similar meaning or similar grammatical function will have similar embedded vectors.

As such embedding can be seen as a measure of semantic distance between words [29].



**Fig. 1. Feed-Forward Neural Network Language Model**

In addition to the basic feed-forward model, many other neural network language models have also been proposed to learn word embedding, among which recurrent neural network language model, CBOW and Skip-gram gained much popularity [30][31]. Recurrent neural network language model employs an Elman recurrent neural network [32] with three layers, i.e., an input layer composed of word indexes, a hidden layer with recursion and a softmax output layer to predict the probability of the next word based on input words. As there exists a recursion in the hidden layer, the recurrent neural network language model can form short term memory and thus overcomes the limitation of traditional feed-forward language model which has to specify the number of previous word  $n$  used in input [31].

In order to reduce computational complexity, some log-linear models have been proposed and CBOW and Skip-gram model are most commonly used. The CBOW model has three layers. The input layer contains  $m$  word indexes which form a word window around the current training word  $W$  in a subsequent way, i.e.  $m/2$  words before  $W$  and  $m/2$  words after  $W$ . This layer is connected by a shared weight matrix to a linear projection layer. The output layer contains the index of current word  $W$ . Whereas the Skip-gram model is a upside-down version of the CBOW where the input is the index of current word which will be put to a log-linear classifier with continuous projection layer, and predicts words within a certain range before and after the current word.

Empirical study shows that Skip-gram outperforms other model in semantic-related tasks [31]. Therefore, in this paper we employ the Skip-gram model to train word embedding to initialise the weights between the input layer and hidden layer of the proposed auto-encoder-based bagging classifier to conduct sentiment analysis.

### III. Methodology

#### A. Auto-encoder-based Bagging Prediction Architecture for Sentiment Analysis

In order to validate the possibility of integrating feature learning and bagging method on sentiment analysis, in this paper, an auto-encoder-based bagging prediction architecture (AEBPA) is proposed, as presented in Fig. 2.

In this architecture, three major parts are included, i.e., the bootstrapping process, auto-encoder-based prediction models and decision aggregating process. Given a sentiment analysis training set  $T$ , firstly  $n$  copies of bootstrapped training sets are generated through bootstrap techniques. Afterwards,  $n$  different auto-encoder-based prediction models are trained and their parameters are stored in hypothesis models  $h_1 \dots h_n$ , respectively. Once the training process is completed, a sentiment analysis test set can be fed into the hypothesis models  $h_1 \dots h_n$  and their prediction results  $p_1 \dots p_n$  are calculated accordingly. Finally, through voting aggregation process, the final verdict is obtained based on all sub-prediction  $p_1 \dots p_n$  made by  $n$  models.

#### B. Bootstrap Process

Bagging is a widely used integration technique to combine several prediction models in order to improve accuracy [33]. Whereas bootstrapping is the first step of bagging in order to generate several copies of the original training set. Given a sentiment training set  $X$  consisting of  $m$  training instances  $X = \{x_1, x_2, \dots, x_m\}$ , it is able to generate each bootstrapped copy of the original training set by sampling with replacement  $m$  elements with equal probability  $1/m$  from the original training set  $X$ . By repeating the procedure for  $n$  times,  $n$  bootstrapped training sets can be generated. Since the sampling is conducted with replacement, there will be repeated examples in the bootstrapped training set, and if  $m$  is large, asymptotically the fraction of unique examples would be:

$$\lim_{m \rightarrow \infty} 1 - (1 - \frac{1}{m})^m = 63.2\% \quad (6)$$

Sampling with replacement is important for bagging. Because of this rule, there would be repeated instances in each bootstrapped training set, therefore the unique training instances in each bootstrapped training set are randomly different. As such the models trained on these bootstrapped training sets have different focuses. Afterwards, the  $n$  bootstrapped training sets can be fed into  $n$  auto-encoder-based prediction models, which will be presented in the following subsection.

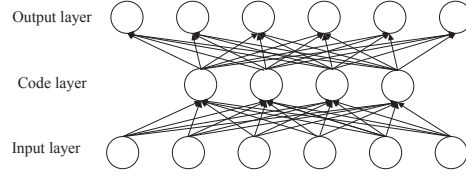


Fig. 3. The Structure Of An Auto-Encoder

#### C. Auto-encoder-based Prediction Model

The proposed AEBPA architecture employs and trains  $n$  auto-encoder-based prediction models simultaneously using  $n$  bootstrapped copies of training sets prepared in subsection III-B. Each model is an auto-encoder-based classification model to integrate dimensionality reduction and unsupervised feature learning into prediction models. The model is composed by two parts: 1) the stacked auto-encoder-based feature learning layers, and 2) the classification layer. The stacked auto-encoder-based feature learning part can have multiple layers in order to obtain higher-level representation of the data and performs feature learning [34], while the classification layer is on top of the feature learning layers and takes the learned features as input and performs classification.

The stacked auto-encoder can be seen as many code layers of auto-encoders stacked together. Its basic component, auto-encoder, has been widely used as an unsupervised feature learning tool [35]. Typically, a basic auto-encoder consists of three layers, i.e., the input layer, the hidden layer (also called code layer), and the output layer [36] which has the same number of units as the input layer, as shown in Fig. 3. This structure is able to achieve unsupervised feature learning by feeding the unlabelled data into the input layer and forcing the output layer to reconstruct the input.

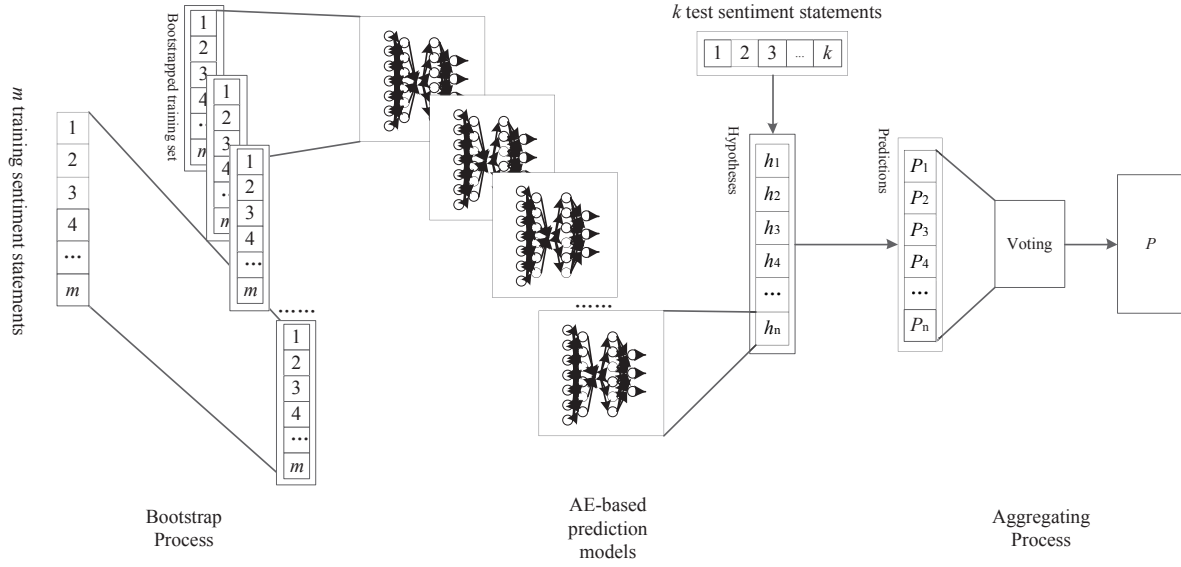
As for the training process of each stacked auto-encoder-based prediction model, a two-stage training approach is adopted. Firstly the proposed model performs a layer-wise unsupervised pre-training with unlabelled data [37]. This procedure aims to obtain unsupervised learned features in code layers. Afterwards the entire network including both code layers and classification layer can be fine-tuned by traditional back-propagation mechanism.

In order to perform layer-wise pre-training, the AEBPA method decomposes the stacked auto-encoder into auto-encoders by taking subsequent layers  $L$  and  $L + 1$  as input layer and hidden layer of an auto-encoder and forcing it to reconstruct the input layer  $L$  at output.

Therefore in the encode phase the input  $x \in \mathbb{R}^m$  will be mapped to a hidden representation (code)  $h \in \mathbb{R}^n$ , which has the form:

$$h = f(x) = s(Wx + b_h) \quad (7)$$





**Fig. 2. AEBPA Architecture**

where  $s$  is the sigmoid activation function  $s(z) = \frac{1}{1+e^{-z}}$ ,  $m$  is the number of input units,  $n$  is the number of hidden units,  $x$  is the input feature vector,  $h$  is the extracted code and the encoder is parametrised by the  $n \times m$  weight matrix  $W$ , and the bias vector  $b_h$ .

Afterwards it is able to force the auto-encoder to map the hidden representation back to a reconstruction  $\hat{x} \in \mathbb{R}^m$ :

$$\hat{x} = g(h) = s(W^T h + b_{\hat{x}}) \quad (8)$$

where  $s$  is the sigmoid activation function  $s(z) = \frac{1}{1+e^{-z}}$ ,  $m$  is the number of output units,  $n$  is the number of hidden units,  $h$  is the extracted code,  $\hat{x}$  is the reconstructed input feature vector, and the parameters are a bias vector  $b_{\hat{x}}$  and the tied weight matrix  $W^T$ .

As for the training process of an auto-encoder, the following cost function is employed to find the parameters  $\theta = \{W, b_h, b_{\hat{x}}\}$ , which minimize  $J$ :

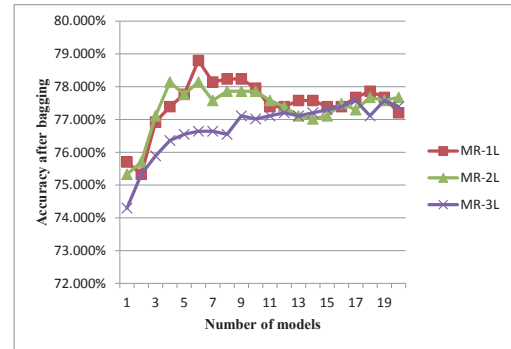
$$J(\theta) = \sum_{x \in \text{trainset}} L(x, g(f(x))) \quad (9)$$

where  $x$  is a feature vector and  $f, g$  are encode and decode function in Eqs. 7 and 8. In this paper, the squared error is utilised as the reconstruction error  $L$ , i.e.,  $L(x, \hat{x}) = \|x - \hat{x}\|^2$ .

Once the training process is completed, it is able to take the code in the hidden layer as unsupervised learned features and construct stacked auto-encoder. Consequently the process for layers  $L + 1$  and  $L + 2$  will be repeated until all code layers in the stacked auto-encoder are pre-trained. Afterwards the entire network will be further fine-tuned to include both the code layers and the classification

layer with labelled data by traditional back-propagation algorithm [38].

The stacked auto-encoder is configured with 0 sparsity penalty and no denoising treatment. All activation functions are sigmoid function and in order to determine the best code layer number, a preliminary study is conducted on one popular dataset (Movie Review, which will be presented in subsection IV-A) with 1, 2, 3 code layers in the proposed model, respectively. The result is shown in Fig. 4.



**Fig. 4. Performance With 1,2 And 3 Code Layers On Movie Review Dataset**

This empirical study can help determine the number of code layers used in the proposed model. As shown in Fig. 4, single code layer outperforms multiple code layers in the best configuration. Therefore, in this paper single code layer is employed.

In this study, the number of units in the input layer  $n_1$  is chosen to be equal to the size of vocabulary because the input vector is in a bag-of-words format. For the Movie Review dataset,  $n_1 = 21176$ , and for the IMDB dataset,  $n_1 = 89527$ . The size of the hidden layer (code layer)  $n_2$  is set to 200, which corresponds to our embedding size. And the size of output layer  $n_3$  is 2 which corresponds to the 2 possible sentiment labels, i.e. positive or negative. For each single model, the training process stops after 100 iterations of the code layer learning and 700 iterations of fine-tuning of the entire network. This generally ensures a desired training error which is less than 0.1%.

#### D. Aggregating process

Once all models are trained, all information will be stored in the parameters of each model and  $n$  hypothesis function  $h_1 \dots h_n$  is obtained to represent the  $n$  models and give prediction when fed with test instances. As mentioned in subsection III-B, each trained model has different focus. Therefore, given a test example  $x$ , the predicted labels  $h_1(x)$ ,  $h_2(x) \dots h_n(x)$  will not always be the same.

Suppose  $y$  is one possible label,  $Y$  is the set containing all possible labels,  $h_i$  is a hypothesis function representing prediction model  $i$ , and  $h^*$  is the combined prediction function, in the aggregating process each trained prediction model can be combined in following way:

$$h^*(x) = \arg \max_{y \in Y} \sum_{i: h_i(x)=y} 1 \quad (10)$$

In this research the number  $n$  is set to vary from 1 to 20. The decision of limiting the number of trained basic models is important in order to complete the experimental study in a reasonable time and for practical applications as indicated by Bauer [39].

#### E. Embedding-Boosted Auto-encoder For Sentiment Analysis

Neural-network-based model may get stuck at many local optima during training [40][41]. Therefore a satisfying initialisation of the weights of the model is crucial to a successfully-trained model. As mentioned in section II, embedding can help better initialise the weights and Skip-gram outperforms other popular models in semantic-related tasks [31]. Thus in this paper the Skip-gram is employed to train word embedding and initialize the weight between the input layer and hidden layer. This paper did not take pre-trained and off-the-shelf embeddings available online because word embedding is domain-specific and training word embeddings with corpus in the domain of study is expected to give better performance.

Skip-gram uses each current word as an input to a log-linear classifier and predicts words within a certain range

before and after the current word, as shown in Fig. 5. As a result it is necessary to maximise the following objective function and employ the statements in testsets for training:

$$\sum_{y(t) \in \text{Range}(w(t))} \log(p(y(t)|w(t); \theta)) \quad (11)$$

where  $W(t)$  is the current word,  $y(t)$  is the word in a context window before and after the current word, and  $\theta$  is the parameter of the model.

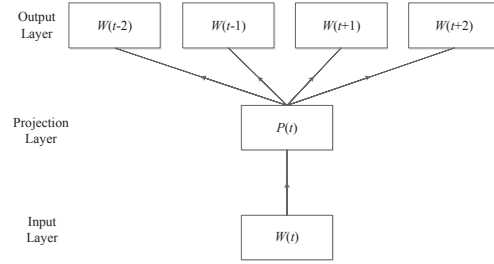


Fig. 5. Skip-Gram Language Model

Once the embeddings are trained, it can be used to initialise the weight between the input layer and hidden layer. The original embedding trained for each word is a real vector of dimension  $m$  (in our situation  $m = 200$ ). We then convert all the embedding vectors into column vectors and concatenate them together according to our dictionary order to form a  $m \times n$  matrix  $W_{emb}$  where  $m$  is the dimension of the embedding size which corresponds to the hidden layer size of the auto-encoder-based neural network, and  $n$  is the size of vocabulary. Afterwards, rather than randomly initialising the weight matrix  $W_{12}$  linking the input layer and the hidden layer, we initialise it by using directly the trained embedding matrix  $W_{emb}$  as  $W_{12}$ . As such, once the model is initialised, the weight matrix  $W_{12}$  already contains semantic and syntactic information. And after the initialisation is done, the embedding-boosted model can be put into the evaluation architecture presented in Fig. 2.

### IV. Experimental Study

#### A. Dataset

To validate and evaluate the proposed auto-encoder-based bagging prediction architecture, two commonly used datasets for sentiment analysis are employed, i.e., Movie Review<sup>1</sup> and IMDB<sup>2</sup>. Both of them contain sentiment reviews about different movies. The details of the datasets are presented in Table I.

<sup>1</sup><https://www.cs.cornell.edu/people/pabo/movie-review-data/>

<sup>2</sup><http://ai.stanford.edu/~amaas/data/sentiment/>

Most of the reviews in Movie Review are short summaries such as: “An idealistic love story that brings out the latent 15-year-old romantic in every one” or “A real movie about real people that gives us a rare glimpse into a culture most of don’t know”. And some shortest review are even just commentary phrases like “thoroughly enjoyable” or “A rare beautiful movie”. Therefore sentiment words are distributed in a highly dense manner in the statement.

On the contrary, the IMDB dataset contains statements which are much longer and have more details. The average length is around 225 words. And typical comments are like “If you like adult comedy cartoons, like South Park, then this is nearly a similar format about the small adventures of three teenage girls at Bromwell High. Keisha, Natella and Latrina have given exploding sweets and behaved like bitches ... The cast is also fantastic, Lenny Henry’s Gina Yashere... I didn’t know this came from Canada, but it is very good. Very good!”. In the statement, the author includes not only commentary information, but also brief scenario of the movie, actors and other information. The sentiment words are scattered very loosely in the statement, which adds difficulty to sentiment analysis models to capture the underlying polarity patterns of a statement.

As for the splitting strategy, for the Movie Review dataset, we adopt the common strategy used in Nakagawa et al.’s work [21] to split 90% of the dataset into training set and the rest 10% into test set, and this also allow us to make comparison of our results with others on a equal basis. For the IMDB dataset, the author of this dataset has already split the dataset into 50%/50% scheme [42], therefore we adopt this splitting as other users of this dataset do.

**TABLE I. Sentiment Analysis Dataset**

Dataset	#training instances	#test instances	positive/negative	average length
Movie Review	9596	1066	0.5/0.5	19
IMDB	25000	25000	0.5/0.5	225

## B. Evaluation Metrics

In order to analyse the performance of different sentiment analysis methods, the accuracy of prediction [21][42] is utilised to evaluate the prediction performance:

$$Accuracy = \frac{\sum_{i=1}^n \mathbb{1}\{t_i == o_i\}}{\#test\ instances} \quad (12)$$

where  $t_i$  stands for the desired label of a test instance and  $o_i$  stands for the output label predicted by the model. As such, the generalisation power of the proposed model towards unseen data is evaluated through its capability to predict accurately the sentiment label of test instances.

## C. Baseline Methods For Sentiment Analysis

In order to compare the potential of the proposed auto-encoder-based bagging approach, several baseline methods are included for comparison, as listed below:

For the Movie Review dataset, Nakagawa et al.’s work [21] shows that the following four methods are most suitable for the Movie Review dataset.

- 1) Bag-of-Words: This method employs bag-of-words features and a simple logistic regression unit to classify if a statement is positive or negative.
- 2) Vote by lexicon: This method employs lexicon developed in [43] [44] and the polarity of a statement is calculated by voting each word’s polarity score.
- 3) Rule-Based Reversal: The polarity of a subjective sentence is deterministically decided based on rules, by considering the sentiment polarities of dependency subtrees. The polarity of the dependency subtree whose root is the  $i$ -th phrase is decided by voting the prior polarity of the  $i$ -th phrase and the polarities of the dependency subtrees whose root nodes are the modifiers of the  $i$ -th phrase.
- 4) Tree-Based CRF: This method employs conditional random fields with hidden variables to classify sentiment polarities.

For the IMDB dataset, Maas et al.’s study [42] reveals that the following four methods show better performance and are more suitable for this dataset.

- 1) LSA: A truncated SVD decomposition is applied to the tf-idf weighted, cosine normalized count matrix, which is a standard weighting and smoothing scheme for VSM model.
- 2) LDA: Topic modeling method proposed by Blei et al. [45] modelling statements as mixtures of latent topics.
- 3) MAAS Semantic: Probabilistic model proposed by Maas et al. [42] which optimises a objective function only containing semantic information.
- 4) MAAS Full: Probabilistic model proposed by Maas et al. [42] which optimises a objective function containing both semantic and sentiment information.

## D. Experimental Results

Tables II and III summarises the results of the proposed approach with comparison to other baseline methods, which shows that the proposed bagging approach can outperform other baseline methods.

**TABLE II. Accuracy Summary For Movie Review Dataset**

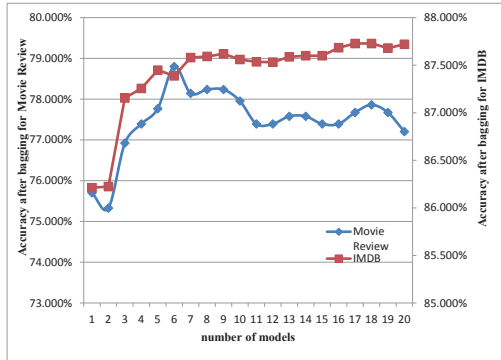
Method	Movie Review
Bag-of-Word	0.764
Voting by lexicon	0.631
Rule-Based Reversal	0.629
Tree-Based CRF	0.773
AEBPA (Single Model without embedding)	0.764
AEBPA (Single Model with embedding)	0.766
AEBPA	0.788

**TABLE III. Accuracy Summary For IMDB Dataset**

Method	IMDB
LSA	0.8396
LDA	0.6742
MAAS Semantic	0.8730
MAAS Full	0.8744
AEBPA (Single Model without embedding)	0.8611
AEBPA (Single Model with embedding)	0.8669
AEBPA	0.8773

Figure 6 presents the result of bagging over the number of models participating in bagging for the two datasets. As the number of models participating in bagging increases, the accuracy after bagging either first increases and then decreases after a optimal value or increases and stagnates around a limit value. In the case of Movie Review, bagging 6 models yields the best result of 78.8% in accuracy. For IMDB, bagging 17 models gives an optimal result of 87.73%. This phenomenon might be partially attributed to the fact that if there are too many models participating in bagging, their outputs would be too divergent, and then the disagreement between different models may lead to a stasis or even a decrease in performance.

The training process is offline, however once the training is done, the prediction process is online and works in real time.

**Fig. 6. Accuracy After Bagging For Sentiment Analysis**

## E. Discussion

As presented in Tables II and III, the proposed approach outperforms other baseline methods. Vote-based method got poor performance partially because it only employs lexicon to vote based on each word's polarity. However, sentiment lexicon is context-dependent, the same word may have different polarity in different domain, and there are no universal lexicon [46][47]. Even worse, the same word in the same domain may have different polarity with respect to different aspects. For example, in a laptop review, 'large' may be negative for the battery aspect while being positive for the screen size aspect.

Bag-of-word method did not achieve best result neither, because the text-based data expressed in a bag-of-word representation are of high dimension. Simply feeding high dimension text-based data into logistic regression learner cannot help capture the underlying sentiment patterns in a statement.

The proposed AEBPA architecture help achieve better result because 1) each single model in the architecture is an auto-encoder-based model which has the capability to perform dimension reduction and extract higher order features from raw representations in an unsupervised manner; 2) Balanced by bagging techniques, the generalisation error of the whole architecture is further decreased; 3) Finally, for the two datasets, word embeddings is used to initialise the weights between input layer and hidden layer in order to avoid the situation where parameters get stuck in local optima during training process.

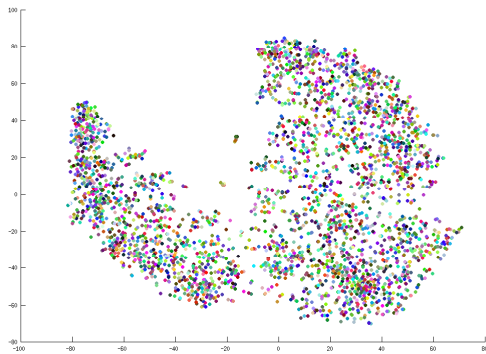
From the comparison result of Single Model without embedding and with embedding in Tables II and III, it is observed that embedding does help boost the performance. To visualise what the proposed model can learn, the embedding learned by the proposed model is further projected using a commonly used dimension reduction tool [48], as shown in Fig. 7, where each word is represented by one point. It is observed that words are roughly separated into two parts which is consistent with the expected classification task since the output labels are in two categories. The lower-left part contains words with positive polarity whereas the right part are negative words. Some typical words are summarised in Table IV. As shown in Fig. 7, words with the same sentiment polarity are clustered near each other and words with close semantic meaning are close to each other. This learned embedding, the dimension reduction and feature learning capability of auto-encoder plus bagging techniques may account for the polarity detection performance.

## V. Conclusion and Future Work

In recent years, many methods for sentiment analysis have been proposed among them machine learning approach such as feature-oriented SVM or more complicated

**TABLE IV. Words Close To Each Other In The Embedding Learned By AEBPA**

Words in the lower-left part	elegant, admirable, amazing, best, charming, unbelievable, wonderful, beneficial, cheer, elegant, encouraging
Words in the right part	loathing, tired, boring, capricious, deceitful, sad, stressful, struggle, stupid, tedious, terribly, trouble, unfaithful, unfavourable



**Fig. 7. Embedding Learned By Auto-Encoder-Based Prediction Model**

probabilistic models are widely employed due to their capability of learning domain specific information. Though these methods have shown their promising potential and robustness, there are still several challenges such as the curse of dimensionality because text-based data are often of high dimension when represented in raw representation. The auto-encoder-based prediction model has been proven useful to overcome this difficulty, while how to reduce generalisation error of the model still keeps an important challenge. There are many ways to reduce generalisation error among which bagging is one of possible methods. To better investigate the feasibility of bagging different auto-encoder-based prediction models for sentiment analysis, in this study, we proposed an auto-encoder-based bagging prediction architecture (AEBPA) for sentiment analysis and conducted experiment on standard datasets. The experimental study shows that the proposed architecture outperforms traditional methods.

Although the proposed approach achieved satisfying results, some challenges still need to be addressed. One of them is how to better initialize the weights of the model. It is worthwhile to employ more data to train a better embedding and in the future it is deserved to use more sentimental corpus out of the experimental datasets to train embedding. As for optimisation process, it is also worthwhile to employ other optimisation algorithms such as Newton [49] and LBFGS [50].

## Acknowledgment

This work was partially supported by the National Natural Science Foundation of China (No. 61103095), and the National High Technology Research and Development Program of China (No. 2013AA01A601). We are grateful to Shenzhen Key Laboratory of Data Vitalization (Smart City) for supporting this research.

## References

- [1] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008.
- [2] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1-167, 2012.
- [3] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment in twitter events," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 2, pp. 406-418, 2011.
- [4] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle," in *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, 2012, pp. 115-120.
- [5] K. Kim and J. Lee, "Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction," *Pattern Recognition*, vol. 47, no. 2, pp. 758-768, 2014.
- [6] E. Boiy and M.-F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," *Information retrieval*, vol. 12, no. 5, pp. 526-558, 2009.
- [7] K. Denecke, "Are sentiwordnet scores suited for multi-domain sentiment classification?" in *Proceedings of the 4th IEEE International Conference on Digital Information Management*. IEEE, 2009, pp. 1-6.
- [8] Y. Yoshida, T. Hirao, T. Iwata, M. Nagata, and Y. Matsumoto, "Transfer learning for multiple-domain sentiment analysis-identifying domain dependent/independent word polarity," in *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.
- [9] E. Keogh and A. Mueen, "Curse of dimensionality," in *Encyclopedia of Machine Learning*. Springer, 2010, pp. 257-258.
- [10] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proceedings of the 28th International Conference on Machine Learning*, 2011, pp. 513-520.
- [11] Y. Bengio, "Deep learning of representations: Looking forward," in *Statistical Language and Speech Processing*. Springer, 2013, pp. 1-37.
- [12] R. Maclin and D. Opitz, "Popular ensemble methods: An empirical study," *arXiv preprint arXiv:1106.0257*, 2011.
- [13] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123-140, 1996.
- [14] G. Liang, X. Zhu, and C. Zhang, "An empirical study of bagging predictors for different learning algorithms," in *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.
- [15] L. Yu, S. Wang, and K. K. Lai, "A neural-network-based nonlinear metamodeling approach to financial time series forecasting," *Applied Soft Computing*, vol. 9, no. 2, pp. 563-574, 2009.
- [16] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*. Springer, 2012, pp. 415-463.

- [17] P. T. Metaxas, E. Mustafaraj, and D. Gayo-Avello, "How (not) to predict elections," in *Proceedings of the 3rd IEEE international conference on privacy, security, risk and trust (PASSAT) and the 3rd IEEE international conference on social computing (Social-Com)*. IEEE, 2011, pp. 165–171.
- [18] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational linguistics*, vol. 37, no. 2, pp. 267–307, 2011.
- [19] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proceedings of the International Conference on Language Resources and Evaluation*, vol. 10, 2010, pp. 2200–2204.
- [20] Y. Choi and C. Cardie, "Learning with compositional semantics as structural inference for subsentential sentiment analysis," in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2008, pp. 793–801.
- [21] T. Nakagawa, K. Inui, and S. Kurohashi, "Dependency tree-based sentiment classification using crfs with hidden variables," in *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics*. Association for Computational Linguistics, 2010, pp. 786–794.
- [22] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*. Association for Computational Linguistics, 2002, pp. 79–86.
- [23] T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, vol. 4, 2004, pp. 412–418.
- [24] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," *Journal of Informetrics*, vol. 3, no. 2, pp. 143–157, 2009.
- [25] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 2011, pp. 151–161.
- [26] G. Thimm and E. Fiesler, "Neural network initialization," in *From Natural to Artificial Neural Computation*. Springer, 1995, pp. 535–542.
- [27] N. Liu, B. Zhang, J. Yan, Z. Chen, W. Liu, F. Bai, and L. Chien, "Text representation: From vector to tensor," in *Proceedings of the 5th IEEE International Conference on Data Mining*. IEEE, 2005, pp. 4–pp.
- [28] Y. Bengio, H. Schwenk, J.-S. Senécal, F. Morin, and J.-L. Gauvain, "Neural probabilistic language models," in *Innovations in Machine Learning*. Springer, 2006, pp. 137–186.
- [29] J. Turian, L. Ratinov, and Y. Bengio, "Word representations: a simple and general method for semi-supervised learning," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010, pp. 384–394.
- [30] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, and S. Khudanpur, "Recurrent neural network based language model," in *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association*, 2010, pp. 1045–1048.
- [31] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
- [32] J. L. Elman, "Distributed representations, simple recurrent networks, and grammatical structure," *Machine learning*, vol. 7, no. 2-3, pp. 195–225, 1991.
- [33] P. Bühlmann, "Bagging, boosting and ensemble methods," in *Handbook of Computational Statistics*. Springer, 2012, pp. 985–1022.
- [34] Y. Bengio, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [35] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," *Journal of Machine Learning Research-Proceedings Track*, vol. 27, pp. 37–50, 2012.
- [36] S. Rifai, G. Mesnil, P. Vincent, X. Muller, Y. Bengio, Y. Dauphin, and X. Glorot, "Higher order contractive auto-encoder," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2011, pp. 645–660.
- [37] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?" *Journal of Machine Learning Research*, vol. 11, pp. 581–616, 2010.
- [38] H. H. Örkücü and H. Bal, "Comparing performances of backpropagation and genetic algorithms in the data classification," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3703–3709, 2011.
- [39] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants," *Machine learning*, vol. 36, no. 1-2, pp. 105–139, 1999.
- [40] S. Ding, C. Su, and J. Yu, "An optimizing bp neural network algorithm based on genetic algorithm," *Artificial Intelligence Review*, vol. 36, no. 2, pp. 153–162, 2011.
- [41] D. Yu and L. Deng, "Accelerated parallelizable neural network learning algorithm for speech recognition," in *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association*, 2011, pp. 2281–2284.
- [42] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2011, pp. 142–150.
- [43] P. J. Stone, D. C. Dunphy, and M. S. Smith, "The general inquirer: A computer approach to content analysis." *MIT press*, 1966.
- [44] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, p. 2001, 2001.
- [45] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.
- [46] Y. Lu, M. Castellanos, U. Dayal, and C. Zhai, "Automatic construction of a context-aware sentiment lexicon: An optimization approach," in *Proceedings of the 20th International Conference on World Wide Web*, 2011, pp. 347–356.
- [47] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *Proceedings of the International Conference on Web Search and Web Data Mining*. ACM, 2008, pp. 231–240.
- [48] L. van der Maaten and G. Hinton, "Visualizing non-metric similarities in multiple maps," *Machine learning*, vol. 87, no. 1, pp. 33–55, 2012.
- [49] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 9–48.
- [50] D. Sheppard, R. Terrell, and G. Henkelman, "Optimization methods for finding minimum energy paths," *The Journal of chemical physics*, vol. 128, no. 13, p. 134106, 2008.



# Towards a Trust, Reputation and Recommendation Meta Model

Gennaro Costagliola, Vittorio Fuccella, Fernando A. Pascuccio  
Dipartimento di Informatica, University of Salerno  
Via Giovanni Paolo II, 84084 Fisciano (SA), Italy  
{gencos, vfuccella, fpascuccio}@unisa.it

## Abstract

*New Trust, Reputation and Recommendation (TRR) models are continuously proposed. However, the existing models lack shared bases and goals. For this reason, in this work we define an innovative meta model to facilitate the definition and standardization of a generic TRR model. Following the meta model, researchers in the field will be able to define standard models, compare them with other models and reuse parts of them. A standardization is also needed to determine which properties should be present in a TRR model.*

*In accordance with the objectives we were seeking, following our meta model, we have: defined a pre-standardized TRR model for e-commerce; identified the fundamental concepts and the main features that contribute to form trust and reputation in that domain; respected the dependence on the context/role of trust and reputation; aggregated only homogeneous trust information; listed and shown how to defend from the main malicious attacks.*

*Trust, Reputation and Recommendation Model; meta model; malicious attacks; context/role sensitivity; main features of trust; homogeneous trust information*

## 1 Introduction

Lately, it has become of paramount importance to obtain information about Trust and Reputation of online service providers as well as of other users. In practice, there is the need for a support to make relatively better *trust-based choices*. Of course, as we are in an area where subjectivity plays a predominant role, the optimal point actually does not exist and the best choice is not easy to spot. In recent years there have been numerous studies aimed at understanding how to manage online trust and reputation. Nevertheless, in our opinion, all of these studies have not gone in the same direction. In fact, according to [1, 2, 3, 4], we recognize the lack of shared bases and goals. Authors in [5]

also recognize the lack of a unified research direction and note that there are no unified objectives for trust technologies and no unified performance metrics and benchmarks.

In fact, there are many models in the literature that treat trust and/or reputation contradictorily. For instance, some models use calculation methods based on the transitivity of trust while some authors demonstrate that trust is not transitive but propagative [6]. Other models calculate trust/reputation without taking into account properties deemed essential by some authors (e.g., context-specific, event-sensitive, etc.) [6, 7].

Lastly, differently from other areas of computer science, there is not a well-defined set of testbeds for comparing models [2]. Validations are not performed through a comparison of the results with other models because often they are neither reproducible nor comparable [8]. Almost always the data are not shared and therefore validations use different data even in the same application domain [9]. It rises from the above reasons the urgency of reaching a standard trust and reputation model.

In this paper we lay the foundation for the formulation of a meta model to be shared with researchers in the field, defining properties, characteristics, methods and best practices to which Trust, Reputation and Recommendation (TRR) models should be compliant. We draw inspiration from similar proposals in the literature [3, 4, 5, 10]. Our meta model is also the result of a critical review in which we have recognized strengths and weaknesses of the most important existing TRR models [7, 11, 12].

However, differently from the above cited works, we define a meta model with real requirements for the definition of TRR models. The main purpose of the meta model is to facilitate the definition of a generic TRR model. In fact, the meta model explains how to create, step-by-step, a compliant model. Among others, a standardization is needed to determine the fundamental properties which must be present in a TRR model, thus avoiding that the models do not take them into account. Designing a TRR model in a standard manner will also facilitate the reuse of some of its parts.

Another goal is to introduce a pre-standardized TRR

model for e-commerce. Obviously our model does not claim to be final, since the intention is to propose a basis on which researchers will be able to discuss and, “speaking the same language”, establish a common objective and select the best proposals [7].

The paper is organized as follows: we firstly describe related work in Section 2. Then, in Section 3, we present our meta model. In Section 4, following our meta model, we introduce a pre-standardized TRR model for e-commerce. Finally, in Section 5 we draw some conclusions and outline future work.

## 2 Related Work

Several papers [6, 7, 13] review the most important TRR models. Conversely, to the best of our knowledge, only a few propose meta models to facilitate the definition of standard models. Many authors, among which [1, 2, 3, 4, 5, 14], emphasize the lack of common understanding and shared description in trust models.

Authors in [3] describe an interesting pre-standardized approach for trust and/or reputation models for distributed and heterogeneous systems. They also survey several representative trust and reputation models, describing their main characteristics, with the objective of extracting some common features from them in order to obtain a set of recommendations for a pre-standardized process. In their view, a generic model should consist of the following five components: *gathering behavioral information*; *scoring and ranking entities*; *entity selection*; *transaction*; *rewarding and punishing entities*.

Authors in [14] deal with the *federated trust management*. Trust management in federated environments, as in service-oriented architecture (SOA), will introduce additional complexity. In these environments, it is necessary that different trust management systems can interoperate. Complexity increases because, as many authors complain, there is no consensus on what constitutes the trust. There is the need for a way of representing trust that may be understood by all parties involved. Authors also stress the need for a shared understanding and they identify important aspects of trust frameworks. In order to systematically study the requirements rising from federated trust management, they classify these problems into five aspects: *trust representation*; *trust exchange*; *trust establishment*; *trust enforcement*; *trust storage*. Then they propose a conceptual architecture for federated trust management.

An approach for building a generic trust model, called *UniTEC*, is also described in [5]. Authors identify the following dimensions of the trust relationship: *trust measure*; *trust certainty*; *trust context*; *trust directness*; *trust dynamics*. Then, they map these concepts on the components of their generic trust model. With this approach, built on

the observation, the outcome of each trust model can be mapped onto *UniTEC* and it is also possible to compare models with each other. However, during mapping to the generic trust model details of the trust model are lost [4].

Authors in [4] created a generally applicable meta model, called *TrustFraMM*, which aims at creating the common ground for future trust research in computer science. As authors declare, their meta model was born from the idea of identifying identical functionalities in different available trust frameworks. Using their meta model any trust framework can be described as a set of standard elements of the *TrustFraMM*. The authors expect to get several common implementations so that it will be possible to apply Model Driven Architecture to trust management. This way, it will be easier for researchers and developers to find new solutions also in domains that have not yet been explored. The proposed meta model is only at its first version. The authors plan to further detail the identified elements taking into account the proposals of the other researchers.

In [1] the *TrustFraMM* meta model is extended to be used in the design process. The authors describe a systematic approach for the design of trust frameworks. The basic idea is that in trust framework design there are typical aspects that restrict the possible solutions. For this reason, the authors believe that, by using tested and approved procedures, the design of a trust framework is an exploratory process. Therefore, a designer can select the elements of *TrustFraMM* suitable for his/her specific implementation.

An investigation of trust-based protocols in mobile ad-hoc networks is reported in [15]. The authors also provide a set of properties and essential concepts that should necessarily be considered by trust framework designers in these environments. In addition, methods for the management of trust evidences are categorized. Although some concepts are only briefly exposed and not explained in detail [4], the work provides some important insights on trust management.

As remarked in [16] and [17], the existing works do not well address how to request and obtain recommendations and how to manage attacks and protection mechanisms. Our meta model, besides identifying some crucial aspects in the building of trust/reputation, addresses researchers on how to “think of” and define a standard TRR model. The meta model “forces” to deal with some fundamental aspects which are often neglected in many of the proposed TRR models.

Lastly, our TRR model for e-commerce provides both an application of the proposed meta model and many starting points about trust and reputation management.

### 3 Trust, Reputation and Recommendation Meta Model

In this section, based on observations and literature, we define a TRR meta model whose objective is to facilitate the definition of TRR models and the identification of standard models for any particular context.

Our meta model has been divided into two parts, which list a series of information that shall be provided to build a TRR model compliant with our TRR meta model. This information will also be useful to make the various TRR models more understandable and classifiable. Part 1 is preliminary for the second and covers the fundamental principles and the basic information on which the TRR model is based. Part 2, instead, specifies the way in which information is handled, i.e. it contains the actual definition of the model.

As mentioned in Section 1, in this section we will only list the requirements of the meta model, and will motivate and explain in more detail the choices made in the next section, where we will apply the meta model to define our TRR model for e-commerce.

#### 3.1 TRR Meta Model Part 1 - Basic principles

1. *Scope*. The community referred to by the TRR model (e.g.: social networks, company, e-commerce, etc.). This information is useful to classify the model;

2. *Goals*. The objectives of the TRR model. This information is necessary to let researchers and developers properly use the model.

3. *Fundamental Concepts*. Basic concepts underlying the TRR model justifying the choices made. This point should be carefully described, as it is very important to have solid and commonly accepted basic principles on trust and reputation otherwise the proposed solutions lack any foundation.

4. *Contexts/Roles Ontology*. All the contexts/roles (e.g. “e-commerce/buyer”, “e-commerce/seller”) in which the TRR model can be used [4, 5, 12]. Although it may be challenging to identify the most suitable contexts/roles ontology, it is still essential to have a shared basis on which to think and discuss. Without a shared ontology of contexts it will be complex to reuse metrics or parts of models defined by other authors. Therefore, following the fundamental concepts at the previous point, the TRR model shall specify:

- a) *Context/Role Main Features (MF)*. The main features that contribute to build the trust (reputation) for each context/role (e.g. “quality”, “reliability”) [6, 7].
- b) *Main Features Values (MFV)*. Value domain which can be assigned to each MF (e.g. “good”, “poor”, [0,1], etc.).
- c) *Main Features Measurement*. The metrics for each MF. They shall respect two conditions:

- i. Negative ratings decrease the value of the MF;
- ii. Positive ratings increase the value of the MF;
- d) *Trust/Reputation Measurement*. The specific metric for trust/reputation for each context/role (it is possible that some contexts/roles share the same metric);
5. *Malicious Attacks*. List of all attacks that can undermine the TRR model [16, 18]. A detailed understanding of the threats that may make the model unreliable helps to define a more effective TRR model. Many models are defined without taking into account the malicious attacks. This point shall be constantly updated as new attacks are periodically identified;

#### 3.2 TRR Meta Model Part 2 - Information Management

1. *Entity management*. The entities shall be distinguishable from each other within the system, otherwise it would be impossible to assign a value to their trust/reputation. For this reason it is necessary that the TRR model defines how the involved entities are managed. In the management of the entities (users, services providers, etc.) the following aspects shall be kept in mind:

- a) *Longevity*. After each interaction with an entity there should be the possibility of having other interactions with it in the future [7]. In practice, it shall be impossible or difficult and above all not convenient to change identifier.
- b) *Privacy*. The protection of privacy has become a crucial aspect and more and more users require that their private data are protected. An unclear privacy policy may discourage the participation of honest users. It should be specified whether and how the privacy of users is protected;
- c) *Anonymity*. It should be specified whether and how the users have the option to be anonymous;
- d) *Initial values*. It shall be specified which initial values are assigned to the trust/reputation of the new entities (newcomers). The assignment of an incorrect or not consistent initial value to newcomers could affect the effectiveness of the system. This is almost always ignored in many models;
2. *Information Gathering*. It shall be specified in which way the values of the main features are gathered for each context/role [3, 4]. Therefore, the TRR model shall specify:
  - a) which *passive* mechanisms (without user intervention) for information gathering are used;
  - b) which *active* mechanisms (the user gives explicit feedback) for information gathering are used;
  - c) how the authenticity of the information is preserved (*Authenticity*).

In any case, the values collected or assigned by the user must be consistent with those reported in Part 1.

3. *Information Storing*. Besides indicating how the information on the MFs is stored, the TRR model shall specify:
  - a) whether one party of a transaction (agent/resource) cannot deny having received/expressed a rating (*Non-Repudiation*);
  - b) whether and how the information is aggregated (*Aggregation*);
  - c) whether and how the oldest information is taken into account (*History*);
  - d) whether and how the oldest information is less influential than the most recent (*Aging*);
4. *Information Sharing*. All users shall have access to the same information (*Democracy*). The TRR model may specify:
  - a) whether all the necessary knowledge (rules, procedures, etc.) to interpret and manage the information shared by the system is made available (*Knowledge*);
  - b) whether the information is easily understood (*Clarity*);
  - c) whether the system is easy to use (*Usability*);
  - d) whether it is possible to trace or contact the raters (*Untraceability*);
5. *Recommendation System*. A mechanism to “advise” the user which entities to interact with in a given situation (context/role). Some systems may only collect and share information about the entities. If the TRR model includes a recommendation system, it shall describe in detail the “decision-making process” used;
6. *Incentive Mechanism*. After a transaction, the users usually have no incentive to give a rating about the other party. To be successful, a TRR model should encourage the participation of honest users and discourage dishonest behavior [7]. Therefore, being this a fundamental aspect, the TRR model should describe in detail the *incentive mechanism* used [19];
7. *Malicious Attacks Resilience*. In general there can not be a system completely immune from attacks by malicious users. Nevertheless, it is possible to make malicious behaviors inconvenient. The TRR model should indicate whether the system is resilient to the attacks listed in the first phase and if there are weaknesses.

## 4 A Pre-Standardized Trust and Reputation Model for E-Commerce

In this section, we propose a pre-standardized TRR model for e-commerce whose definition has been carried out following our TRR meta model. Here we outline the features that, in our view, must necessarily be part of a standard model. The purpose of this section is to provide an example of application of our meta model, which is helpful for creating models for various contexts (e.g. *product review*; *expert sites*; *autonomous system*; *wireless sensor networks*; *mobile ad-hoc networks*; *mobile agent system*; etc.).

### 4.1 TRR Model for E-Commerce Part 1 - Basic principles

1. *Scope*. The model is applicable in the context of selling products online (e-commerce).
2. *Goals*. Providing an entity with reliable information on the conduct of the other party in the context/role in which they will interact.
3. *Fundamental Concepts*. We adopt the definition of *trust* based on the former encounters between two agents: “*Trust is a subjective expectation an agent has about another’s future behavior based on the history of their encounters*” [20]. Here an “*encounter is an event between two agents within a specific context*”.

Furthermore, we adopt the following definition of Reputation: “*what is generally said or believed about an entity’s character or standing*” (*word of mouth*) [7].

Reputation clearly has a global aspect whereas trust is viewed from a local and subjective angle.

In the light of the above arguments, a recommendation system, besides relying on reputation, should allow an entity to consider any private information in its possession (direct trust). For example, an entity may build its own *trust-chain*, i.e. a periodically updated local store which includes trust data of the other entities of which the entity has direct knowledge.

As regards trust and reputation, in agreement with [6, 7, 21], we consider of primary importance one aspect that is often overlooked by many authors: *Trusting Ann as a doctor is not the same as trusting her as an aircraft pilot*.

This is also true for the reputation and suggests that trust and reputation are dependent on the context and the role (*Context/Role-Sensitive*). Therefore there can be no single evaluation method for trust and/or reputation that is applicable in all contexts/roles and thus it is essential to identify a *Contexts/Roles Ontology*.

In addition, each context/role has its own peculiar characteristics (main features) that contribute to a greater extent to build trust (*subjective expectation*) or reputation (*word-of-mouth*) of an entity. For instance, a high level of trust (reputation) in an online store might arise from its positive ratings received on *product quality*, *assortment*, etc.

Another major point is the way in which inhomogeneous values are commonly aggregated. In fact, the aggregation of all the features values in a single trust (reputation) score causes a significant loss of information, and produces an unreliable result. For example, we believe that aggregating the ratings regarding *product quality* with those regarding *assortment* has neither a logic nor a theoretical justification. For this reason, in our model we maintain separate values for each main feature.

In this section we cannot ignore a series of other fundamental properties of trust and/or reputation [6, 7]:

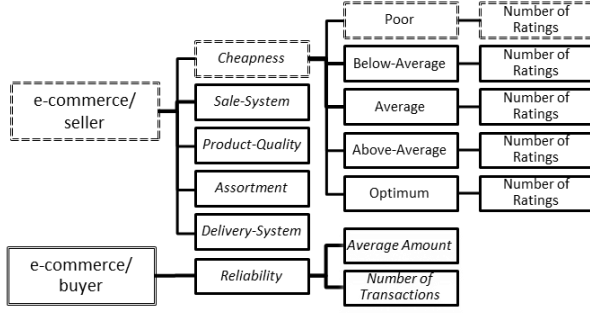


Figure 1: Categorization of the information according to: context/role, main feature and main feature values.

(a) *Subjective*: subjective nature of ratings leads to personalization of trust/reputation evaluation.

(b) *Relational*: as two members interact with each other frequently, their relationship strengthens, and trust (reputation) will increase if the experience is positive and decrease otherwise.

(c) *Dynamic*: trust and reputation decay with time, hence new experiences are more important than old ones.

(d) *Propagative*: if Ann knows Bob who knows Clair, and Ann does not know Clair, then Ann can derive some amount of trust on Clair based on how much she trusts Bob and how much Bob trusts Clair.

(e) *Non-Transitive*: if Ann trusts Bob and Bob trusts Clair, this does not imply that Ann trusts Clair. Propagation does not imply transitivity.

(f) *Asymmetric*: Ann trusts Bob does not imply that Bob trusts Ann.

(g) *Slow*: high trust and good reputation take a long time to build, i.e., they grow slowly.

(h) *Event sensitive*: a single high-impact event may destroy trust (reputation) completely.

(i) *Indirect Trust*: trust can be based on second-hand information about an entity that one does not know directly.

(j) *Direct Trust*: first-hand information should always be the most reliable.

4. *Contexts/Roles Ontology*. In the context of e-commerce we identify the following contexts/roles: *e-commerce/seller* and *e-commerce/buyer*.

a) *Context/Role Main Features (MF)*. We identified the following main features (see Fig.1):

- e-commerce/seller: *Cheapness*, *Assortment*, *Delivery System*, *Sale System* and *Product Quality*;
- e-commerce/buyer: *Reliability*;

b) *Main Features Values (MFV)*. The following values can be assigned to all main features in the e-commerce/seller context: *poor (P)*, *below average (BA)*, *average (A)*, *above average (AA)* and *optimum (O)*. The amount of the transaction (values in **R**) is assigned to the *Reliability* main feature in the e-commerce/buyer context.

c) *Main Features Measurement*. For each MFV in e-commerce/seller, the Number of Ratings (NoR) received is stored. Since each rating will increase the NoR of the MFV, the following properties are valid:

- i. Every negative rating decreases the value of the feature. In fact, it increases the NoR associated with the corresponding negative value;
- ii. Every positive rating increases the value of the feature. In fact, it increases the NoR associated with the corresponding positive value.

For each MFV in the e-commerce/buyer, two values are kept: number of transactions (necessary to recalculate the average) and the average amount spent (the actual measure of the Reliability).

- i. Transactions with a spent amount lower than average decrease the value of the feature;
- ii. Transactions with a spent amount higher than average increase the value of the feature;

d) *Trust/Reputation Measurement*. As shown in the following, the model does not deal with the computation of a value for trust/reputation. It only collects information on the past behavior of the entities, that are made available for subjective evaluations.

5. *Malicious Attacks*. The most common attacks are the following (detailed definitions are in [16, 18, 22]):

- a) *Whitewash*: a user with a poor reputation, obtains a new identity to erase his/her previous reputation;
- b) *Sybil Attacks*: a dishonest user attempts to obtain multiple identities to cheat the reputation system;
- c) *Traitor*: a user with low reputation behaves well until s/he reaches a good reputation. Then s/he resumes his/her dishonest behavior;
- d) *Fake transaction*: a user creates ad-hoc transactions only to express a rating. Although transaction costs may be required, it could be equally convenient to bear the costs in relation to the benefits achieved;
- e) *Slander*: a user acquires many identities (sybil attack) and then provides negative feedback to decrease the reputation of the victims;
- f) *Promote*: a user acquires many identities (sybil attack) and provides positive feedback to increase the reputation of a target;
- g) *Slander+Promote*: both Slander and Promote are exploited at the same time;
- h) *Self-promote*: a user gives positive feedback on subjects in whose s/he is not interested to increase his/her own reputation, and then provides dishonest feedback on the victims;
- i) *Oscillation*: some users acquire many identities (sybil attack) which are divided into two groups with different roles. A group focuses in giving dishonest feedback on target, while the other focuses in increasing its reputation by providing honest feedbacks. The roles of the

two groups dynamically alternate;

- j) *Ballot stuffing*: a user gives many ratings so as to affect the reputation of the target;
- k) *RepTrap*: some users acquire many identities (sybil attack) forming a coordinated group to become “majority”. The group gives many negative ratings on the targets (users with a few feedbacks, called “traps”) to exceed the majority of feedbacks. In this way, the system will judge the negative feedback expressed by dishonest users as “consistent” with the reputation of the target and, on the other hand, as “inconsistent” the feedback provided by honest users.  
The final effect will be that the honest users will have their reputation lowered and the dishonest users will have it increased.  
Many traps will be disseminated in the system. The side effect is also to reduce the total number of honest users with high reputation;
- l) *Denial of Service*: a DOS is caused to avoid the calculation and dissemination of reputation;
- m) *Exit*: a user who has decided to leave the system is no longer interested in his own reputation and can behave dishonestly without worrying about the consequences;
- n) *Context/Role Sliding*: a user attempts to gain a good reputation in the contexts/roles where it is easier and cheaper to get it and then have malicious behaviors in other contexts/roles exploiting the high reputation gained;
- o) *Orchestrated*: a user organizes attacks using various of the strategies listed above;

## 4.2 TRR Model for E-Commerce Part 2 - Information Management

1. *Entity management*. This model works with any Identity Management. Necessary and sufficient condition is that the entities are *distinguishable* from each other.

- a) *Longevity*. Even though the entities can change their identifier, it is not advantageous: the newcomers are recognized because they have a few ratings.
- b) *Privacy*. Neither data on users nor data on content of transactions are stored, thus privacy is preserved. Furthermore, the average value of purchases it is not a “sensitive” information (see next point 3);
- c) *Anonymity*. The users are uniquely identified by an identifier and therefore remain anonymous;
- d) *Initial values*. Newcomers have no initial assigned value;

2. *Information Gathering*. Following a transaction, the buyer (seller) can express his own evaluation of the seller (buyer) (*relational property*). Users, depending on the context/role, shall express a rating on the specific MFs identified in Part 1.

- a) The ratings expressed by the seller against the buyer (the amount of a successful transaction) could be automated and considered a passive mechanism for information gathering;
- b) The ratings expressed by the buyer against the seller (explicit feedbacks) are an active mechanism for information gathering;
- c) Information authenticity can be preserved using SAML assertions;

3. *Information Storing*. As mentioned in Part 1, a rating produces a list of values, one for each MF in the context/role. In general, with respect to an entity and to the context/role under which the interaction took place, the NoR received for each MFV is stored. These values are categorized according to: context/role, MF and MFV (see Fig.1).

- a) *Non-Repudiation*: since ratings can be only given following a transaction, neither party can repudiate the ratings concerning it;
- b) *Aggregation*. In the *e-commerce/seller* context/role, as a result of a rating, the NoR of the corresponding MFV is increased by 1. For instance, if a buyer assigns *poor* to the *Cheapness* of an online store, the NoR of the MFV *poor* related to the MF *Cheapness* is increased by 1 (see the dashed boxes in Fig.1). In this way we avoid to aggregate inhomogeneous information. In the *e-commerce/buyer* context/role, following a rating the value of the MF *Reliability* is updated by increasing by 1 the total number of transactions and recalculating the average amount.
- c) *History*. The history can be managed by storing the time in which a rating is expressed. As time passes, the past ratings may need to be aggregated. In this case, in the *e-commerce/seller* context/role, the aggregation of the values of the MF is obtained by adding the NoR of the corresponding MFV.

In the *e-commerce/buyer* context/role, the history is managed similarly with the only difference that the aggregation of the values of the *reliability* of the buyer is obtained by adding the total number of transactions and then re-calculating the average;

- d) *Aging*. The oldest information should be given less importance than the most recent. Who uses it should decide how to weight it (*dynamic and subjective properties*);
4. *Information Sharing*. All users, without distinction, can access to the same information (*Democracy*). Only the average amount spent, used to measure the *Reliability* of the buyer, is published, while the number of transactions is kept confidential. Furthermore:

- a) *Knowledge*. To interpret and manage the information shared by the system it is sufficient to know how the data are stored (Fig.1);
- b) *Clarity*. The information is evidently easy to understand;
- c) *Usability*. The user is required to make an assessment



on the interaction by simply choosing values from lists;  
 d) *Untraceability*. The ratings are stored as described in point 3. Therefore it is not possible to identify the raters;  
 5. *Recommendation System*. This model is only concerned with gathering, aggregating, storing and sharing ratings without including a Recommendation System. At the same time, however, the ratings expressed on the past behavior of users (*indirect trust*) are made available divided into two contexts/roles. This enables an easy build of a custom Recommendation System (*subjective property*) effective to identify the right entities to interact with in a given situation (*propagative property*).

Moreover, as already mentioned, the user's private information may also be managed by means of the trust-chains (*direct trust*). We believe that the *Recommendation System* in its decision-making process should ensure that:

- the recent *positive ratings* should have low impact on reputation (*reputation lag*);
- even a single recent *negative rating* has a strong impact on reputation (*event sensitive*);

6. *Incentive Mechanism*. We adopt the incentive mechanisms based on financial rewards presented in [23]. In addition, dishonest users are discouraged, as shown in the following.

7. *Malicious Attacks Resilience*. The aim of our model is to make the user aware of the history of the counterpart: it will be the user, in a subjective way, to decide how much trust to place in it from time to time.

Particular attention must be paid to the way in which the reliability of the buyer is handled. Firstly, we highlight that the transactions always have a cost: that of the purchased good/service. Therefore, a buyer who wants to increase his reputation maliciously should carry out transactions spending a lot of money. This is a pretty strong disincentive that protects the system from many attacks. Another technique might perform a behavioral analysis to identify users who make many low value transactions.

In line with some consolidated solutions [18, 24, 25], we can deduce that the effectiveness of the following attacks is reduced for the following reasons:

- (a) *Whitewash* and *Sybil Attacks*: newcomers are not assigned any initial value, therefore who interacts with a newcomer will be aware of this and will be able to take all the necessary precautions.
- (b) *Traitor*: the number of ratings with negative values will be known to the users, thus the objective of the attacker will be nullified.
- (c) *Fake transaction*, *Promote*, *Ballot stuffing* and *Exit*: the danger of attack is severely curtailed from the application of mechanisms to discourage dishonest behavior. Furthermore a rating can only be expressed only as a result of a transaction, whose cost discourages malicious user to make fake transactions.

(d) *Slander*: if an entity receives a lot of negative ratings from newcomers it would understand it is under attack and would avoid future interactions with them. In addition, the same mechanism of Promote is also effective in this case.

(e) *Slander+Promote*, *Self-promote*, *Oscillation* and *Rep-Trap*: being curtailed the danger of Slander and of Promote, a simultaneous attack is not very effective as well.

(f) *DOS*: the system for the management of the ratings must provide mechanisms (e.g. the use of message queues and/or of a decentralized system) to handle the sudden increase in requests.

(g) *Context/Role Sliding*: this threat is avoided by maintaining separate transactions for contexts/roles;

(h) *Orchestrated*: by limiting all of the above threats, this threat is limited too.

## 5 Conclusions and Future Research

In this paper we firstly defined a new meta model which simplifies and standardizes the definition of a generic TRR model. Based on this, we have also proposed a pre-standardized TRR model for e-commerce and listed some of the fundamental properties that must necessarily be taken into account in the construction of a TRR model. We identified the peculiar characteristics (main features) that contribute to a greater extent to form the trust in an entity in the context of e-commerce. The main malicious attacks against a Trust and Reputation system were listed, and it was demonstrated the resistance of our model to them. Moreover, the collection and sharing of information is decoupled from that of the calculation of the trust/reputation. In this way, users can exploit different systems to derive the trust/reputation of an entity.

In the future we will continue to work on the meta model and the TRR model for e-commerce in order to achieve a standardization of both. In this regard, the meta model has been designed to be open to the contribution of the scientific community. We are also trying to establish a recommendation system for e-commerce that making use of information provided from our TRR model can help the user to make trust-based choices. In addition, more efforts are needed to define other standard TRR models for other contexts. Lastly, we are working on the implementation of an environment for simulating and validating the proposed TRR models.

## References

- [1] M. Vinkovits and A. Zimmermann. *Defining a Trust Framework Design Process*, volume 8058 of *Lecture Notes in Computer Science*. 2013.
- [2] K. K. Fullam, T. B. Klos, G. Muller, J. Sabater,

- A. Schlosser, Z. Topol, K. S. Barber, J. S. Rosenschein, L. Vercouter, and M. Voss. A specification of the agent reputation and trust (ART) testbed: experimentation and competition for trust in agent societies. In *Proceedings of AAMAS 2005*, page 512–518.
- [3] F.G. Mármol and G.M. Pérez. Towards pre-standardization of trust and reputation models for distributed and heterogeneous systems. *Computer Standards & Interfaces*, 32(4):185–196, 2010.
- [4] M. Vinkovits and A. Zimmermann. TrustFraMM: Meta Description for Trust Frameworks. In *Proceedings of PASSAT 2012*, pages 772–778.
- [5] M. Kinateder, E. Baschny, and K. Rothermel. Towards a Generic Trust Model – Comparison of Various Trust Update Algorithms. In *Proceedings of iTrust 2005*, number 2477, pages 177–192.
- [6] W. Sherchan, S. Nepal, and C. Paris. A Survey of Trust in Social Networks. *ACM CSUR 2013*, 45(4):47:1–47:33.
- [7] A. Jøsang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.
- [8] R. Kerr and R. Cohen. TREET: The trust and reputation experimentation and evaluation testbed. *Electronic Commerce Research*, 10(3-4):271–290, 2010.
- [9] S. Magin and S. Hauke. Towards engineering trust systems: Template-based, component-oriented assembly. In *Proceedings of PST 2013*, pages 348–351.
- [10] F.G. Mármol and G.M. Pérez. Trust and reputation models comparison. *Internet Research*, 21(2):138–153, 2011.
- [11] J. Sabater and C. Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, 2005.
- [12] D. Artz and Y. Gil. A survey of trust in computer science and the Semantic Web. *Web Semantics*, 5(2):58–71, 2007.
- [13] I. Pinyol and J. Sabater-Mir. Computational trust and reputation models for open multi-agent systems: a review. *Artificial Intelligence Review*, 40(1):1–25, 2013.
- [14] Z. Wu and A. C. Weaver. Requirements of federated trust management for service-oriented architectures. In *Proceedings of PST 2006*, page 10.
- [15] K. S. Ramana and A. A. Chari. A survey on trust management for mobile ad hoc networks. *International Journal of Network Security & Its Applications*, 2(2):75–85, 2010.
- [16] Y. Sun and Y. Liu. Security of online reputation systems: The evolution of attacks and defenses. *IEEE Signal Processing Magazine*, 29(2):87–97, 2012.
- [17] Y.L. Sun, Z. Han, W. Yu, and K.J.R. Liu. A Trust Evaluation Framework in Distributed Networks: Vulnerability Analysis and Defense Against Attacks. In *INFOCOM*, pages 1–13, 2006.
- [18] K. Hoffman, D. Zage, and C. Nita-Rotaru. A Survey of Attack and Defense Techniques for Reputation Systems. *ACM CSUR 2009*, 42(1):1–31.
- [19] R. Jurca and B. Faltings. An incentive compatible reputation mechanism. In *Proceedings of AAMAS 2003*, pages 1026–1027.
- [20] L. Mui, M. Mohtashemi, and A. Halberstadt. A Computational Model of Trust and Reputation for E-businesses. In *Proceedings of HICSS 2002-Vol.7*, pages 188–.
- [21] I. Pinyol, J. Sabater-Mir, P. Dellunde, and M. Paolucci. Reputation-based decisions for logic-based cognitive agents. In *Proceedings of AAMAS 2012*, volume 24, pages 175–216.
- [22] D. Fraga, Z. Bankovic, and J.M. Moya. A Taxonomy of Trust and Reputation System Attacks. In *Proceedings of TrustCom 2012*, pages 41–50.
- [23] N. Miller, P. Resnick, and R. Zeckhauser. Eliciting honest feedback in electronic markets. In *Proceedings of SITE 2002*.
- [24] Y. Sun, Z. Han, and K.J.R. Liu. Defense of trust management vulnerabilities in distributed networks. *Communications Magazine, IEEE*, 46(2):112–119, 2008.
- [25] F.G. Mármol and G.M. Pérez. Security threats scenarios in trust and reputation models for distributed systems. *Computers & Security*, 28(7):545–556, 2009.

# Proceedings for Visual Languages and Computing



# Spatial Relations Between 3D Objects: The Association Between Natural Language, Topology, and Metrics

Jennifer L. Leopold Chaman L. Sabharwal Katrina J. Ward  
leopoldj@mst.edu chaman@mst.edu kjw26b@mst.edu  
Computer Science Department  
Missouri University of Science & Technology  
Rolla, MO 65409 USA

**Abstract**—With the proliferation of 3D image data comes the need for advances in automated spatial reasoning. One specific challenge is the need for a practical mapping between spatial reasoning and human cognition, where human cognition is expressed through natural-language terminology. With respect to human understanding, researchers have found that errors about spatial relations typically tend to be metric rather than topological; that is, errors tend to be made with respect to quantitative differences in spatial features. However, topology alone has been found to be insufficient for conveying spatial knowledge in natural-language communication. Based on previous work that has been done to define metrics for two lines and a line and a 2D region in order to facilitate a mapping to natural-language terminology, herein we define metrics appropriate for 3D regions. These metrics extend the notions of previously defined terms such as *splitting*, *closeness*, and *approximate alongness*. The association between this collection of metrics, 3D connectivity relations, and several English-language spatial terms was tested in a human subject study. As spatial queries tend to be in natural language, this study provides preliminary insight into how 3D topological relations and metrics correlate in distinguishing natural-language terms.

**Keywords**—*Spatial Reasoning, Natural-Language Processing, Region Connection Calculus, 3D Images*

## I. INTRODUCTION

In tandem with increases in pervasive mobile computing and the proliferation of 3D image data comes the need for advances in automated spatial reasoning. One of the particular challenges is the need for a practical mapping between qualitative and quantitative spatial reasoning and human cognition, the latter being expressed principally through natural-language terminology. With respect to human understanding, errors about spatial relations typically tend to be metric rather than topological [1, 2]; however, topology alone has been found to be insufficient for conveying spatial knowledge in natural-language communication [3, 4]. The consensus is that topology matters while metrics refine [5]. To accommodate natural-language spatial queries, an effective interface between automated spatial reasoning and natural language requires an appropriate blend of natural language, topology, and metrics.

Based on the work that has been done to define metrics for two lines [4] and a line and a 2D region [3] with topological

relations in order to facilitate a mapping to natural-language terminology, herein we define metrics appropriate for two 3D regions and the topological connectivity relations used in VRCC-3D+ [6, 7, 8]. These metrics extend the notions of what previous authors [3, 4, 9] have referred to as: *splitting* (i.e., how much is in common between two objects), *closeness* (i.e., how far apart parts are), and *approximate alongness* (i.e., a combination of splitting and closeness). The association between this collection of metrics, 3D connectivity relations, and several English-language spatial terms was tested in a human subject study. The results of that study provide preliminary insight into how the 3D topological relations and metrics correlate in distinguishing natural-language terms.

The paper is organized as follows. Section II briefly discusses the region connection calculus, VRCC-3D+, and the topological relations pertinent to this study. Section III defines the metric relations for splitting, closeness, and approximate alongness, which are similar in concept to those that have been proposed for a line and a 2D region [3], but are significantly redefined to be appropriate for objects in 3D space. Section IV identifies dependencies between the topological relations and the metrics, as well as intra-relationships within the metrics. Section V examines associations between the topological relations, the metrics, and various natural-language terms based on the results of a human subject experiment. Section VI outlines directions for future work, followed by a summary and conclusions in Section VII.

## II. TOPOLOGICAL RELATIONS

### A. Mathematical Preliminaries

$R^3$  denotes the three-dimensional space endowed with a distance metric. Here the mathematical notions of *subset*, *proper subset*, *equal sets*, *empty set* ( $\emptyset$ ), *union*, *intersection*, *universal complement*, and *relative complement* are the same as those typically defined in set theory. The notions of *neighborhood*, *open set*, *closed set*, *limit point*, *boundary*, *interior*, *exterior*, and *closure* of sets are as in point-set topology. The interior, boundary, and exterior of any region are disjoint, and their union is the universe.

A set is *connected* if it cannot be represented as the union of disjoint open sets. For any non-empty bounded set  $A$ , we use symbols  $A^c$ ,  $A^i$ ,  $A^b$ , and  $A^e$  to represent the universal complement, interior, boundary, and exterior of a set  $A$ ,

respectively. Two regions A and B are equal if  $A^i \equiv B^i$ ,  $A^b \equiv B^b$ , and  $A^e \equiv B^e$  are true. For our discussion, we assume that every region A is a non-empty, bounded, regular closed, connected set without holes; specifically,  $A^b$  is a closed curve in 2D, and a closed surface in 3D.

### B. Region Connection Calculi

Much of the foundational research on qualitative spatial reasoning is based on a region connection calculus (RCC) that describes 2D regions (i.e., topological space) by their possible relations to each other. Most notable is the RCC8 model [10] which defines the following eight relations (illustrated in Fig. 1): disconnected (DC), externally connected (EC), partial overlap (PO), equality (EQ), tangential proper part (TPP), non-tangential proper part (NTPP), converse tangential proper part (TPPc), and converse non-tangential proper part (NTPPc). Topological relations in a region connection calculus are typically defined using first-order logic (as in the work of Randell, Cui, and Cohn [10]) or using the 9-Intersection model [11] which looks at whether the intersections between the interiors, exteriors, and boundaries of two regions are empty or non-empty.

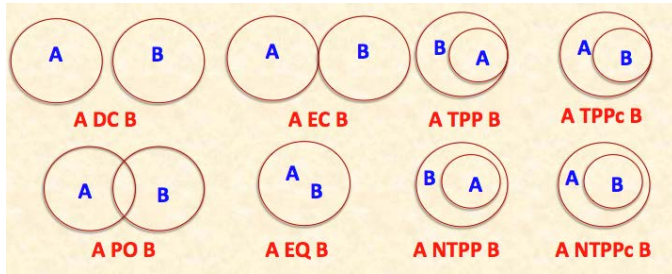


Figure 1. RCC8 relations.

Whereas a 2D object is in a plane, a 3D object is in space. The simplest examples of 3D objects are a pyramid, a cube, a cylinder, and a sphere. A concave pyramid is a complex, simply connected 3D object. Since concave objects can be partitioned into convex objects, for all practical purposes, we work with convex objects. For the rest of this discussion, we will base our analysis on convex objects; in particular, spheres are used in our natural-language human study.

VRCC-3D+ [6, 7, 8] is the implementation of a region connection calculus that qualitatively determines the spatial relations between 3D objects, both in terms of connectivity and obscuration. The VRCC-3D+ connectivity relations are named the same as in RCC8; however, the VRCC-3D+ connectivity relations are calculated in 3D rather than 2D. Fifteen obscuration relations also are defined in VRCC-3D+. Considered from a 2D projection, each VRCC-3D+ obscuration relation is a refinement of basic concepts of no obscuration, partial obscuration, and complete obscuration. A composite VRCC-3D+ relation specifies both a connectivity relation and an obscuration relation. Herein our discussion is limited to the VRCC-3D+ connectivity relations, which heretofore will be referred to as topological relations; application of this work to the VRCC-3D+ obscuration relations is beyond the scope of this paper. For a more in-

depth discussion of VRCC-3D+, including how it compares to the other RCC models, see [6, 7, 8].

### III. METRIC PROPERTIES

Metric relations focus on the quantitative differences in spatial features between the two regions or objects being compared; typically, these relations are expressed as scaled (normalized) volumes, areas, distances, lengths, or size differences. Three metric concepts were introduced in the study that compared a line and a 2D region [3]: (1) splitting, which determines how much is in common between two objects; (2) closeness, which determines how far apart parts of the two objects are; and (3) approximate alongness, which combines splitting and closeness along the boundaries of objects. Metric relations for these concepts were defined in another study [4] to be suitable for comparing two lines. Metrics for additional spatial concepts such as angular direction and overlap have been considered for objects in 2D space [4]. Here we adopt and adapt some of the basic metrics to apply to objects in 3D space. The intersection between the convex volumes can be a volume, a surface, a line segment, or a point. We use normalized metric values, (i.e. dimensionless units) to distinguish between them. For each of our metric relations  $M(A,B)$  for objects A and B there is a converse metric relation denoted  $Mc(A,B)$ , defined by  $Mc(A,B) = M(B,A)$ .

#### A. Splitting

Interior volume splitting (IVS) considers the scaled (normalized) part of one object that is split by the other object. Here boundary does not matter as the volume of the boundary is zero.

$$IVS(A,B) = \frac{\text{volume}(A \cap B)}{\text{volume}(A)}$$

Exterior volume splitting (EVS) describes the proportion of one object's interior that is split by the other object's exterior. Again, boundary does not matter.

$$EVS(A,B) = \frac{\text{volume}(A \cap B^e)}{\text{volume}(A)}$$

Observe that  $\text{volume}(A) = \text{volume}(A \cap B) + \text{volume}(A \cap B^e)$ , hence  $EVS(A,B) = 1 - IVS(A,B)$ .

We define another splitting metric to specifically examine the proportion of the boundary of one object that is split by the boundary of the other object; we denote this metric BS for boundary splitting. It should be noted that there are two versions of the equations for this metric. If the intersection of the objects' boundaries is a line, then the metric should be computed as a scaled length; otherwise, the intersection must be a surface area and the metric should be computed as a normalized area.

$$BS(A,B) = \frac{\text{area}(A^b \cap B^b)}{\text{area}(A^b)} \text{ or } \frac{\text{edgeLength}(A^b \cap B^b)}{\text{edgeLength}(A^b)}$$



### B. Closeness

For the 3D object shown in Fig. 2(a), let  $N_{Er}(A^b)$  be the exterior 3D  $r$ -neighborhood of  $A^b$  of radius  $r>0$  (Fig. 2(b)), and let  $N_{Ir}(A^b)$  be the interior 3D  $r$ -neighborhood of  $A^b$  of radius  $r>0$ ; see Fig. 2(c). The smaller the value of  $r$ , the closer the objects are; thus  $N_{Er}(A^b) \cap B \neq \emptyset$  determines the minimum volume common to the exterior neighborhood of  $A^b$  and  $B$ . Let  $\delta$  be the least upper bound of  $r$  for which  $N_{Er}(A^b) \cap B \neq \emptyset$ . Then  $\delta$  is the closest distance between the exterior neighborhood of  $A^b$  from  $B$ . Let  $\Delta_E A = N_{Er}(A^b)$  be the minimum volume in a pre-specified exterior  $r$ -neighborhood ( $r>\delta$ ) of  $A^b$ . Similarly let  $\Delta_I A = N_{Ir}(A^b)$  be the minimum volume in a pre-specified interior  $r$ -neighborhood ( $r>\delta$ ) of  $A^b$ . The value of  $r$  is specified by the application. In general, for numerical calculations, it is approximately one percent of the sum of the radii of two spheres. Intuitively,  $r$  accounts for the thickness of the boundary for the object.

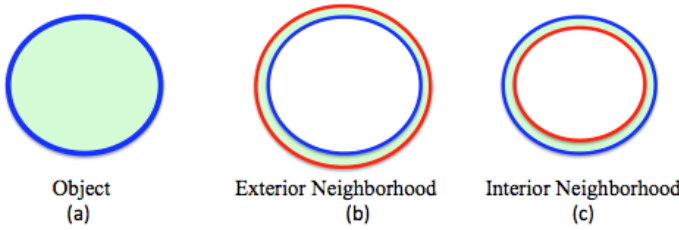


Figure 2. (a) An object, (b) the exterior neighborhood of the boundary of the object, and (c) the interior neighborhood of the boundary of the object.

Considering the interior neighborhood of an object, we define interior volume closeness (IVC) as follows:

$$IVC(A,B) = \frac{\text{volume}(\Delta_I A \cap B)}{\text{volume}(\Delta_I A)}$$

Similarly, we can consider the exterior neighborhood of an object, and can define a metric for exterior volume closeness (EVC) as follows:

$$EVC(A,B) = \frac{\text{volume}(\Delta_E A \cap B)}{\text{volume}(\Delta_E A)}$$

This metric is a measure of how much of the interior neighborhood of  $A^b$  covers  $B$ ; this is the extent to which the interior neighborhood of  $A$  is closer to  $B$ .

### C. Approximate Alongness

Approximate alongness is essentially a combination of splitting and closeness. There are three types of metrics in this case: (1) normalized boundary of an object common to the interior neighborhood of the boundary of the other object; (2) normalized boundary of an object common to the exterior neighborhood of the boundary of the other object; and (3) normalized boundary of an object common to the other entire object. Similar in concept to the inner and outer approximate alongness metrics that were proposed for a line and a 2D region [3], we can define metrics to assess the relative amount of the boundary of one object that is shared with the

neighborhood of the other object. Interior boundary (IB) is the proportion of the boundary of one object along the interior neighborhood of the other object. Exterior boundary (EB) is the proportion of the boundary of one object along the exterior neighborhood of the other object. As was the case for boundary splitting (BS), there are two versions of the equations for each of these metrics. If the intersection is a line, then the metric should be computed as the normalized length; otherwise, the intersection must be a surface area and the metric should be computed as the normalized area.

$$IB(A,B) = \frac{\text{area}(A^b \cap \Delta_I B)}{\text{area}(A^b)} \text{ or } \frac{\text{edgeLength}(A^b \cap \Delta_I B)}{\text{edgeLength}(A^b)}$$

$$EB(A,B) = \frac{\text{area}(A^b \cap \Delta_E B)}{\text{area}(A^b)} \text{ or } \frac{\text{edgeLength}(A^b \cap \Delta_E B)}{\text{edgeLength}(A^b)}$$

We define another alongness metric, which we shall call boundary alongness (BA), to assess how much of the scaled boundary of one object is shared with the entirety of the other object. Here the edge length of intersection is the length of the arc or line segment of intersection. The edge length of the boundary (i.e., the denominator) corresponds to the edge length of the super arc or line segment that contains the intersection edge (i.e., the numerator).

$$BA(A,B) = \frac{\text{area}(A^b \cap B)}{\text{area}(A^b)} \text{ or } \frac{\text{edgeLength}(A^b \cap B)}{\text{edgeLength}(A^b)}$$

## IV. CONSTRAINTS AND DEPENDENCIES AMONG TOPOLOGICAL RELATIONS AND METRICS

For a topological relation, if the value of a metric varies as the two objects vary then there is a dependency between that topological relation and metric. For example, if objects  $A$  and  $B$  are disconnected (i.e.,  $DC(A,B)$ ), then  $A \cap B$  always will be empty, and consequently both  $IVS(A,B)$  and  $IVSc(A,B)$  always will be 0. Hence the values of  $IVS(A,B)$  and  $IVSc(A,B)$  are not dependent on  $A$  and  $B$ ; that is, there is no dependency between the relation  $DC(A,B)$  and metrics  $IVS(A,B)$  and  $IVSc(A,B)$ . In contrast, if objects  $A$  and  $B$  partially overlap (i.e.,  $PO(A,B)$ ), then  $A \cap B$  will be non-empty; the values of both  $IVS(A,B)$  and  $IVSc(A,B)$  will vary depending upon the particular configuration of the partially overlapping objects  $A$  and  $B$ , so we say that there is a dependency between the relation  $PO(A,B)$  and the metric  $IVS(A,B)$  (as well as  $IVSc(A,B)$ ). Table 1 lists these dependencies where a highlighted box indicates that there is a dependency relation between the metric (listed in the top row) and the topological relation (listed in the left column).

When a metric is not dependent upon a topological relation, the value of that metric always will be unchanged; it turns out to always be equal to 0 or always be equal to 1 regardless of the particular objects that are in that topological configuration. However, when a metric is dependent upon a topological relation, we can deduce the most restrictive

Table 1. Dependencies among VRCC-3D+ topological relations and metric relations. Highlighted boxes denote metric relations that are dependent upon a topological relation.

	IVS(A,B)	IVSc(A,B)	EVS(A,B)	EVSc(A,B)	BA(A,B)	BAC(A,B)	BS(A,B)	BSc(A,B)	EVC(A,B)	EVCc(A,B)	IVC(A,B)	IVCc(A,B)	IB(A,B)	IBc(A,B)	EB(A,B)	EBc(A,B)
DC(A,B)																
EC(A,B)																
EQ(A,B)																
PO(A,B)																
TPP(A,B)																
TPPc(A,B)																
NTPP(A,B)																
NTPPc(A,B)																

Table 2. Value constraints among VRCC-3D+ topological relations and metric relations.

	IVS(A,B)	IVSc(A,B)	EVS(A,B)	EVSc(A,B)	BA(A,B)	BAC(A,B)	BS(A,B)	BSc(A,B)	EVC(A,B)	EVCc(A,B)	IVC(A,B)	IVCc(A,B)	IB(A,B)	IBc(A,B)	EB(A,B)	EBc(A,B)
DC(A,B)																
EC(A,B)																
EQ(A,B)																
PO(A,B)																
TPP(A,B)																
TPPc(A,B)																
NTPP(A,B)																
NTPPc(A,B)																
	always 0	always 1	> 0	≥ 0												

constraint on the range of values for that metric; namely, we know whether the metric will produce a value  $> 0$  or  $\geq 0$  regardless of the size and shape of the two objects being compared. For example, if object A is a non-tangential proper part of object B (i.e.,  $NTPP(A,B)$ ), then  $A^b$  intersected with the exterior neighborhood of  $B^b$  is empty, so  $EB(A,B)$  will always be equal to 0, regardless of the particular objects A and B that are in that spatial configuration. However,  $B^b$  intersected with the exterior neighborhood of  $A^b$  is non-empty, so  $EBc(A,B)$  will always be  $\geq 0$  depending on the pre-defined radius  $r$ . Table 2 lists these constraints (i.e., 0, 1,  $> 0$ , and  $\geq 0$ ), color-coded as specified by the legend in the table.

On examining Table 1 for all sixteen metrics (eight metrics and their converses), it is determined that some of the metrics have the same set of dependencies for the topological relations. For example, in Table 1,  $IVS(A,B)$  and  $EVS(A,B)$  have identical values in all rows. Two metrics are *equivalent* if they have the identical dependencies corresponding to all topological relations. A set of equivalent metrics is called an *equivalence class* of metrics. A set of equivalence classes forms a *partition* of the set of metrics. Upon complete examination of Table 1 in terms of the dependencies, the set of metrics can be partitioned into eleven equivalence classes:  $\{IVS(A,B), EVS(A,B), IVC(A,B)\}$ ,  $\{IVSc(A,B), EVSc(A,B), IVCc(A,B)\}$ ,  $\{BA(A,B)\}$ ,  $\{BAC(A,B)\}$ ,  $\{BS(A,B), BSc(A,B)\}$ ,  $\{EVC(A,B)\}$ ,  $\{EVCc(A,B)\}$ ,  $\{IB(A,B)\}$ ,  $\{IBc(A,B)\}$ ,  $\{EB(A,B)\}$ ,  $\{EBc(A,B)\}$ . Table 3 shows the number of topological relations that differ (in terms of dependencies) for each pair of these equivalence classes. This can be used as the basis for creating a conceptual neighborhood graph (CNG) to get a sense of the conceptual closeness of the metrics; see Fig. 3. The CNG is a connected minimal spanning graph. Edges selected to form the CNG are highlighted in Table 3.

The metrics at mutual distance 1 are: (1) metrics that refer to interior volume separation and interior volume closeness, and (2) metrics for exterior volume and boundary (interior or exterior) closeness. Of the remaining metrics, the Snapshot Model [12] is used to include all metrics at distance 2 in the connected minimal spanning graph.

The CNG based on topological-metric dependencies has no direct correlation with the CNG for RCC8 relations, which can be constructed based on the 9-Intersection model; in the latter, two RCC8 relations are conceptual neighbors if one relation can be deformed into the other (i.e., via translation, rotation, or scaling) without encountering any other relation in between. The topological-metric dependencies CNG is useful for easily identifying conceptual similarity between the various metrics; for example, the concept of interior volume splitting (IVS) is more similar to the concept of interior boundary (IB) than it is to exterior boundary closeness (EVC).

A more practical use of the topological-metric dependency CNG is as a means of measuring the semantic distance between two natural-language terms. The similarity between terms  $t_1$  and  $t_2$  could be measured by finding two nodes in the graph that correspond to metrics applicable to both terms, and finding the minimal-cost path between those nodes. This would need to be done for each pair of metrics that apply to the two terms, summing all those metric path costs, and dividing by the number of paths. If the result for terms  $t_1$  and  $t_2$  is smaller than the result for terms  $t_1$  and  $t_3$ , then terms  $t_1$  and  $t_2$  are conceptually more similar than terms  $t_1$  and  $t_3$  (i.e.,  $t_1$  and  $t_3$  are more likely to be considered synonyms than are  $t_1$  and  $t_2$ ).

## V. ASSOCIATION BETWEEN THE NATURAL-LANGUAGE TERMS, TOPOLOGICAL RELATIONS, AND THE METRICS

The primary objective of this work is to define a collection of metrics and topological relations that will facilitate a mapping to natural-language terminology for 3D objects, and thereby establish a practical mapping between qualitative and quantitative spatial reasoning and human cognition. To that end, insight can be gleaned from conducting human subject experiments in which human perception of images and words can be associated with the mathematical notions represented by the topological relations and metrics.

### A. Related Experiments

Human subject experiments were conducted by Shariff, Egenhofer, and Mark [3] wherein subjects were given several

Table 3. Topological-metric dependency differences for metric equivalence class pairs.

	IVS, EVS, IVC	IVSc, EVSc, IVCc	BS, BSc	IB	IBc	BA	BAC	EVC	EVCC	EB	EBc
IVS, EVS, IVC	0	4	3	4	2	2	4	6	2	3	5
IVSc, EVSc, IVCc	4	0	3	2	4	4	2	2	6	5	3
BS, BSc	3	3	0	1	1	1	1	3	3	2	2
IB	4	2	1	0	2	2	2	2	4	3	1
IBc	2	4	1	2	0	2	2	4	2	1	3
BA	2	4	1	2	2	0	2	4	2	3	3
BAC	4	2	1	2	2	2	0	2	4	3	3
EVC	6	2	3	2	4	4	2	0	4	3	1
EVCC	2	6	3	4	2	2	4	4	0	1	3
EB	3	5	2	3	1	3	3	3	1	0	2
EBc	5	3	2	1	3	3	3	1	3	2	0

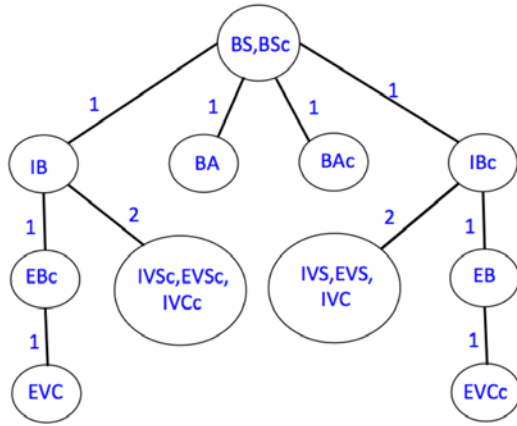


Figure 3. CNG for metrics based on topological-metric dependencies.

sentences each containing a natural-language spatial term such as “connects” and asked to draw a picture depicting that specified spatial relation between a park (i.e., a 2D region) and a road (i.e., a line or a curve). The drawings were then analyzed to measure the metrics that had been defined for a 2D region and a line. The subjects also were asked which of fifteen line-region topological relations they most closely associated with the natural-language spatial term in each sentence. Among the conclusions drawn from that study were that: (1) for the majority of the natural-language spatial terms, the topological relation is a more important influence (i.e., distinguishing feature) than any of the metrics, and (2) several natural-language spatial terms fall under the same topological relation, but have different metric values [3].

A similar experiment was conducted by Xu [4] whereby human subjects were given pictures of two lines and sentences containing a natural-language spatial term; the collection of terms used were not exactly the same as those used in the line-region study [3]. The subject was asked to rank his/her agreement as to whether the term described the spatial

configuration in the picture. As in the line-region study, the actual values of the metrics for each spatial configuration used in the experiment were measured and included in the dataset. However, it is important to note that the collection of metrics used in the line-line study was considerably different than those examined in the line-region study; the only metrics that were conceptually common to both studies were for three types of splitting. The line-line study also differed from the line-region study in its consideration of topological relations: (1) the collection of topological relations simply consisted of the entries in the 9-Intersection matrix (i.e., the intersection of each object’s interior, exterior, and boundary with that of the other object), and (2) each natural-language spatial term was pre-classified with applicable topological relations; the human subjects were not asked whether or not they thought that a certain topological relation applied to a particular natural-language term. One of the conclusions drawn from this study was that in most cases using topological and metric properties together produces better results for distinguishing natural-language terms than using only topological properties [4].

## B. Design and Analysis of Experiment with 3D Objects

### B.1. Design of the Experiment

To investigate the association between natural-language spatial terms, topological relations, and metrics for 3D objects, we conducted a human subject experiment with design and analysis aspects similar to those of the aforementioned studies. 119 human subjects (88 male, 31 female) were given a test consisting of 48 questions. Each question contained an image of two 3D spheres, one blue and one green, in the spatial configuration of disconnected (DC), externally connected (EC), partial overlap (PO), non-tangential proper part (NTPP), or tangential proper part (TPP); see Fig. 4. The converse relations TPPc and NTPPc were not tested as they would just be the reverse cases of TPP and NTPP. The relation EQ was not tested because we anticipated that the subjects might be confused if they could not detect both a green sphere and a blue sphere (i.e., because the spheres are “equal”). Each question also included a sentence of the form “The blue sphere *term* the green sphere.” where *term* was one of the 26 natural-language spatial terms listed in Table 4; with the exception of “disconnected”, none of the topological relations (e.g., “tangential proper part”) was used as a term. We did use some of the same terms that were used in the line-region [3] and line-line [4] studies. The subject was asked to rank his/her agreement as to whether the term described the spatial configuration in the picture, choosing from seven rankings: ‘strongly agree’, ‘agree’, ‘somewhat agree’, ‘neutral’, ‘somewhat disagree’, ‘disagree’, and ‘strongly disagree.’ We subsequently categorized the answers of ‘strongly agree’, ‘agree’, and ‘somewhat agree’ as ‘yes’; all other answers were categorized as ‘no.’ Not every natural-language term and topological relation pair was tested; for example, in the case of two disconnected spheres, we deemed that it would have been highly unlikely that anyone would have said that one sphere “encloses” the other sphere. The frequency of ‘yes’ and ‘no’ responses for the cases that were tested are shown in Table 4.

In most cases, there was a very clear consensus as to whether the majority of the subjects agreed or disagreed that a

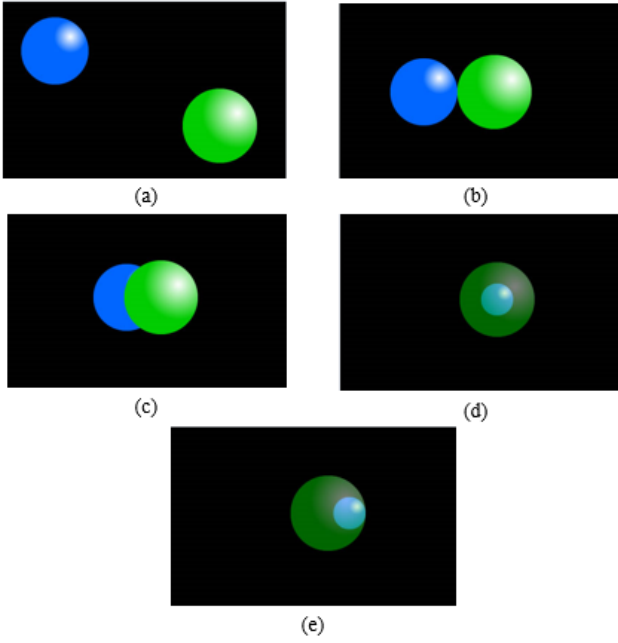


Figure 4. Images used in the experiment: (a) disconnected (DC), (b) externally connected (EC), (c) partial overlap (PO), (d) non-tangential proper part (NTPP), and (e) tangential proper part (TPP).

particular natural-language term corresponded to a topological relation (as indicated by the blue-highlighted entries in Table 4 showing  $\geq 72\%$  frequency, and to a lesser extent the orange-highlighted entries in Table 4 showing  $\geq 65\%$  frequency). However, there were six cases where the decision was fairly evenly split (i.e., approximately 50% agreeing/disagreeing and approximately 50% disagreeing/agreeing): “is connected to”, “cuts across”, and “intersects” for partial overlap; “enters” and “goes into” for tangential proper part; and “goes into” for non-tangential proper part. These cases (highlighted in green in Table 4) are indicative of a natural-language term that is ambiguous for describing a particular topology.

As was concluded in the line-region study [3], we found that there were some natural-language spatial terms for which more than one topological relation applied (e.g., “is contained within”, “encloses”, “is inside of”, “is outside of”, and “is within”). It was noted in the line-region study that in such cases the metric values associated with a natural-language term could be different [3]. We computed the metric values for each of the five spatial configurations used in our study. Using k-means clustering (*SimpleKMeans* in the WEKA software, <http://www.cs.waikato.ac.nz/ml/weka>), we too found that some of the natural-language terms that mapped to more than one topological relation had different values for their metrics.

## B.2. Analysis of the Experiment

Analysis of the 3D object experiment dataset was conducted with the objective of investigating the extent of the association between the natural-language terms, the topological relations, and the metrics. Each row of the dataset consisted of a topological relation (DC, EC, PO, TPP, or NTPP), the eleven metric equivalence classes with the dependency values for that particular topological relation, the calculated value of the metric for the depicted spatial

configuration, and a natural-language spatial term. For this part of the analysis, we removed all responses (rows) where the subject did not agree that the spatial term described the topological relation that had been depicted between the two 3D objects (spheres in this case). We also performed the testing on smaller datasets, each containing the results for only a single natural-language term. Analysis was done both with and without considering the topological relation as an attribute, for both the smaller datasets and the complete dataset as a whole.

We first used multinomial logistic regression (*SimpleLogistic* in the WEKA software) to see which of the eleven metric equivalence classes would be the most likely predictors of (and hence the most important for distinguishing) the natural-language terms. Our results showed that the classes  $\{IVS(A,B), EVS(A,B), IVC(A,B)\}$ ,  $\{IVSc(A,B), EVSc(A,B), IVCc(A,B)\}$ , and  $\{BS(A,B), BSc(B,A)\}$  were most significant in distinguishing natural-language terms based on both the metric dependency constraints and the metric values. All other metric classes either had zero weight or were dependent on other metric equivalence classes. Using logistic regression alone, the resulting model was able to correctly classify the natural-language terms using these metrics 79% of the time. These three metric classes were equally important; when we removed any one or two of them, the model would use one of the other metrics in this group and achieve the same accuracy. However, if we removed all three, the accuracy dropped noticeably. When we included the topological relations in the dataset, it made no difference in the results under this model. When we tested classification of the natural-language terms in the dataset using the topological relation alone without the metric attributes, the number of correctly classified terms dropped to only 13% using the logistic model.

We then examined the C4.5 [13] decision trees that could be built for each of our datasets using *J48* in the WEKA software. Recall that each node of a C4.5 decision tree contains the attribute of the dataset that most effectively splits the instances into one class or another [13]. What we determined in every case was that, in the absence of the topological relation, these same three metric classes (e.g.,  $\{IVS(A,B), EVS(A,B), IVC(A,B)\}$ ,  $\{IVSc(A,B), EVSc(A,B), IVCc(A,B)\}$ , and  $\{BS(A,B), BSc(B,A)\}$ ) always were included when building the decision trees, and that the accuracy of correctly classified natural-language terms ranged from 76% to 98% for datasets containing a single natural-language term. An example for a case where we tested the dataset as a whole and eliminated the topological term as an attribute is shown in Fig. 5. Although the accuracy was low on the total dataset, this example demonstrates the significance of the three metric classes (e.g.,  $\{IVS(A,B), EVS(A,B), IVC(A,B)\}$ ,  $\{IVSc(A,B), EVSc(A,B), IVCc(A,B)\}$ , and  $\{BS(A,B), BSc(B,A)\}$ ) for defining multiple natural-language terms. The notation of the form (N/E) next to a node in the tree represents N as the number of instances that reached that node and E as the number of instances that differed in the value identified at that node. The first rule in the tree shown in Fig. 5 would be interpreted as “if the average value of the metrics in class  $\{IVSc(A,B), EVSc(A,B), IVCc(A,B)\} = 0.333333$  and the average value of the metrics in class  $\{BS(A,B), BSc(A,B)\}$



Table 4. Frequency of responses for natural-language spatial terms used in human subject experiment.

	DC - yes	DC - no	EC - yes	EC - no	PO - yes	PO - no	TPP - yes	TPP - no	NTPP - yes	NTPP - no
is adjacent to	27%	73%	91%	9%						
is connected to			76%	24%	50%	50%				
is contained within							81%	19%	85%	15%
crosses			13%	87%	72%	28%				
cuts					30%	70%				
cuts across					47%	53%				
cuts through					35%	65%			29%	71%
is disconnected from	91%	9%								
divides					20%	80%			17%	83%
encloses							77%	23%	83%	17%
enters							50%	50%		
goes across					69%	31%				
goes into					29%	71%	50%	50%	52%	48%
goes through									29%	71%
goes up to	9%	91%	41%	59%						
is in							82%	18%		
is inside of							82%	18%	87%	13%
intersects			24%	76%	49%	51%	40%	60%	38%	62%
merges with					34%	66%			31%	69%
is near	42%	58%	94%	6%						
is outside of	87%	13%	75%	25%						
splits					18%	82%				
touches			88%	12%						
transects					20%	80%	21%	79%	19%	81%
transverses					19%	81%			10%	90%
is within							79%	21%	86%	14%

```

=== Classifier model (full training set) ===

J48 pruned tree
-----
IVSc = 0.333333
| BS = 0: disconnected (304.0/196.0)
| BS = 0.059: isNear (596.0/484.0)
| BS = 0.0715: isNear (0.0)
| BS = 1.75: isNear (0.0)
IVSc = 0.342: isin (668.0/570.0)
IVSc = 0.389667: crosses (588.0/502.0)
IVSc = 0.666667: isinside (672.0/569.0)

Number of Leaves :    7
Size of the tree :    9

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   490      17.3267 %
Incorrectly Classified Instances 2338      82.6733 %
Kappa statistic                  0.1297
Mean absolute error              0.0575
Root mean squared error          0.1696
Relative absolute error          89.8974 %
Root relative squared error      94.8492 %

```

Figure 5. *J48* pruned tree for distinguishing natural-language terms based on metric equivalence class values.

= 0 then *A is disconnected from B*” where “is disconnected from” is a natural language term (not the topological relation DC). Due to attribute naming constraints in WEKA, metric class {IVSc(A,B), EVSc(A,B), IVCc(A,B)} was named simply *IVSc*, and *BS* was used for metric class {BS(A,B), BSc(A,B)}. The

certainty of this particular rule is not high (35.5%); of 304 data instances that reached this node in the decision tree built using *J48*, only 196 of those instances actually agree with this decision. In general, the datasets that contained multiple natural-language terms had much lower classification accuracy (than the datasets that contained a single natural-language term) because there were instances where multiple natural-language terms mapped to multiple topological relations, and hence multiple metric equivalence classes.

Our *J48* analysis found that every natural-language term used some combination of the aforementioned three metric classes in its decision. If we removed these metrics from the dataset, the classification accuracy drastically dropped. Adding the topological relations to a dataset did not change the accuracy results; however, we did note that if we removed {EVS(A,B)} from a dataset, *J48* would replace it in the tree with a topological relation and reduce the tree size, making it slightly more efficient computationally.

What the results of this experiment tell us is that there are three metric equivalence classes that can fairly accurately define the majority of the natural-language terms in this dataset for 3D objects based on metric values (when the dataset contains a single natural-language term); similar results were obtained based simply on metric dependencies with topological relations. The other metrics were either dependent on these three metric equivalence classes, or were of very little significance in the final decision. In fact, two of the metric equivalence

```

=== Classifier model (full training set) ===

J48 pruned tree
-----

RCC8 = DC: disconnected (304.0/196.0)
RCC8 = EC: isNear (596.0/484.0)
RCC8 = NTPP: isInside (672.0/569.0)
RCC8 = PO: crosses (588.0/502.0)
RCC8 = TPP: isIn (668.0/570.0)

Number of Leaves :    5

Size of the tree :      6

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   490      17.3267 %
Incorrectly Classified Instances 2338      82.6733 %
Kappa statistic                  0.1297
Mean absolute error              0.0575
Root mean squared error          0.1696
Relative absolute error          89.8974 %
Root relative squared error      94.8492 %

```

Figure 6. J48 pruned tree for distinguishing natural-language terms using only topological relations (no metrics).

classes,  $\{EB(A,B)\}$  and  $\{EBc(A,B)\}$ , were never used in any of the decision trees and had zero weight in the logistic regression analysis. We also found that using the topological relation with the metrics made no difference in terms of accuracy, although it did make computation of the decision (i.e., the natural-language term determination) slightly more efficient in a few cases, namely those involving  $EVS(A,B)$ . Using only the topological relations (no metrics) did not improve the results for distinguishing one natural-language term from another; see Fig. 6 for the J48 decision tree that results from making decisions based only on the five topological relations tested in this experiment (DC, EC, NTPP, PO, and TPP).

These results are in contrast to those reported in the line-line and line-region studies wherein topology was found to play a more important role in distinguishing natural-language terms. At this preliminary point in our research, we cannot attribute this difference simply to the 3D nature of the spatial topology; further experimentation is necessary whereby, for example, measurement of our metrics for more examples of each spatial configuration should be analyzed for possible correlations. At this stage we also cannot conclude that only the metrics in the equivalence classes  $\{IVS(A,B), EVS(A,B), IVC(A,B)\}$ ,  $\{IVSc(A,B), EVSc(A,B), IVCc(A,B)\}$ , and  $\{BS(A,B), BSc(B,A)\}$  are valid for 3D objects.

## VI. FUTURE WORK

Some of the metrics presented herein are simplifications; they do not consider all relevant aspects of the topology for 3D objects. While the 9-Intersection model is sufficient to determine the topological connectivity, it is not sufficient to determine the qualitative extent of connectivity. For example,  $PO(A,B)$  embodies that  $A \cap B$ ,  $A^b \cap B$ , and  $A \cap B^b$ , are all nonempty,

but it does not quantify the precise or qualitative commonality; we need metrics. Since the metrics measure the commonality in addition to the topological connectivity, for some of the metric equations we need to consider additional parameters such as the interior, exterior, boundary, interior neighborhood, and exterior neighborhood for each object. This will be explored in our subsequent future research.

In the future we also will present new metrics for the obscuration relations of VRCC-3D+, and investigate how the connectivity and obscuration relations combined with the metrics affect distinguishing natural-language spatial terms. With that integration of connectivity and obscuration, we expect practical applications of this work will include robotic navigation via voice control as well as natural-language user interfaces for 3D spatial querying.

## VII. SUMMARY

Topology alone has been found to be insufficient for conveying spatial knowledge in natural-language communication. Based on previous work that has been done to define metrics for two lines and a line and a 2D region in order to facilitate a mapping to natural-language terminology, herein we defined metrics appropriate for 3D regions. The association between this collection of metrics, 3D connectivity relations, and several English-language spatial terms was tested in a human subject study. We found three metric equivalence classes that could define the natural-language terms in our experiment dataset for 3D objects. In contrast to the results reported for line-line and line-region studies, we found that using the topological relation with the metrics made no difference in terms of the accuracy in defining natural-language terms. This work is too preliminary to attribute this as a phenomenon of 3D topology. However, we believe our work is an interesting starting point in a world that is being saturated with 3D data and thus is in need of automated spatial reasoners with a natural-language interface.

## REFERENCES

- [1] A. Tversky, 1981, Distortions in Memory for Maps. *Cognitive Psychology*, 13, pp. 407-433.
- [2] L. Talmy, 1983, How Language Structures Space. In *Spatial Orientation: Theory, Research, and Application*, H. Pick and L. Acredolo (Eds) (New York: Plenum Press), pp. 225-282.
- [3] R. Shariff, M.J. Egenhofer, and D.M. Mark, 1998, Natural-Language Spatial Relations Between Linear and Areal Objects: The Topology and Metric of English-Language Terms. *International Journal of Geographical Information Science*, 12, pp. 215-246.
- [4] Jun Xu, 2007, Formalizing Natural-Language Spatial Relations Between Linear Objects with Topological and Metric Properties. *International Journal of Geographical Information Science*, 21:4, pp. 377-395.
- [5] M.J. Egenhofer and D.M. Mark, 1995, Naïve Geography. In *Proceedings of the International Conference on Spatial Information Theory: A Theoretical Basis for GIS (COSIT'95)*, A.U. Frank and K. Kuhn (Eds.) (Berlin: Springer-Verlag), pp. 1-15.
- [6] C.L. Sabharwal, J.L. Leopold, and N. Elloe, 2011, A More Expressive 3D Region Connection Calculus. In *Proceedings of the*



2011 International Workshop on Visual Languages and Computing (in conjunction with the 17<sup>th</sup> International Conference on Distributed Multimedia Systems (DMS'11))), August 2011, Florence, Italy, pp. 307-311.

- [7] J. Albath, J.L. Leopold, C.L. Sabharwal, and A.M. Maglia, "RCC-3D: Qualitative Spatial Reasoning in 3D", *Proceedings of the 23<sup>rd</sup> International Conference on Computer Applications in Industry and Engineering (CAINE 2010)*, Las Vegas, NV, Nov. 8-10, 2010, pp. 74-79.
- [8] C.L. Sabharwal and J.L. Leopold, "Smooth Transition Neighborhood Graphs for 3D Spatial Relations", *Proceedings of IEEE Workshop on Computational Intelligence for Visual Intelligence (CIVI 2013)*, Singapore, Apr. 16-20, 2013, pp. 8-15.
- [9] K. Nedas, M.J. Egenhofer, D. Wilmsen, 2007, Metric Details of Topological Line-Line Relations. *International Journal of Geographical Information Science*, 21:1, pp. 21-48.
- [10] D.A. Randell, Z. Cui, and A.G. Cohn, 1992, A Spatial Logic Based on Regions and Connection. In *Proceedings of the 3<sup>rd</sup> International Conference on Knowledge Representation and Reasoning*, October 1992, Cambridge, MA (San Francisco: Morgan Kaufmann), pp. 156-176.
- [11] M.J. Egenhofer, J. Sharma, and D. M. Mark, 1993, A Critical Comparison of the 4-Intersection and 9-Intersection Models for Spatial Relations: Formal Analysis. *11th Int. Symposium on Computer-Assisted Cartography*, American Society of Photogrammetry and Remote Sensing, pp.1-11.
- [12] M. Santos and A. Moreira, 2009, Conceptual Neighborhood Graphs for Topological Spatial Relations. In *Proceedings of the 2009 World Congress on Engineering*, Vol. I, July 2009, London, UK, pp. 12-18.
- [13] J.R. Quinlan, 1993, C4.5: Programming for Machine Learning. San Mateo, CA: Morgan Kaufmann.

# Nonoverlapped View Management for Augmented Reality by Tabletop Projection

Makoto Sato and Kaori Fujinami

Department of Computer and Information Sciences,  
Tokyo University of Agriculture and Technology

Email: jawdp3@gmail.com, fujinami@cc.tuat.ac.jp

## Abstract

*Augmented reality (AR) by a projector allows easy association of information by using a label with a particular object. When a projector is installed above a workspace and pointed downward, supportive information can be presented; however, a presented label is deformed on a non-planar object. Also, a label might be projected in a region where it is hidden by the object, i.e., a blind area. In this paper, we propose a view management technique to allow interpretation by improving the legibility of information. Our proposed method, the Nonoverlapped Gradient Descent (NGD) method, determines the position of a newly added label by avoiding overlapping of surrounding labels and linkage lines. The issue of presenting in a shadow area and a blind area is also addressed by estimating these areas based on the approximation of objects as a simple solid. An experimental evaluation showed that the visibility of the labels was improved with this method.*

## 1 Introduction

Augmented reality (AR) presents computational information to the real world. A large amount of work has examined desktop tasks using AR in tabletop projections [8,9,11,12,20]. Labels with a textual or graphical form are often utilized in AR to provide information. Users obtain information once after recognizing a label that corresponds to a physical object. Legible presentation is necessary to communicate a message correctly, and view management improves label visibility. Investigation on view management methods have been studied for the see-through type AR, such as through a head-mounted display or a hand-held display [1,2,5,10,14,17]. However, few studies show view management methods for projection type AR [6,15,18].

In see-through type AR, labels are presented on a display by superimposition with a video captured image. However, in projection type AR, labels are projected in the real world. In this case, the following two issues should be taken into account. First, labels are deformed when they are overlapped by objects. The linkage line that connects a label with a target object is also deformed. Second, from the user's perspective, a label may be hidden by a tall object. In both issues, a user can only see a part of information or might even not notice the presence of the label, which prevents interpretation. This is critical in applications in which speed and correctness of associating labels with a target object are important, such as in a chemistry experiment [16]. These issues are unique to projection type AR that deals with the three-dimensional relationship between a target object, a projector, and a user. In contrast, a view management method for see-through type AR does not need to consider these issues, because all things happen in two-dimensional space. Existing see-through view management methods cannot be directly applied to projection type AR.

In this paper, we propose a new view management technique for projection type AR. Figure 1 summarizes the variations of view management that include (a) no view management, (b) the existing method, Gradient Descent (GD), for see-through type AR, (c) our proposed method, Nonoverlapped Gradient Descent (NGD), without overlap, and (d) and (e) our proposed method with gradual improvements. The core idea in our method is to determine the position of a label by changing the distance of a line to a target object, *linkage line*, until no overlapping is detected (c). Here, "overlapping" includes not only each label and line, but also an object with a label or line. Overlapping of the presentation with an object is regarded as projecting in a *shadow area* of an object. Meanwhile, a *blind area* of an object from a user's perspective is estimated to avoid hidden projection from a user (d). So, estimating

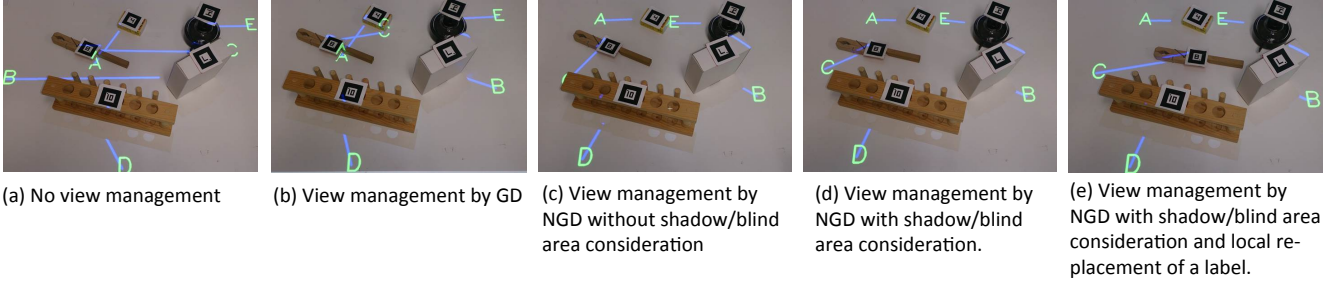


Figure 1: Various label placements: (a) no view management, which makes it difficult to read letters and find association of labels and target objects, (b) traditional method (GD) for see-through type AR [1] that is realized with linkage lines with equal length, (c) our basic approach (NGD), which is an improved version of (b) that changes the length of linkage lines until no overlap is detected, (d) our approach (NGD) with an extension to three-dimensional space that considers shadow and blind areas of an object, and (e) our approach (NGD) with local label placement computation for a newly moved object. Note that a linkage line is a line that connects a label with an object to clearly indicate the association.

both shadow and blind areas allows improved legibility of information. These areas are estimated by approximation into a cylinder or a cuboid based on the shape of the area on the surface. The approximation allows fast detection of overlapping by a two-dimensional geometric computation. As can be seen in Figure 1-(d), labels are placed without any overlapping and have high legibility of information, whereas (a), (b), and (c) do not have high legibility. Furthermore, our method recalculates label placement for an object that changes its position only (e).

The rest of the paper is organized as follows: Section 2 examines related work. Design considerations in the view management method for projection type AR are shown in Section 3. Section 4 proposes a novel view management method (NGD), which intends to avoid any overlapping as an extension of the traditional method designed for see-through type AR. The integrated view management system is presented in Section 5, followed by a basic performance evaluation of the system and future work in Section 6 and Section 7, respectively. Finally, Section 8 concludes the paper.

## 2 Related Work

Much work has been done to improve the visibility of *label placement*. Makita et al. [10] proposed a technique for a dynamic environment where target objects move. Their approach employs *lines* to associate information to an object in a head-mounted display. The position of information presentation is determined by minimizing the cost function represented by three factors: 1) the distance of the line, 2) the area of overlapping information with a target object, and 3) the amount of movement of the information between frames. Grasset

et al. [5] investigated a label placement method for an AR browser that estimates a critical region from an image frame and avoids placing rendering information on it. Azuma et al. [1] proposed four algorithms that determine the position of a label with less overlap. In this paper, we adapt their method to a projection type AR environment.

Little work has been done on projection type AR. Projection type AR for a wearable projector was proposed by Uemura et al. [18]. Their method allows a person who performs some tasks against a wall to obtain clear wall-projected information without overlap by his/her hands. A label placement technique for a non-planar and textured surface was proposed by Iwai et al. [6]. Image distortion was also addressed. However, our assumed target objects for annotation are thin, tall, sharp objects that stand on a table, such as a test tube, a gas burner and a pipette. Objects are often made of glass, so this technique is not suitable for direct projection because the legibility of the information is degraded. The work of Siriborvornratanakul et al. [15] determines an appropriate projection area for a handheld projector without overlapping in a dynamic cluttered environment. Their approach, based on image processing, can avoid overlapping of a projected image with physical objects; however, their approach is limited to projection in a situation where the projector and the user are on the same side. In the case of projection on a table, a blind area might appear according to the height of an object and the spatial relationship between the objects and the user. Furthermore, their approach aims at finding an area for projection and does not handle label placement.

### 3 Design Considerations

In this section, we present the requirements for view management of projection type AR, more specifically, for tabletop projection.

#### 3.1 Association of a label with a target object

A label is usually associated with a physical object. Direct projection of a label onto a non-planar and textured surface object, e.g., [6], seems to be natural; however, the applicability depends on the material and the shape of an object. A transparent object as well as an odd-shaped object may significantly degrade the legibility of the information. As shown in Figure 2-(a), the two labels ‘D’ and ‘E’ projected on a transparent and an odd-shaped object, respectively, cannot be read. Therefore, we project the label on the flat surface on which the object is placed. However, an issue of ambiguous presentation still exists. Here, label ‘A’ is presented next to a glass beaker, which is easily associated; however, in the case of label ‘C’ (upper left of Figure 2-(a)), a user might be confused with which of three surrounding objects the label is actually associated. Thus, we decided to project labels on a table with a *line* connected to the bottom of an object. We call this a *linkage line*.

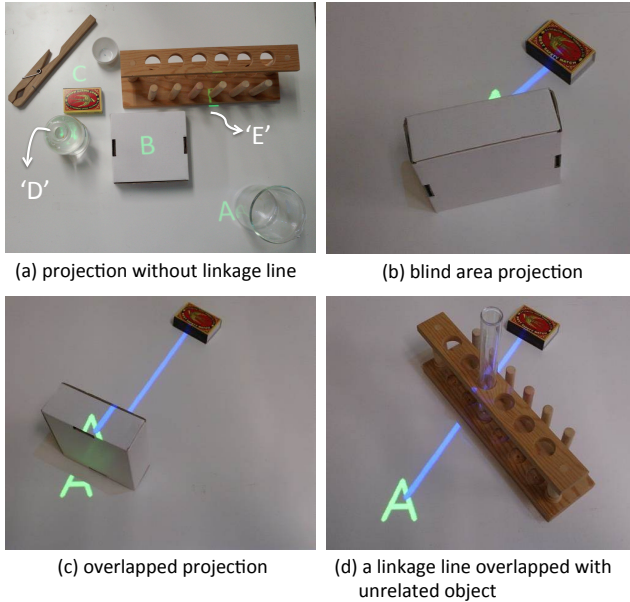


Figure 2: Issues in projecting information near or on 3D objects

#### 3.2 Blind area projection

No information is hidden by objects in see-through type AR because information is computationally superimposed with a captured image on a device screen. However, it is not applicable to the case of tabletop projection. Information can be projected onto a blind area, from the user’s perspective, on a table. As shown in Figure 2-(b), information for the small box (match box) is projected on an area of a table that is hidden by a large box (white cardboard box). A user can see only a small portion of the projected information. Thus, if the information is projected onto such an area, the information may not be communicated. A user may not even notice the presence of information when the projected information enters into a full blind area. Such a blind area is considered as the *critical projection area*. View management for tabletop projection type AR should avoid the critical projection area.

#### 3.3 Overlapped projection

In Figure 2-(c), label ‘A’ is overlapped with a large object, while Figure 2-(d) shows a situation where a linkage line is projected on an unrelated object, i.e., a test tube rack, placed between a label and its target object. Such deformation of the label and the linkage line may degrade both the time of interpretation of the label with the object and the correctness of interpretation. This is also a uniqueness of projection type AR. Thus, a view management method for see-through type AR is not applicable to projection type AR. Instead, a label should be positioned by taking into account the shape and the size of objects near the target object. Also, the linkage line should be directly drawn between the label and the target object. Azuma et al. investigated a view management method that resolves not only the overlapping of labels but a label and an object for see-through type AR [1]. We extend their method so that it can handle the issues in projection type AR.

### 4 Nonoverlapped Gradient Descent Method

We introduce the method proposed by Azuma et al. [1] called the Gradient Descent (GD) method and then we present our extension. A preliminary comparative experiment on the two methods is also presented.

## 4.1 GD method

Azuma et al. identified the following three factors that increase the visibility of see-through type AR:

- Less overlapping of labels
- Short distance between a label and its target object
- Small amount of movement of labels between frames.

Azuma et al. proposed four algorithms: Adaptive Simulated Annealing (ASA), Greedy, Clustering, and GD. We focused on GD because it allowed the most correct association in the user study. In GD, the distance between a label and a target object is constant. The processing flow is as follows. First, the candidate position of label placement is determined every 10 degrees around a target object. Then, the *cost* of overlapping is calculated. Finally, the most distant candidate from other objects and labels is selected if there are more than two-least cost candidates. The cost function has four elements with different weights, as follows:

- Between labels: 10
- Between lines: 2
- Label and object: 1
- Label and line: 1.

The overlapping detection in the approach of Azuma et al. is divided into five pairs: 1) between labels, 2) between a label and an object, 3) between linkage lines, 4) between a label and a linkage line, and 5) between an object and a linkage line. Their method classifies both a label and an object as rectangles. This classification simplifies the overlapping detection with the geometric calculation of line vs. rectangle, rectangle vs. rectangle, or line vs. line. Note that their technique does not focus on the overlapping between a label and an object, as the weight of “1” suggests. Although, the increase of the weight of “label and object” might improve the overlapping in such a case. As pointed out in Section 3, a label is deformed and degrades its legibility when it is overlapped with an object. We suppose overlapping has the most impact on projection type AR. Thus, we extend GD with variable distance between a label and an object to eliminate overlapping, which we call NGD.

## 4.2 Nonoverlapped Gradient Descent (NGD) method

As described above, the key idea is the variable length of the linkage line. Similar to GD, candidate

label positions are set around a target object by 10 degrees once overlapping is detected (A and B in Figure 3). If there is no area without overlap around the object (C), the distance of the linkage line is increased (D). This process is repeated until no overlapping is found. The selection rule in the case of multiple label positions with the same distance follows that of GD (E). Steps A, B, and E are identical to those of GD, which is also true for the cost function.

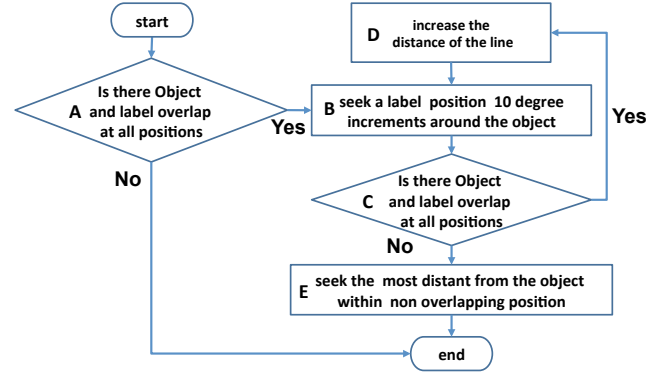


Figure 3: Processing flow of NGD. The processes ‘A’ and ‘B’ are identical to those of GD.

Figure 4 shows the resultant views of the two methods. As shown, the distance of the linkage line is constant, and overlapping of lines and objects (rectangles) is observed in GD (Figure 4-(a)). In contrast, NGD with variable length shows no overlapping (Figure 4-(b)).

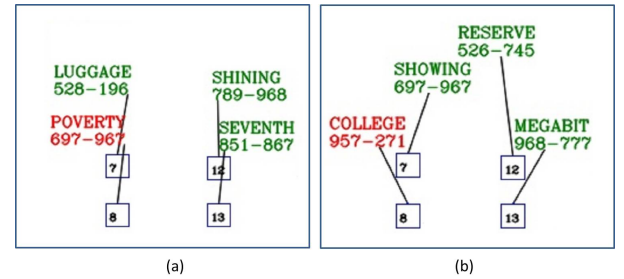


Figure 4: Resultant views of GD (a) and NGD (b). Rectangles indicate objects, while the labels in red are what the subjects of comparative study in Section 4.3 were asked to read in the experiment.

## 4.3 Comparative study of GD and NGD

We carried out a comparative study on the correctness and legibility of association, as well as on the speed of association. Fourteen university students in their



20s (ten men, four women) participated in the experiment. We followed the experimental scheme of Azuma et al. [1]. However, one exception is that we utilized tabletop projection rather than see-through AR. This is because we intended to compare the two methods under the possible deformation of lines and labels caused by overlapping. Here, the critical projection areas, i.e., the shadow and the blind areas, are not considered; they can be regarded as the bottom of an object when they are handled.

Figure 5 shows a snapshot of the experimental projection. Twenty cubes were placed on a table (W 66 × D 50 [cm]). The cubes were small enough (2 [cm] square) to avoid the critical projection area. Each cube has a label that consists of alphanumeric words. The label is linked to the center of a dedicated object. Subjects were asked to read a pair of red words and the number on the cube they thought that were associated with the words. They read four pairs in one trial, and a total of 20 trials were conducted for each subject. To avoid order bias, the subjects were divided into two groups: one group were tested with GD for the first 10 trials and with NGD for the latter 10 trials, while the other group started with NGD followed by GD. The position of the cubes did not change; however, the positions of the labels and numbers were randomly determined for each trial to avoid memory effects. The words were also changed for each trial to avoid habituation, although the length of the words was equalized to seven.

Three performance metrics were collected in this experiment: the time to association, the correct association rate, and the correct reading rate. The time to association, i.e., task completion time, was measured by differentiating the start time of a particular trial from the last word of the fourth pair, i.e., number. The difference between the correct association rate and the correct reading rate is that the correct association rate just checks whether the association is successful regardless of the failure of reading the words. In contrast, the correct reading rate shows a more rigid performance of association by counting only the successful readings.

#### 4.4 Result

The average task completion time of a subject with GD and NGD was 10.1 sec and 8.0 sec, respectively (Figure 6-(a)). The result of a paired t-test for the two groups with a significance level of 5% showed that the NGD completion time was significantly shorter ( $t(26) = 6.9 \times 10^{-5}$ ,  $p < 0.05$ ). Although individual differences were seen in the reading speed, we consider that this does not have a major impact on the outcome because

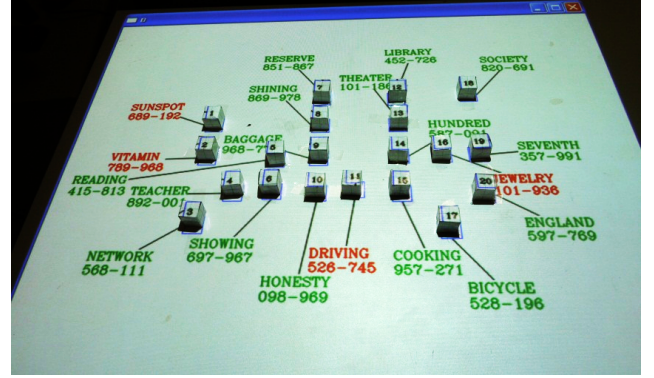


Figure 5: Snapshot of the projected labels in NGD.

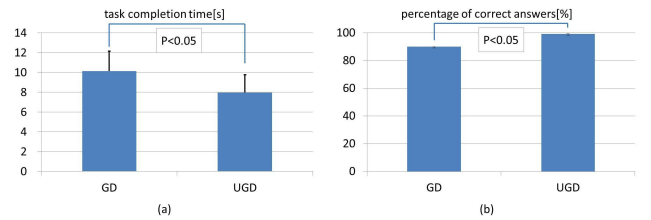


Figure 6: Comparison with GD and NGD: (a) task completion time and (b) correct association rate for a subject

all subjects were tested with both GD and NGD.

The average correct association rate of a subject for GD and NGD was 90.0% and 99.0%, respectively (Figure 6-(b)). Similar to the task completion time, the result of the paired t-test showed that the correct association rate of the NGD method was significantly higher ( $t(26) = 2.4 \times 10^{-5}$ ,  $p < 0.05$ ). In contrast, the correct reading rate for GD and NGD was the same at 97.9%.

#### 4.5 Discussion

As shown above, NGD is superior to GD in the speed and the correctness of association. These results suggest that avoiding overlapping played a key role. Makita et al. [10] defined a cost function by the length of the linkage line and the overlapping for view management of a head-mounted display. However, our result indicates that the impact of the length on the legibility is smaller than that of overlapping. In other words, we can ignore the negative impact of the length of a linkage line and proceed with NGD unless we intend to develop a room-size application.



## 5 View Management with NGD

A view management system is proposed based on NGD. The system considers various object properties.

### 5.1 Handling various shapes and sizes of objects by models

NGD uses various object properties such as position, shape and size to avoid overlapping of a label. In the experiment of Section 4.3, the objects were uniform. Also, they were small enough so that we could ignore the shadow and blind areas described below; however, to be applicable to real-world applications, we need to take into account the object properties.

An object is modeled into either a cuboid or a cylinder that is circumscribed to the object to allow fast overlapping detection. An alternative method is to represent a target object with a cloud of points, *point cloud*, from a depth camera. The point cloud approach can lead to precise estimation of the critical projection area and efficient use of the desktop real estate; however, we consider that this approach requires more computational power than the model-based one due to predominantly large number of points to represent the contour of the object. So, we decided to take the modeling approach.

Basically, an object can be modeled into a cylinder (Figure 7 left); however, a long, thin object on the surface, such as a fork or a pen, is modeled as a rectangular solid (Figure 7 right). The cylindrical approximation requires a large base area with a diameter of the length of the object. This consumes more area, as the *critical projection area* cannot be used to place a label. Therefore, the candidate positions for a label near a target object decrease, and the search time increases due to incrementing the search range. Furthermore, a critical situation may occur when a small object is placed inside a large circle defined by a long, thin object (see Figure 8). Here, the label for the small object will not be placed anywhere because the linkage line between the center of the small object and its label crosses the base circle for a cylinder, which means overlapping cannot be avoided. Thus, cuboid approximation addresses this issue.

To focus on overlap handling rather than 3D object recognition and shape measurement, we applied a visual marker approach, where necessary information is retrieved by the ID of a marker from an external database. The types of information linked to an ID are as follows: 1) the type of applicable model, i.e., cylinder or cuboid; 2) the size of the base surface of the object; and 3) the height of the object. We assume

cylindrical approximation of tall object

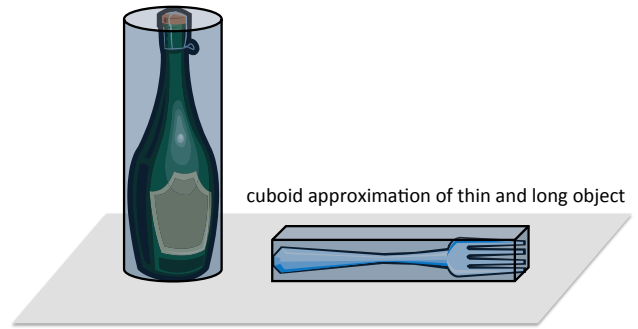


Figure 7: Model approximation



Figure 8: Drawn circle indicating the base area of a test tube clamp when cylindrical approximation is adopted.

that these information are input into the database in advance. The two-dimensional position on the surface is obtained by a camera, which is regarded as the center of gravity of an object, and the linkage line is drawn toward a label.

### 5.2 Finding the shadow area and the blind area

A shadow area appears due to the shade of an object from the projector's light source. In other words, light is overlapped with an object. Thus, avoiding a shadow area in projection leads to projection without overlapping of a label with the object. A number of studies proposed the *hard shadow technique* to estimate the shadow area based on a point light source [3, 4, 13, 19]. We adopted the *planar projection shadow* method [3], because it can calculate the shadow area very quickly when a shadow is cast on a planar surface and the number of objects is small. In the model approximation approach, the shadow area is drawn by connecting the feature points of an object projected on the surface from the light source. Here, the feature points of a particular model are analytically calculated.

For cylindrical modeling, two points, which are on the diameter of the upper surface of a cylinder, are obtained. The line segment formed by the two points is

perpendicular to a line connecting the light source with the center of the upper surface of the cylinder. Then, the two points on the cylinder are projected on the surface, i.e., table, by calculating the intersections of the surface and the lines that connect the light source and the points. The projected points constitute the diameter of a circle, which is combined with the rectangular area that corresponds to the body of the cylinder. Here, although the shadow is actually a part of an ellipse, we regard it as a circle for simplicity of calculation. The right-hand part of Figure 9 illustrates an example. Note that we assume that the shape of the shadow of the cylinder's body is a rectangle for the ease of computation. The points forming the silhouette of the cuboid are obtained in a similar way to those of the cylinder. Here, four vertices of the upper surface of the cuboid are projected on the surface. Thus, overlapping of a projected label with an object (see Figure 2-(c)) can be avoided by placing the label at a position outside the shadowed area on the tabletop coordinates.

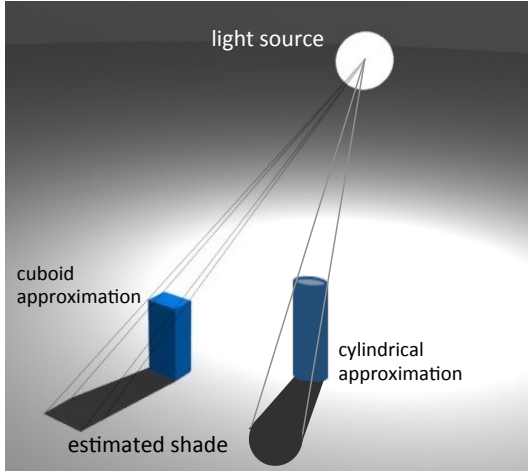


Figure 9: Estimation of shadow areas of rectangular solid (left) and cylinder (right)

As pointed out in Section 3.2, the blind area appears based on the spatial relationship between an object, a projector, and the viewpoint (Figure 2-(b)). The case of a projected label being hidden from a person by an object (Figure 2-(b)) can be handled in the same manner as the case with the shadow area projection. The light source by the projector can be replaced with the viewpoint of a person. The calculation of the blind area from the user's perspective is equivalent to that of the shadow area by the projector, so a label can be seen by a person if it is placed outside the blind area. Our proposed method, NGD, is adopted for avoiding both the shadow area by the projector and the blind area from the user's viewpoint.

### 5.3 Detecting overlap by geometric computation

Labels and linkage lines are projected on a planar surface, i.e., a table, which allows us to limit the calculations to two-dimensional coordinates. The shadow area and the blind area of a cuboid is represented by a polygon, whereas that of a cylinder is a combination of a circle and a rectangle. In addition, a label is represented as a rectangle. Now, we can focus on detecting the overlapping of lines, rectangles, and circles. The detection consists of the following processes:

- Between labels: The containment of a vertex of one rectangle in the other
- Label and object: The containment of a vertex of a label's rectangle in a polygon or a circle of an object
- Between linkage lines: The intersection of one line segment with the other
- Between a label and linkage line: The intersection of one of four sides of a rectangle of a label with a line segment
- Between an object and linkage line: The intersection of a side of a polygon for an object with a line segment or that of a circle with a line segment.

Containment checking of a vertex is actually realized as follows. In the case of a circle for cylinder approximation, the intersection is checked based on the distance between a vertex and the center of the circle. Meanwhile, a cuboid represented by a polygon is decomposed into line segments. So, the existence of the intersection between a segment of a label and a segment of the polygon proves the fact of overlapping of a label by an object. The overlapping detection method even works in the case that all four vertices of a label are contained in a critical projection area by a large object. This is because the linkage line between one of the vertices and the center of the target object should cross the border of the area. An exception is that the target object itself is contained in the critical projection area, which is examined in terms of the effect of the cuboid approximation in Section 6.2.

### 5.4 System configuration and Implementation

Now that we have designed the major system functionalities, they can be integrated into a system. Figure 10 shows a block diagram of label placement from image acquisition to label rendering, which also shows the relationship between major functionalities. A captured

image frame is sent to the component responsible for object identification and localization (marked as ‘A’ in Figure 10). Object and label information DB (‘F’) is updated by the extracted information. As described in Section 4.1, Azuma et al. [1] suggested that a small amount of label movement between frames increased the visibility. However, to avoid the frequent change of label positions, the system checks whether there is sufficient movement in any object (‘B’), in which 10 pixels of displacement of an object triggers a new process of label placement. Otherwise, the position of the label is not changed, but the length of the linkage line changes (Figure 1-(e)). This is, of course, applicable if no overlapping is detected. In ‘C’, the shadow and blind areas for all objects are estimated by the method described in Section 5.2. Then, NGD is applied to a moved object (‘D’), which allows the reduction of computation. Finally, the position of a label and the center of mass of a target object are used to draw the linkage line as well as the label itself (‘E’). This flow is iterated about every 7.33 [msec]. We used ARToolkit as a tool for extracting the position and ID of an object and OpenCV for rendering labels and lines.

As shown in Figure 1, NGD with shadow and blind area consideration (d) significantly improves the legibility of information compared with the presentation without view management (a) or with just resolving the overlap with labels (b). In (e), an object to which label ‘C’ is linked moved to the right from the position in (d); however, the position of the label is not changed to avoid degradation of the comprehension of information. An internal image of the shadow and blind estimation is shown in Figure 11, where the white and the black colored areas indicate the shadow and the blind areas, respectively. The system finds an appropriate position for a label by avoiding these areas.

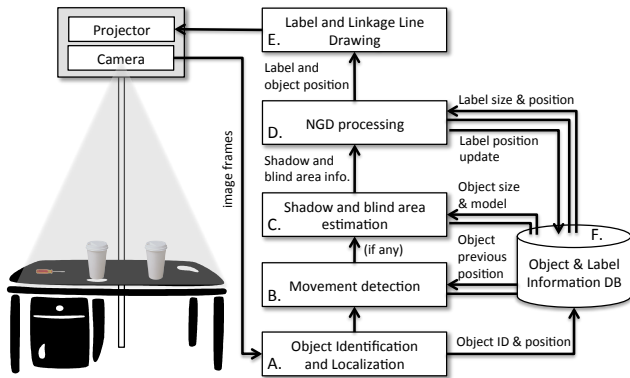


Figure 10: Block diagram of view management system based on NGD

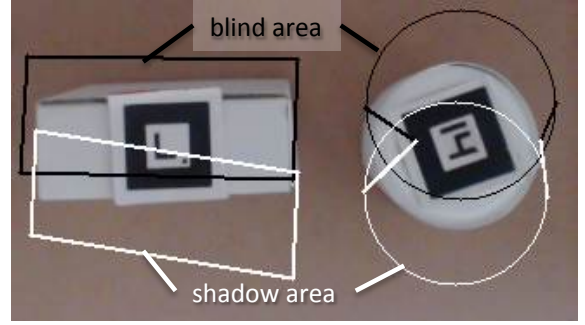


Figure 11: Internal image of shadow and blind area estimation. Note that the white and the black colored areas indicate the shadow and the blind areas for the objects, respectively.

## 6 Evaluation

We carried out evaluations of the processing speed of major functionalities and the effect of model approximation.

### 6.1 Processing time

The elapsed time for major functionalities was measured in a prototype system running on a PC (OS: Windows 7 64 bit, CPU: 2.8 GHz Intel Core 2 duo, RAM: 4 GB). Five objects (a test tube clamp, a pipette and a matchbox as cuboids and a spirit lamp and a beaker as cylinders) were used. The average elapsed time was calculated over 20 trials. Table 1 shows these average times, in which the elapsed time for estimating shadow and blind areas is for one object, while the elapsed time for overlapping detection includes that for all objects. The total elapsed time was less than 1.0 [msec], which is due to the simple containment and the intersection checking method described in Section 5.3.

Table 1: Average processing time [msec]

Functionality		Elapsed time
Estimating shadow and blind areas of cylinder	cylinder	0.57
	cuboid	0.97
Overlapping detection	blind area	0.01
	label with an object	0.05

### 6.2 Effect of cuboid approximation

As described in Section 4.2, NGD utilizes the area defined by a model of an object to remove overlapping.

The aim of introducing the cuboid approximation is to reduce the critical projection area for a long, thin object, and to allow an object and a label to be placed near such an object. Here, we examine the effect of the cuboid approximation in comparison with the cylindrical approximation. Five subjects moved the three types of objects, i.e., a test tube clamp, a pipette, and a matchbox, as they chose. These objects were modeled as both cylinders and cuboids. The number of times that the centroid of an object was placed within an area of the model was counted. Five subjects tried 20 times for each type of approximation.

Table 2 shows the results per subject (A to E). Containment in the cylindrical area occurred in every trial, whereas almost no case was observed in the cuboid approximation. However, this does not suggest that a cylindrical model is not necessary. As shown in Table 1, the processing speed of the shadow and blind areas for the cuboid approximation took slightly longer than that of the cylinder. Although the difference is less than 0.5 [msec], it increases as the number of objects increases because the processing speed of these areas depends on the number of objects. Therefore, an appropriate model selection is a good option for improving real-time processing performance in the case that the number of objects is expected to be large in a particular application.

Table 2: Ratio of an object contained within an area defined by a model for another object by subject (A–E)

Situation/Subject	A	B	C	D	E
Within a cylinder	1.00	1.00	1.00	1.00	1.00
Within a cuboid	0.05	0.00	0.00	0.00	0.00

The size of an object is currently measured and registered into the system by hand, which is a burdensome task. The type of model approximation is also predefined; however, sometimes static approximation is not adequate. For example, a bottle lying on its side on a table can be modeled as a cuboid, while a cylindrical approximation is suitable when it is standing upright on a table. Thus, dynamic measurement and model selection allows more precise detection of overlapping. A depth sensor, e.g., Microsoft Kinect, would be a solution for this challenge [7]. A depth sensor provides the shape of an object as a cloud of points, and so it could remove the model approximation process; however, it may sacrifice the simplicity of shadow and blind area estimation. Therefore, a hybrid approach is worth considering as an alternative, in which a depth sensor is employed to identify an appropriate model, as well as to obtain the size of the model.

## 7 Future Work

The estimation of shadow and blind areas plays an important role in the usefulness of NGD. Therefore, we will carry out an in-depth evaluation on the correctness of the estimation by comparing the area of an actual shadow or a blind area with an estimated area. The comparative study in Section 4.3 was conducted under controlled conditions. To see whether the effectiveness can be scaled up to a real-world application, we are planning to conduct another comparative user study under a particular scenario, e.g., a chemistry experiment, that contains objects with various shapes, sizes, and layouts on a table. Furthermore, automatic model selection and size measurement based on depth information is under investigation.

## 8 Conclusion

We proposed a novel view management technique for tabletop projection type AR, in which we investigated the importance of considering the shape, the size, and the material of tabletop objects for placing a label in a meaningful manner. The issue we dealt with was overlap of a label with other objects, labels, and linkage lines. Also, projection on the critical projection area, i.e., the shadow and blind areas, was another issue to further improve the legibility of projected information.

To address the issue of overlapping labels, we proposed the Nonoverlapped Gradient Descent (NGD) method, which was designed as an extension of the GD method of Azuma et al. [1]. The difference between GD and NGD is that NGD varies the length of the linkage line between an object and the label. Although Azuma et al. suggested a short distance between a label and its target object for better legibility of information, we prioritized the overlap issue over the distance-derived issue in projection type AR due the three-dimensional nature of the display environment. The result of a preliminary user study showed that NGD was superior to GD in the time and the correctness of associating a label with its corresponding object.

Regarding the issue of the critical projection area, we proposed a method to estimate the area by model approximation of either a cylinder or a cuboid based on preregistered information for each object. We showed that cuboid approximation utilized the desktop real estate efficiently.

We consider that the proposed method would open a door for applying a tabletop projector-based AR to a critical domain in which the speed and the correctness of associating labels with a target object is important, such as in a chemistry experiment [16].

## Acknowledgements

This work was supported by MEXT Grant-in-Aid for Scientific Research (C) No. 24500142.

## References

- [1] R. Azuma and C. Furmanski. Evaluating Label Placement for Augmented Reality View Management. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '03, pp. 66–75, 2003. IEEE Computer Society.
- [2] B. Bell, S. Feiner, and T. Höllerer. View Management for Virtual and Augmented Reality. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, pp. 101–110, 2001. ACM.
- [3] J. Blinn. Me and My (Fake) Shadow. *IEEE Comput. Graph. Appl.*, 8(1):82–86, 1988.
- [4] F. C. Crow. Shadow algorithms for computer graphics. In *Proceedings of the 4th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '77, pp. 242–248, 1977. ACM.
- [5] R. Grasset, T. Langlotz, D. Kalkofen, M. Tatzgern, and D. Schmalstieg. Image-driven view management for augmented reality browsers. In *Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '12, pp. 177–186, 2012. IEEE Computer Society.
- [6] D. Iwai, T. Yabiki, and K. Sato. View management of projected labels on nonplanar and textured surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 19(8):1415–1424, 2013.
- [7] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pp. 559–568, 2011. ACM.
- [8] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner. The reacTable: Exploring the Synergy Between Live Music Performance and Tabletop Tangible Interfaces. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, TEI '07, pp. 139–146, 2007. ACM.
- [9] S. K. Kane, D. Avrahami, J. O. Wobbrock, B. Harrison, A. D. Rea, M. Philipose, and A. LaMarca. Bonfire: A Nomadic System for Hybrid Laptop-tabletop Interaction. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, pp. 129–138, 2009. ACM.
- [10] K. Makita, M. Kanbara, and N. Yokoya. View management of annotations for wearable augmented reality. In *Proceedings of IEEE International Conference on Multimedia and Expo*, ICME 2009, pp. 982–985, 2009.
- [11] S. Morioka and H. Ueda. Cooking Support System Utilizing Built-in Cameras and Projectors. In *Proceedings of 12th IAPR Conference on Machine Vision Applications*, MVA'12, pp. 271–274, 2011.
- [12] B. Piper, C. Ratti, and H. Ishii. Illuminating Clay: A 3-D Tangible Interface for Landscape Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pp. 355–362, 2002. ACM.
- [13] W. T. Reeves, D. H. Salesin, and R. L. Cook. Rendering Antialiased Shadows with Depth Maps. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '87, pp. 283–291, 1987. ACM.
- [14] F. Shibata, H. Nakamoto, R. Sasaki, A. Kimura, and H. Tamura. A View Management Method for Mobile Mixed Reality Systems. In *Proceedings of the 14th Eurographics Conference on Virtual Environments*, EGVE'08, pp. 17–24, 2008. Eurographics Association.
- [15] T. Siribornvornratanakul and M. Sugimoto. Clutter-aware Adaptive Projection Inside a Dynamic Environment. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, VRST '08, pp. 241–242, 2008. ACM.
- [16] A. Sökan, M. Hou, N. Shinagawa, H. Egi, and K. Fujinami. A Tangible Experiment Support System with Presentation Ambiguity for Safe and Independent Chemistry Experiments. *Journal of Ambient Intelligence and Humanized Computing*, 3(2):125–139, 2012.
- [17] K. Tanaka, Y. Kishino, M. Miyamae, T. Terada, and S. Nishio. An Information Layout Method for an Optical See-through Head Mounted Display Focusing on the Viewability. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '08, pp. 139–142, 2008. IEEE Computer Society.
- [18] K. Uemura, K. Tajimi, Y. Kajiwarra, N. Sakata, M. Billingham, and S. Nishida. Annotation view management for wearable projection. In *Proceedings of the 20th International Conference on Artificial Reality and Telexistence*, ICAT2010, pp. 202–205, 2010.
- [19] L. Williams. Casting Curved Shadows on Curved Surfaces. In *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '78, pp. 270–274, 1978. ACM.
- [20] M. Wu and R. Balakrishnan. Multi-finger and Whole Hand Gestural Interaction Techniques for Multi-user Tabletop Displays. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, UIST '03, pp. 193–202, 2003. ACM.

# Markerless Hand Gesture Interface Based on LEAP Motion Controller

Danilo Avola, Andrea Petracca  
Giuseppe Placidi, Matteo Spezialetti  
Dep. of Life, Health and Environmental Sciences  
University of L'Aquila  
Via Vetoio Coppito, 67100, L'Aquila, Italy  
Email: (danilo.avola, andrea.petracca)@univaq.it  
Email: (giuseppe.placidi, matteo.spezialetti)@univaq.it

Luigi Cinque, Stefano Levialdi  
Dep. of Computer Science  
Sapienza University  
Via Salaria 113, 00198, Rome, Italy  
Email: (cinque, levialdi)@di.uniroma1.it

## Abstract

*Hand gesture interfaces provide an intuitive and natural way for interacting with a wide range of applications. Nowadays, the development of these interfaces is supported by an increasing number of sensing devices which are able to track hand and finger movements. Despite this, the physical and technical features of many of these devices make them unsuitable for the implementation of interfaces oriented to the everyday desktop applications. Conversely, the LEAP motion controller has been specifically designed to interact with these applications. Moreover, this latter device has been equipped with a hand skeletal model that provides tracking data with a high level of accuracy.*

*This paper describes a novel approach to define and recognize hand gestures. The proposed method adopts free-hand drawing recognition algorithms to interpret the tracking data of the hand and finger movements. Although our approach is applicable to any hand skeletal model, the overall features of that provided by the LEAP motion controller have driven us to use it as a reference model. Extensive preliminary tests have demonstrated the usefulness and the accuracy of the proposed method.*

**Keywords:** hand gesture definition, hand gesture recognition, feature extraction, LEAP motion controller.

## 1. Introduction

The diffusion of consumer sensing devices has promoted the development of novel Human-Computer Interfaces (HCIs) able to track body and/or hand movements. These interfaces process the captured sensing information to provide body and/or hand models through which recognize poses, movements, and gestures. The accuracy of the model (and of the recognition process) depends on the spe-

cific application. In fact, the interfaces designed for applications in the field of rehabilitation require greater accuracy than those designed for entertainment. These interfaces can be classified as Natural User Interfaces (NUIs), or Haptic Interfaces (HIs). The term NUIs is referred to interfaces in which the interaction between human and machine is controlled by poses and movements of the body (and its parts) without using any tool or wearing any device. While, the term HIs is related to interfaces in which the interaction is controlled by signals emitted from sensors equipping a some kind of body suite and/or glove. Although HIs can be extremely accurate, their use is restricted to some specialized fields (e.g., robotic, manufacturing) due to high costs, and hard customization. Moreover, different contexts (e.g., serious games) require that users can use these interfaces without physical constraints, cumbersome devices, or uncomfortable tools. In our context, we were interested in managing a hand skeletal model to interact (by one or two hands) with everyday desktop applications (e.g., data management, icon browsing), for this reason we were focused in investigating the more suitable current NUI. A first example of NUI is represented by the early Computer Vision (CV) based Motion Capture (MoCap) systems which were equipped with one or more RGB cameras [22, 15]. Although these systems still prove their effectiveness and efficiency, most of them can not be considered genuine NUIs since their tracking algorithms are based on visual expedients (e.g., markers, coloured suites and gloves). The recent introduction of high-resolution and high-speed RGB cameras has supported the implementation of newer markerless CV based MoCap systems able to track even subtle movements of the hand articulation [17, 20]. Despite this, the use of these systems (early and newer) is not particularly suitable to interact with everyday desktop applications. In fact, as well known, these systems require more than a camera (i.e., view) to recognize and track a body and its hands. This last aspect introduces some hard technical issues, in-



cluding calibration, synchronization, and data processing. The recent proliferation of consumer Time of Flight (ToF) and Structured Light (SL) range imaging cameras [7] has allowed us to solve some of the above issues. In particular, these devices have allowed developers to implement NUIs having a 3D perception of the observed scene by using a single device. In fact, these cameras provide both a set of RGB maps and a set of depth maps for each second captured within their Field of View (FoV). Although this new generation of NUIs is profitably used in different application fields (such as: entrainment, rehabilitation, and movement analysis), some practical (e.g., size, shape) and technical (e.g., resolvable depth, depth stream) aspects do not promote their use in interacting with the mentioned class of applications. The LEAP motion controller [13] is a latest NUI that allows us the implementation of advanced hand gesture recognition systems, its physical and technical features make it an ideal tool to interact with any kind of desktop application. In fact, the LEAP motion controller is a light and tiny device that has been designed to be placed on a desk. Its working area is defined by the 8 cubic feet of space above itself, the device has been cleverly conceived to track palm and finger movements since this is the common hand pose of a user while interacting with a desktop application. The new version of the LEAP motion API (version 2.0, [14]) introduces a new hand skeletal model that provides additional information about hands and fingers and also improves overall tracking data. Their model allows the device to predict the positions of fingers and hands that are not clearly in view, the hands can often cross over each other and still be tracked. Despite this, the LEAP motion controller remains a device having a single viewpoint, this implies that occlusions (or inaccurate evaluations) can occur when users perform complex hand poses or subtle motions, especially those involving non-extended fingers.

This paper describes a novel approach to define and recognize hand gestures. The skeletal model of the LEAP motion controller provides tracking data in which palm and finger movements are expressed by a set of 3D spatial information (i.e., space  $(x,y,z)$ ) over the time (i.e., time  $t$ ). Our main idea has been to project these 3D spatial and temporal information within their 2D reference planes (i.e., planes  $(x,y)$ ,  $(x,z)$ ,  $(y,z)$ , over time  $t$ ). In this way, any 3D hand gesture can be interpreted through the analysis of its projections on the related 2D reference planes. Each projection can be seen as a freehand drawing whose features can be extracted through algorithms belonging to the sketch recognition field [11, 4]. This paper is focused in showing the approach independently from the specific device, for this reason here we are not interested in occlusion resolution.

The rest of the paper is structured as follows. Section 2 discusses basic background about freehand drawing recognition. Section 3 presents an overview of the framework,

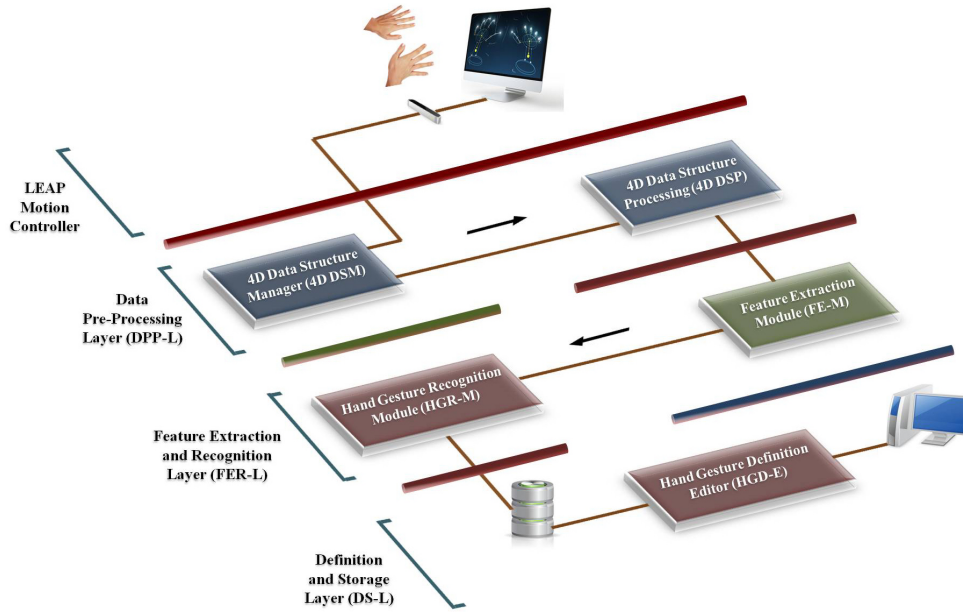
and introduces the preliminary tests. Finally, Section 4 concludes the paper and shows future directions.

## 2. Background

In this paper we present an ongoing research work, this implies that in this section we are not interested in comparing the proposed approach with others, our intent is to provide a survey of the works that more than others have contributed to define our method. The freehand drawing processing faces different issues including shape recognition and style identification. The first one regards the ability of a system in distinguishing a set of hand-drawn symbols constituting a graphical library. The second one concerns the ability of the above system in recognising each hand-drawn symbol independently of the style (e.g., bold, solid) used by users in tracing it. Both issues can be addressed by studying a set of mathematical features (i.e., feature vector) able to characterize every hand-drawn symbol, its style, including a certain degree of perturbation.

### 2.1. Freehand Drawing Processing

A first remarkable work that has supported the implementation of our feature vector is presented in [9]. In this paper the authors propose a robust and extensible approach to recognize a wide range of hand-drawn 2D graphical symbols. Their method recognizes symbols independently of sizes, rotations, and styles. Their feature vector is based on the computation of the convex hull and of three special polygons from it derived: largest triangle, largest quadrilateral, and enclosing rectangle. Different works have inherited the above feature vector to customize and extend the library of symbols. Among others, the works presented in [3] and [2] provide an interesting viewpoint to generalize the definition and recognition of any set of hand-drawn 2D graphical symbols. The first work describes the Feature calculation Bid Decision (FcBD) system, which adopts an agent-based architecture to introduce new symbols inside a defined library. Their system implements some new measures and adds them to the feature vector presented in [9]. Moreover, it provides a practical strategy to manage conflicts that occur when two similar symbols are introduced. The second work can be considered an application of that just described. In this case, the authors have adopted the FcBD system to define a fixed library of symbols, then they have implemented a simple CV based MoCap system equipped with a single RGB camera to recognize them. The users could perform hand gestures by using a suitable tool or a coloured glove. This approach can be considered an early version of that we propose in the present paper, where the use of the RGB camera reduces the efforts to a



**Fig. 1. The framework design is composed by three layers: Data Pre-Processing Layer (DPP-L), Feature Extraction and Recognition Layer (FER-L), and Definition and Storage Layer (DS-L).**

single reference plane (i.e.,  $(x,y)$ ). An agent-based architecture has been also designed in [8], where the authors define a framework for interpreting hand-drawn symbols in a context-driven fashion, exploiting heterogeneous techniques for the recognition of each symbol. Their framework has been adopted to derive AgentSketch, a multi-domain sketch recognition system able to work in on-line and off-line mode. As shown by the just introduced works, many authors are focused on conceiving freehand drawing recognition systems to identify more than a single set of graphical symbols. In fact, the implementation of a recognizer to identify a specific set of symbols is a time-consuming operation. Often, the introduction of a single new symbol may require the whole re-implementation of the system. Another work that we have analysed is shown in [1], where the authors presents SketchREAD, a multi-domain sketch recognition engine capable of recognizing hand-drawn diagrammatic sketches. Their system is based on a suitable description of the shapes according to the related domain, moreover SketchREAD does not require of training data or re-implementation processes. Two last meaningful works are described in [10] and [5]. The first introduces LADDER which has been the first language to describe how sketched diagrams in a domain are drawn, displayed, and edited. The last details a framework through which users can define every set of hand-drawn 2D symbols. The symbols are defined and recognized by using a novel Sketch Modeling Language: SketchML. The proposed framework adopts the SketchML to formalize and manage the gestures.

### 3. The Framework Architecture

As shown in Figure 1, this section describes the framework architecture that implements the proposed approach to define and recognize hand gestures. Although the method is applicable to any hand skeletal model, that provided by the LEAP motion controller has addressed our requirements since its real time tracking data has a high level of accuracy. The rest of the section is structured as follows. A first sub-section will present main details about the device and hand skeletal model. The second, third, and fourth sub-sections will explain each one of the three pipeline layers that compose the architecture, respectively. Finally, a last sub-section will discuss the preliminary tests. Overall, the framework process the hand gestures to extract the information of the trajectory of palm and fingertip movements. These information are treated as graphical objects and subsequently interpreted by means of freehand drawing algorithms. This interpretation provides an univocal classification of each gesture. Note that, in this phase we are interested in recognizing a set of separated gestures, so we have not introduced mechanisms to distinguish the beginning and end of gestures. Moreover, even the issue of disambiguation of similar gestures is left to future extension of the present work. Despite this, it should be observed that a first level of disambiguation is guaranteed by the same algorithms that perform the freehand drawing recognition, since also in that case it is important the ability of a system in distinguishing between two similar shapes (e.g., ellipse and circle).

The first layer, Data Pre-Processing Layer (DPP-L), takes as input each frame generated by the LEAP motion controller, and extracts the 3D spatial (i.e.,  $(x,y,z)$ ) and temporal (i.e.,  $t$ ) information of the hand model. Subsequently, these 3D spatial information are projected inside their 2D reference planes ( $(x,y)$ ,  $(x,z)$ , and  $(y,z)$ ) and the same time ( $t$ ) is associated to each plane. The second layer, Feature Extraction and Recognition Layer (FER-L), takes as input the 2D spatial and temporal information of each plane, and adopts, on each of them, a freehand drawing recognition algorithm. The purpose of this algorithm is to provide as result three shapes (one for each plane) according to a default set of shapes stored within a repository. The combined interpretation of these shapes provides a classification of the hand gesture. The last layer, Definition and Storage Layer (DS-L), is responsible for the definition and storage of the shapes contained within the repository. Actually, the working of the whole framework starts from this layer, since in a first phase one or more libraries of symbols have to be created. When a user builds a library, each introduced symbol is processed by an approach similar to that used to recognize it, where the three computed shapes are related to each other and stored within the repository as identifying features of the performed gesture.

### 3.1. LEAP Motion Controller

The LEAP motion controller is a device equipped with three IR emitters and two CCD cameras that has been designed to support hand gesture based interfaces having a high accuracy in detecting hand and fingertip position. Although the raw data is currently not conventionally accessible, the provided hand model incorporates a rich set of information [21]. In particular, each frame is connected to a complex set of data structures and methods of which we report a main subset:

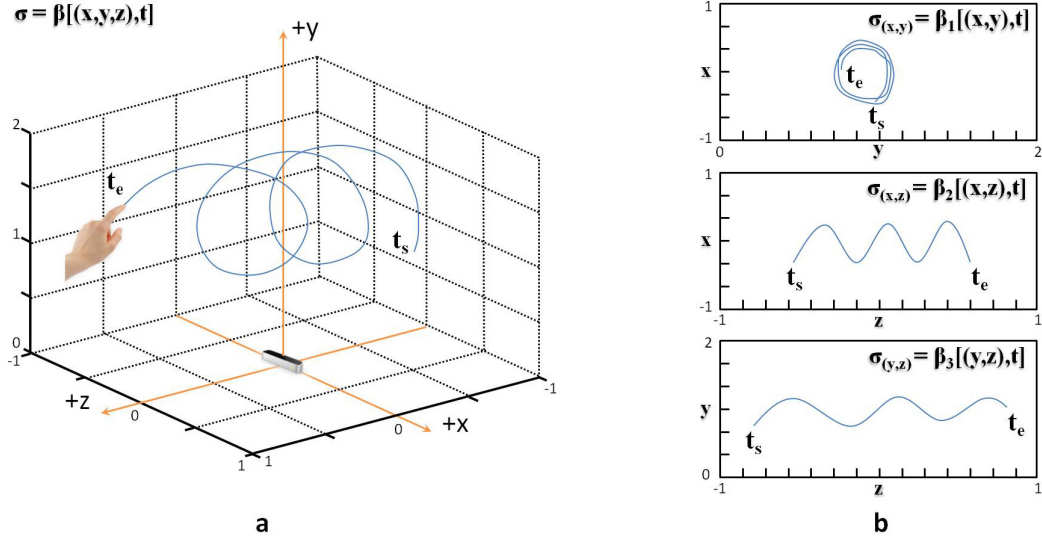
- **Frame Data:** FrameID, Timestamp, Hands, Fingers, Tools, and Gestures;
- **Hand Data:** HandID, Direction, Palm normal, Palm position, Palm velocity, Sphere center, Sphere radius, Translation, Rotation axis, Rotation angle, and Fingers IDs;
- **Finger and Tool Data:** PointableID, Belongs to (hand or tool), Classified as (finger or tool), Length, Width, Direction, Tip position, and Tip velocity;
- **Gesture Data:** containing spatial and temporal information about a fixed set of hand gestures: circles, swipes, key taps, and screen taps.

The LEAP motion controller employs a right-handed Cartesian coordinate system with origin centered at the top of the device. All the distances are computed in millimetres, the time in seconds (or microseconds), and the angles

in radians. The frame data contains quantitative information about the current frame, including how many hands and fingers have been detected. The device also recognizes tools as pencils or pens which can be used to interact with applications. The frame data also reports if a hand movement belongs to a default gesture (e.g., swipes), in fact the current version of the hand model provides a minimal set of gestures which are not editable or expandable. We have adopted all these information to set up the starting state of our data structure which is initialized when at least a hand or a tool are detected. The hand data contains the physical characterization of a detected hand. It reports the direction from the palm position toward the fingers (direction), the normal vector to the palm (palm normal), the center position of the palm from the device (palm position), and the rate of change of the palm position (palm velocity). In addition, it provides the center and the radius of a virtual sphere fit to the curvature of the hand (sphere center and sphere radius, respectively). Finally, the hand data also reports the change of position of a hand between the current frame and a specified frame (translation), the axis of rotation derived from the change in orientation of the hand between the current frame and a specified frame (rotation axis), and the angle of rotation around the specified axis derived from the change in orientation of this hand between the current frame and a specified frame. The finger and tool data reports the length and width of each finger or tool detected (i.e., object projection on the  $(x,z)$  plane). In addition, it contains the direction in which each finger or tool is pointing (direction), the tip position from the device origin (tip position), and the rate of change of the tip position (tip velocity). Hands and related fingers are connected each other by a simple identification (IDs) mechanism. The palm and fingertip tracking data represent the core information of the proposed definition and recognition algorithm. Finally, we did not have enabled the gesture recognition engine of the device (disabled for default) since we did not need of those information.

### 3.2. Data Pre-Processing Layer

The Data Pre-Processing Layer (DPP-L) is composed of two sub-modules: the 4D Data Structure Manager (4D DSM), and the 4D Data Structure Processing (4D DSP). The first sub-module takes as input the stream of frames coming from the device, and seeks information about hands (i.e., palm and fingertips) or tools. When at least a hand or a tool is found the sub-module creates and initializes our data structure. Hereinafter we will not treat the interaction with a tool, since this case is similar to one of a hand with a single extended finger. The second sub-module takes as input the 3D spatial and temporal information of the palm and fingertips, and projects each of them within of three related 3D reference planes. In order to explain the working



**Fig. 2.** An example of gesture performed by using a hand and a single finger: (a) its representation within the 3D space,  $(x,y,z)$ , over the time,  $(t)$ , (b) its projections on the 2D reference planes,  $(x,y)$ ,  $(x,z)$ , and  $(y,z)$ , over the time,  $(t)$ .

of the DPP-L, in this section we describe the processing of a gesture performed by a hand with a single extended finger. The generalization of the method is due to the fact that palm and fingers are treated separately during the recognition process, and only in the final step all the recognition information are correlated to interpret the gesture. As shown in Figure 2a, when a user performs the above gesture, the 4D DSM computes the information contained within the set of frames, and supports the definition of the stroke  $\sigma$ :

$$\sigma = \beta[(x, y, z), t] \quad (1)$$

Where,  $\beta$  represents a function of the 3D spatial coordinates  $((x,y,z)$ , derived by fingertip position data) over the time  $(t)$ , derived by fingertip velocity data). In particular,  $t_s$  and  $t_e$  represent the start and the end times of the gesture, respectively. Subsequently, the stroke is supplied to the 4D DSP which, as shown in Figure 2b, derives the three projected strokes according to their reference planes:

$$\sigma_{(x,y)} = \beta_1[(x, y), t] \quad (2)$$

$$\sigma_{(x,z)} = \beta_2[(x, z), t] \quad (3)$$

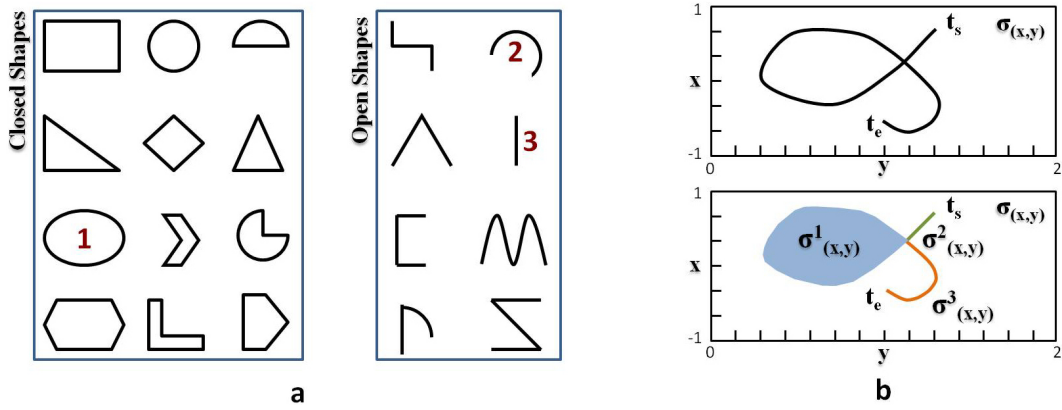
$$\sigma_{(y,z)} = \beta_3[(y, z), t] \quad (4)$$

Where,  $\beta_i$  ( $i = 1, \dots, 3$ ) represents a function of the 2D spatial coordinates (planes  $(x,y)$ ,  $(x,z)$ , and  $(y,z)$ , respectively) over the time  $(t)$ , same time for each one). The just introduced example describes the simplest case in which the single index finger is extended in a stationary pose. Note that, in case of the whole hand the framework is initialized when at least a open hand is centered over the device with all five extended fingers. Moreover, here occlusions are not treated.

### 3.3. Feature Extraction and Recognition Layer

The Feature Extraction and Recognition Layer (FER-L) is composed of two sub-modules: the Feature Extraction Module (FE-M), and the Hand Gesture Recognition Module (HGR-M). The first sub-module takes as input the three projected strokes associated to the center position of the palm (i.e., a group of three correlated strokes), and all five fingertips (i.e., five groups each one having three correlated strokes). Note that, our notion of hand gesture is based on tracking data of palm and fingers. Although the gesture ambiguity resolution is not a focus of the present paper, preliminary empirical observations have highlighted that the tracking of the palm (in addition to that of the fingers) represents a supplementary feature to univocally define and recognise a gesture. The above sub-module analyses each projected stroke, of each group, to identify the type of the shape that has been indirectly “drawn”. In this way, the movement of each fingertip, as well as that of the center of the palm, can be identified by three correlated shapes. The final step is to consider each group of three shapes with the others to provide an univocal interpretation of the hand gesture. Our freehand drawing recognition algorithm is based on the same feature vector shown in [9] and extended in [3], below we report the main geometrical measures through which the feature vector has been implemented:

- **original measures:** convex hull, largest triangle, largest quadrilateral, and enclosing rectangle;
- **extended measures:** angle ratio, and sketch perimeter.



**Fig. 3. Shape library supporting the freehand drawing definition and recognition algorithm: (a) closed and open shapes, (b) recognition of a stroke, on the plane  $(x,y)$ , composed of three shapes: a closed shape (1), and two open shapes (2 and 3).**

The above feature vector is able to recognize a fixed set of closed and open shapes. Although the format of the current paper does not allow us a complete dissertation of the used freehand drawing recognition algorithm, in Figure 3a we report the vectorial representation of the main shapes that it recognizes. Note that, the imported feature vector is able to recognize shapes independently of their sizes, rotations, and styles. This means that the proposed framework is robust enough in interpreting univocally the same hand gesture performed by users having different “tracing” styles. As shown in Figure 3b, each projected stroke can be composed of more than a shape. This last aspect has been solved by adopting an algorithm able to detect how many closed and open “regions” compose a single stroke [5]. Subsequently, the same algorithm compares each “region” with the shape library to identify the set of shapes that compose the stroke. From the work presented in [5] we have also inherited the SketchML, a structured language to describe, by constructs, the shapes, their features, and their constraints. In this way, we have introduced within the framework a suitable tool to manage each open and closed shape. As a result, the SketchML can be seen as the language to represent each treated hand gesture. When a user performs a hand gesture above the LEAP motion controller, each fingertip and the center of the palm are represented by a set of shapes. These shapes are described by the SketchML language. The second sub-module takes as input these descriptions, and matches them with all ones previously acquired during the hand gesture editing phase.

### 3.4. Definition and Storage Layer

The Definition and Storage Layer (DS-L) includes the Hand Gesture Definition Editor (HGD-E), which allows

users to define any hand gesture library. When the framework is in editing mode, a user can perform a hand gesture and a related feature vector is created. The framework allows users to repeat the gesture several time to obtain a more reliable vector. However, we have implemented a parameter (i.e., a mathematical norm) to set the accuracy of the framework. This means that when the difference between two consecutive feature vectors is lower than the norm, the hand gesture is accepted and stored within the gesture library.

### 3.5. Preliminary Tests

In this paper we present an ongoing research work, for this reason the preliminary tests have been summarized as a set of technical qualitative observations aimed in confirming the usefulness of the proposed approach. First of all, the idea to transform the 3D spatial and temporal information into a set of 2D projected strokes is simple and affective. The use of a freehand drawing recognition algorithm to classify the projected strokes works properly. We have initially tested the framework by performing (on different reference planes) the set of 2D graphical symbols inherited from the works that have provided the feature vector (see [9] and [3]). Subsequently, we have performed a set of combined hand gestures (as that shown in Figure 2a). In both cases the framework has stored on the database a suitable set of SketchML descriptions able to identify univocally the hand gestures. Also in this case the paper format does not allow us an exhaustive explication of the experiments, besides in this phase we were not interested in designing a practical hand gesture library for a specific application, out intent has been to check the work of the method and of the framework. However, the limit of the approach is the set of open and closed shapes which could be insufficient to rec-

ognize each projected stroke. To overcome this aspect we are working on a novel freehand gesture recognition algorithm that will continue the work begun in [5].

## 4. Conclusions and Future Work

The LEAP motion controller is a powerful device to develop hand gesture interfaces. Its physical and technical features make it an ideal tool to interact with any kind of desktop application. The new version of the LEAP motion API equips the device with a hand skeletal model that provides tracking data with a high level of accuracy. Finally, their model allows the device to predict the position of fingers and hands even if they are partially occluded.

This paper describes an ongoing research work to define and recognize hand gestures. The proposed method adopts freehand drawing recognition algorithms to interpret the center of the palm and fingertip movements. Although the approach is applicable to any hand skeletal model, that provided by the LEAP motion controller satisfies each technical requirement. Preliminary tests have highlighted that the method identifies univocally hand gestures.

Currently, we are working on different application fields. In the first one, we are adopting the proposed method to interpret the body and arm model provided from other devices (Microsoft Kinect [12] and MYO [16], respectively). In the second one, we are defining novel body and hand models to support Self-Avatars (SAs) [6], and Virtual Gloves (VGs) based applications [18, 19]. Finally, we are exploring the possibility to use multiple devices to obtain a reliable skeletal hand model without prediction mechanisms.

## References

- [1] C. Alvarado and R. Davis. Sketchread: A multi-domain sketch recognition engine. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, pages 23–32, New York, NY, USA, 2004. ACM.
- [2] D. Avola, P. Bottoni, A. Dafinei, and A. Labella. Color-based recognition of gesture-traced 2d symbols. In *DMS*, pages 5–6, 2011.
- [3] D. Avola, P. Bottoni, A. Dafinei, and A. Labella. Fcbd: An agent-based architecture to support sketch recognition interfaces. In *DMS*, pages 295–300, 2011.
- [4] D. Avola, L. Cinque, and G. Placidi. Sketchspore: A sketch based domain separation and recognition system for interactive interfaces. In A. Petrosino, editor, *Image Analysis and Processing ICIAP 2013*, volume 8157 of *Lecture Notes in Computer Science*, pages 181–190. Springer Berlin Heidelberg, 2013.
- [5] D. Avola, A. Del Buono, G. Gianforme, S. Paolozzi, and R. Wang. Sketchml a representation language for novel sketch recognition approach. In *Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments*, PETRA '09, pages 31:1–31:8, NY, USA, 2009. ACM.
- [6] D. Avola, M. Spezialetti, and G. Placidi. Design of an efficient framework for fast prototyping of customized human-computer interfaces and virtual environments for rehabilitation. *Comput. Methods Prog. Biomed.*, 110(3):490–502, June 2013.
- [7] S. Berman and H. Stern. Sensors for gesture recognition systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 42(3):277–290, May 2012.
- [8] G. Casella, V. Deufemia, V. Mascardi, G. Costagliola, and M. Martelli. An agent-based framework for sketched symbol interpretation. *J. Vis. Lang. Comput.*, 19(2):225–257, Apr. 2008.
- [9] M. J. Fonseca and J. A. Jorge. Experimental evaluation of an on-line scribble recognizer. *Pattern Recognition Letters*, 22(12):1311 – 1319, 2001. Selected Papers from the 11th Portuguese Conference on Pattern Recognition - {RECPAD2000}.
- [10] T. Hammond and R. Davis. Ladder: A language to describe drawing, display, and editing in sketch recognition. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH '06, New York, NY, USA, 2006. ACM.
- [11] L. B. Kara and T. F. Stahovich. An image-based, trainable symbol recognizer for hand-drawn sketches. *Computers & Graphics*, 29(4):501–517, Aug. 2005.
- [12] Kinect. <http://www.xbox.com/it-it/kinect>, 2014.
- [13] LEAPMotion. <https://www.leapmotion.com/>, 2014.
- [14] LEAPMotionAPI. <https://developer.leapmotion.com/>, 2014.
- [15] T. B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2):90–126, 2006.
- [16] MYO. <https://www.thalmic.com/en/myo/>, 2014.
- [17] I. Oikonomidis, N. Kyriazis, and A. Argyros. Markerless and efficient 26-dof hand pose recovery. In R. Kimmel, R. Klette, and A. Sugimoto, editors, *Computer Vision ACCV 2010*, volume 6494 of *Lecture Notes in Computer Science*, pages 744–757. Springer Berlin Heidelberg, 2011.
- [18] G. Placidi. A smart virtual glove for the hand telerehabilitation. *Comput. Biol. Med.*, 37(8):1100–1107, Aug. 2007.
- [19] G. Placidi, D. Avola, D. Iacoviello, and L. Cinque. Overall design and implementation of the virtual glove. *Comp. Biol. Med.*, 43(11):1927–1940, Nov. 2013.
- [20] D. Tang, T.-H. Yu, and T.-K. Kim. Real-time articulated hand pose estimation using semi-supervised transductive regression forests. *Comp. Vis., IEEE International Conference on*, 0:3224–3231, 2013.
- [21] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5):6380–6393, 2013.
- [22] Y. Wu and T. S. Huang. Vision-based gesture recognition: A review. In *Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*, GW '99, pages 103–115, London, UK, UK, 1999. Springer-Verlag.



# Intermodal Image-Based Recognition of Planar Kinematic Mechanisms

Matthew Eicholtz, Levent Burak Kara  
Department of Mechanical Engineering  
Carnegie Mellon University  
Pittsburgh, PA USA  
meicholt@andrew.cmu.edu, lkara@cmu.edu

**Abstract**—We present a data-driven exploratory study to investigate whether trained object detectors generalize well to test images from a different modality. We focus on the domain of planar kinematic mechanisms, which can be viewed as a set of rigid bodies connected by joints, and use textbook graphics and images of hand-drawn sketches as input modalities. The goal of our algorithm is to automatically recognize the underlying mechanical structure shown in an input image by leveraging well-known computer vision methods for object recognition with the optimizing power of multiobjective evolutionary algorithms. Taking a raw image as input, we detect pin joints using local feature descriptors in a support vector machine framework. Improving upon previous work, detection confidence depends on multiple context-based classifiers of varying image patch size and greedy foreground extraction. The likelihood of rigid body connections is approximated using normalized geodesic time, and NSGA-II is used to evolve optimal mechanisms using this data. The present work is motivated by the observation that textbook diagrams and hand-drawn sketches of mechanisms exhibit similar object structure, yet have different visual characteristics. We apply our method using various combinations of images for training and testing, and the results demonstrate a trade-off between solvability and accuracy.

**Keywords**—computer vision; evolutionary multiobjective optimization; kinematic simulation; object recognition

## I. INTRODUCTION

The design of complex mechanical linkages is a challenging task involving the coordination of multiple rigid bodies to achieve a desired dynamic profile (see Fig. 1 for examples). The ability to visualize the kinematics of a mechanism is a valuable skill to improve mechanical intuition during design analysis and synthesis [1], yet current simulation tools may be insufficient for fast kinematic visualization. Currently, engineers will likely resort to one of three options. First, they may use mental simulations to infer mechanical behavior [2], but this is ineffective for people with low spatial ability [3] and is generally difficult for complex mechanisms [4]. Second, specialized software [5-6] may be used for simulations, but this task is often too time-consuming to be practical (e.g. students solving a dynamics homework problem, professional engineers brainstorming potential design concepts) and may require advanced programming skills, which hinders novice users. Third, engineers often use hand-drawn sketches

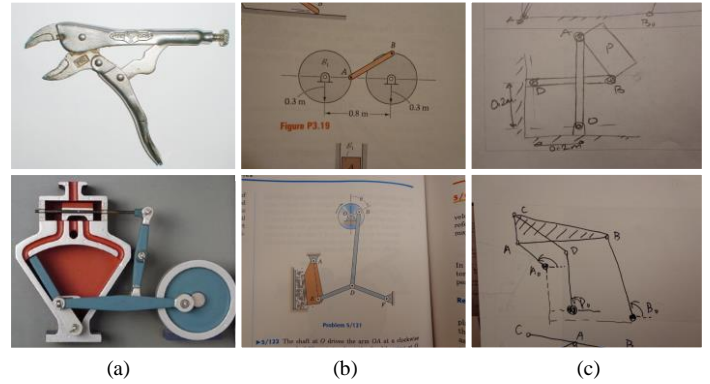


Figure 1. Example mechanisms in (a) natural images of real-world objects, (b) textbook graphics, and (c) hand-drawn sketches. Each mechanism contains a set of rigid bodies connected by kinematic pairs (e.g. revolute or prismatic joints) that constrain their motion. The present work focuses on automatically recognizing the number, location, and connectivity of joints in textbook graphics and hand-drawn sketches; this information is all that is required to fully specify the allowable motion of each rigid body.

to convey design ideas and visualize dynamic properties, perhaps abstracting the mechanism to a simpler form or using key annotations and arrows to demonstrate motion.

In a previous work [7], we developed an algorithm to bridge the gap between ineffective mental simulations and impractical computer simulations by automatically recognizing the underlying mechanical structure in a single image. At the heart of our approach was a novel combination of vision-based object recognition with multiobjective evolutionary optimization. The fundamental principle of the method was to consider mechanisms as a collection of connected joints, where each pairwise joint connection indicated that two joints were fixed to the same rigid body. We limited our study to planar mechanisms, in which the motion of every rigid body is constrained to the plane perpendicular to the viewer, and only considered examples made up entirely of revolute joints. With this representation, the task involved locating probable joints in an image using a sliding window object detector, assessing the likelihood of all pairwise joint connections using normalized geodesic time and maximizing image consistency and mechanical feasibility using the NSGA-II algorithm. The algorithm enabled the evolution of a small set of feasible mechanical structures based on local features in a single image, and only required a set of training images for joint detection.

We initially implemented the approach on textbook graphics due to their relative simplicity and wide availability.

In the present work, outlined in Fig. 2, we shift our focus to include sketches as valid input data to our algorithm. This is motivated by the idea that sketches are more directly related to design synthesis than textbook graphics. Someone creating a new mechanism may not be able to find a clean image depicting their design concept; indeed, they may not even know what they are looking for yet. With our technology, we hope to enable users to rapidly explore the design space using pencil and paper without being encumbered by existing designs.

We represent sketch data as an image, so that no modifications to the original algorithm are explicitly required to accommodate the new input modality. Regardless, we propose a couple key enhancements to the joint detection scheme in order to boost performance; details are provided later in this paper. Despite being of the same “form” as the textbook images used previously, we still consider sketches to be from a different *modality* because they were created in a different manner than textbook graphics. The evidence in support of this proposition is clear from the examples pictured in Fig. 1. Textbook graphics use consistent shapes, colors, and textures, while sketches are typically messier, have curvier lines, and include artifacts such as overtracing, tonal variation in stroke intensities, and cross hatching, among others [8]. Furthermore, depictions of mechanisms in textbooks may be surrounded by irrelevant text, annotations, highlighting, or other mechanisms that clutter the image; sketches, on the other hand, can be created without such distracting visual elements.

Even though they may be strikingly different in certain visual characteristics, textbook graphics and sketches of mechanisms adhere to the same structural principles. This poses an interesting problem: can we successfully use one input modality for training and the other for testing? More specifically, are we required to have a set of training sketches in order to correctly recognize test sketches of mechanisms? The answer may have important implications for future tools involving the recognition of visual objects with different input modalities.

The remainder of the paper is structured as follows: section II highlights related work in sketch recognition, computer vision, and evolutionary algorithms. Improvements made to our original algorithm are provided in section III. Experimental methods, including results and discussions, are given in section IV, followed by concluding remarks in section V.

## II. RELATED WORK

### A. Object Detection

Object detection is a mature field of research in computer vision, spanning countless real-world applications. A typical object detector extracts salient features from sample images, learns a discriminative model from those features, and then scans test images using the model to locate instances of the object. Arguably the most critical step in developing a detection algorithm is feature selection. There are many well-known feature descriptors with reported success [9-11]; in the

present work, locally normalized histograms of oriented gradients (HOG) over a grid of regions in the image are used. We follow the method outlined in the original work [9], which includes training a soft linear support vector machine (SVM) and mining hard negatives from sample images for subsequent re-training. We selected the HOG descriptor because it is a popular, dense, local feature set that has been successful for detecting various objects. However, our algorithm is not dependent on this choice; any feature descriptor and classifier can be incorporated into the overall recognition pipeline.

To our knowledge, kinematic mechanisms are a novel domain for object recognition. However, there is a breadth of ongoing research in recognizing similar objects comprising structured parts. Practical applications include face recognition [12], pose estimation [13], and 3D surface estimation [14]. The key difference, though, between previous work in this area and our present domain is that mechanisms do not have well-defined structural or spatial dependencies. For example, in face recognition, it is straightforward to learn that the forehead is not located below the mouth or that a nose should exist between the eyes; with kinematic mechanisms, it is less clear if a specific joint should be connected to another. Little knowledge is gained about the likelihood of other objects in the image just from knowing one object’s location.

Object recognition across multiple modalities is a less well understood problem in computer vision. The most relevant works relate to face photo-sketch recognition [15-17], which attempts to match hand-drawn sketches of faces with samples in an image database. Various methods are used to find discriminating features between the two modalities; some even transform one modality to another (e.g. convert all photos to pencil sketches) in an effort to reduce the variance among the dataset. Our present work, by contrast, must not only recognize an object (mechanism) across modalities, but also should detect parts (joints) that make up the object across modalities.

### B. Sketch Recognition

There are two important aspects of sketch recognition that relate to the present work: representation and complexity. With regard to representation, two classes of techniques have emerged in the literature. *Stroke-based* methods treat each sketch as a sequence of time-stamped strokes, each containing a series of sample points in space. While some works share similarities to our domain [18-23], stroke-based methods are ill-suited for our recognition framework, which was designed to work on images. Still, there are interesting parallels; for instance, [22] uses a graph representation to combine “low-level primitives into high-level shapes using geometrical rules”. We also implement graphs in our recognition pipeline, but instead connect low-level joints to form high-level mechanisms based (partially) on mechanical feasibility rules. The other class of sketch recognition techniques is *image-based* approaches, including the present work, which neglect temporal information and only consider the spatial layout of pixels. This poses the additional challenge of grouping relevant pixels, depending on the object being recognized. With regard to sketch complexity, it is important to distinguish between isolated symbol recognizers and detecting objects in freehand sketches, which is a more challenging problem. The task of

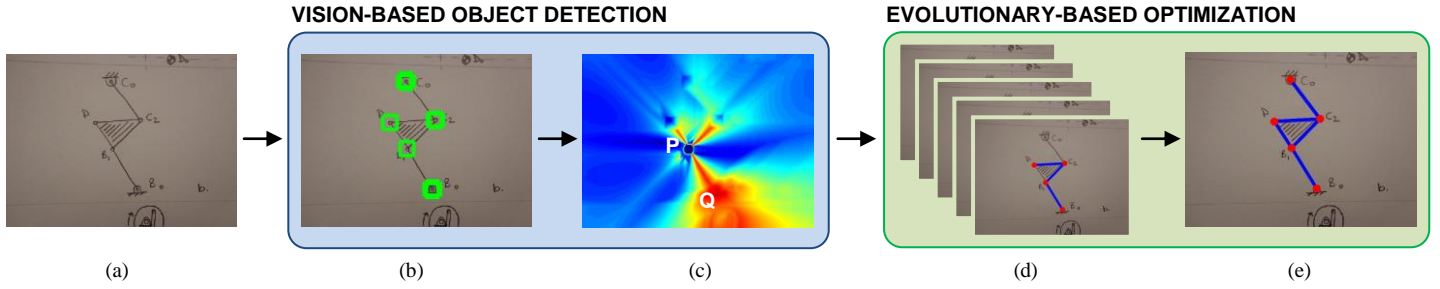


Figure 2. Overview of recognition pipeline. (a) The system takes as input a raw image. (b) To recognize the underlying mechanical structure, the first step is to detect all pin joints in the image; here, line thickness correlates positively with detection confidence. (c) Next, we compute the likelihood that pairwise detected joints are on the same rigid body using normalized geodesic time. In effect, this metric looks for short paths through dark regions in the image. Since there is a dark line connecting point P to Q in the original image, the normalized geodesic time from P to Q is high (indicated by the red color around Q); refer to [7] for more details. (d) Finally, the resulting data is optimized in an evolutionary framework, evolving a set of solutions (mechanisms) using conventional genetic operators and hopefully finding (e) the true solution over time.

symbol recognition can be treated as a template matching problem; some examples of successful approaches in this area include [23-29]. In some sense, the joint recognition algorithm used here is similar to a sliding window symbol recognizer. However, we do not use part templates and instead learn a discriminative model based on local image features.

It is widely agreed that robust sketch recognition algorithms require a large corpus of training images. Yet, acquiring such a large number of sample sketches can be a tedious task. One recent work [30] demonstrated the ability to automatically generate synthetic images from a small set of labeled examples. Another [31] investigates the effect of using isolated symbols for training a recognizer designed for freehand sketches. Our work contributes to this area by hypothesizing that sketch images may not be necessary for training if examples from another modality are more readily-available.

### C. Evolutionary Multiobjective Optimization

Multiobjective evolutionary algorithms (MOEAs) are widely used in real-world applications that require optimization of several, often conflicting, objectives (fitness criteria). MOEAs operate by stochastically sampling the search space of candidate solutions and iteratively applying genetic operators such as crossover and mutation to evolve optimal solutions. To handle conflicting objectives, for which there is no single optimal solution, many MOEAs use the idea of Pareto dominance to rank solutions [32-34]. An individual solution is said to dominate another solution if it is at least as good for all objectives and better (more fit) for at least one objective. The Pareto front is defined as the set of all nondominated solutions. The algorithm selected for our approach, called the nondominated sorting genetic algorithm, was first introduced two decades ago (NSGA [32]) and improved several years later (NSGA-II [33]). We use the latter version, which is characterized by fast computation of nondominated sorting and inclusion of crowding distance to preserve diversity and showed promising results in prior work [7].

For the present domain, the feasibility of a predicted mechanism is governed by mechanical principles. These principles can be formulated as a series of constraints; in this way, large regions of the search space may become infeasible because one or more of the constraints fail. A critical step in

MOEA design is determining how to handle such constraints. Constraint handling methods can be broadly categorized into two groups: (i) those that always prefer feasible solutions (hard constraints) [33,35] and (ii) those that treat constraints as objectives (soft constraints) [36]. We employ the latter method in order to allow infeasible, yet strong, solutions to persist because they may be near the constraint boundaries.

## III. TECHNICAL APPROACH

The proposed framework for mechanism identification in images from various modalities largely relies on work previously developed in [7]. In this section, we provide a brief overview of the algorithm pipeline, followed by detailed descriptions of key modifications made to the original work. Unless stated otherwise, we use the same methods and parameters as [7].

### A. Overview

The recognition framework (Fig. 2) has two primary stages: (i) vision-based detection of mechanical components, and (ii) evolutionary-based optimization of the mechanism structure. The algorithm was developed to be general in nature; any image type is a valid input to the system, any feature descriptor and classification method can be used to detect joints, any metric can be used to compute pairwise joint connection likelihood, and any genotype representation and genetic operators can be tested in the evolutionary algorithm. A basic outline of the algorithm is listed below; recent improvements are highlighted and will be discussed in the following sections.

### B. Using Multiple Context-Based Classifiers

Previously, a fixed-window SVM classifier was used to detect likely pin joints in an image. The recall was generally high (i.e. very few false negatives), but the precision was sometimes low (i.e. too many false positives). Furthermore, the evolutionary algorithm does not optimize joints based on a simple binary decision; instead, it relies on the strength of classification, which we define as the distance to the SVM decision boundary. With this in mind, it should be clear to see that even a high-precision, high-recall classifier can be problematic for the optimization routine if even one false positive in an image has strong confidence. Also, previous

---

**Algorithm 1 – Main**

---

Pre-training:

1. Acquire sample images containing planar kinematic mechanisms.
2. Manually label all pin joints and pairwise joint connections in each image.
3. Separate data into training and testing sets.

Training:

1. Extract positive examples of pin joints in training images.
2. Augment positive examples by reflecting image patches about vertical/horizontal axes and rotating by {90,180,270} degrees.
3. Extract random negative examples from training images. Use tolerance of 32 pixels to ensure negative patches do not contain pin joints.
4. Compute HOG features for positive and negative image patches.
5. Train a soft ( $C=0.01$ ) linear SVM.
6. Randomly extract 1000 additional patches per training image, classify using initial SVM, and add hard negatives to dataset.
7. Re-train the SVM.
8. Repeat training process using larger window sizes.

Testing:

1. Apply SVM to test image with sliding window of fixed size.
  2. Suppress non-local maxima using mean shift algorithm [35].
  3. Apply multiple classifiers to detected joints to compute weighted confidence. Discard detections with confidence less than zero.
  4. Apply foreground extraction to image. Discard background detections.
  5. Store detected pin joint locations and confidence values.
  6. For all pairs of detected pin joints, compute normalized geodesic time, which indicates the likelihood that those two joints are located on the same rigid body.
  7. Store connection likelihood matrix.
  8. Run NSGA-II using pin joint locations, associated confidence levels, and connection likelihood as input. Fitness evaluation includes image consistency measures and binary constraints for mechanical feasibility. The output is a set of Pareto-optimal solutions.
  9. Discard solutions that are infeasible.
  10. Remove duplicate solutions.
  11. Prioritize the remaining solutions (currently using strength of joint connections).
  12. Locate the ground truth (if applicable).
- 

experiments revealed that execution time strongly depends on chromosome length, which is a function of joint detections. Therefore, it is highly desirable to decrease the number of false positive detections and increase the confidence of true positives relative to false positives.

To address this challenge, we implemented two significant modifications to the detection scheme. First, we incorporated multiple classifiers with varying window size with the idea that larger window sizes would pick up more global context cues regarding true joints. This design decision was primarily motivated by the observation that many false positives demonstrated strong local correlation to pins (e.g. text containing the letter ‘o’), but lacked similarity in a global context (e.g. a pin usually has two rigid bodies emanating from its center, while the letter ‘o’ does not). For the current implementation, we used a root detection window size of 48 pixels, and two additional context classifiers with window sizes of 64 and 80 pixels, respectively. We increase the appropriate HOG descriptor parameters such that all image patches have the same number of features (in this case, 1764). In this manner, the larger classifiers have the same dimensionality, but coarser spatial binning due to increased window size. Contrary to some other approaches involving multi-scale classification, we do not run all classifiers on the full image. Instead, the root classifier is first used to find local maxima, after which patches

centered at those locations are fed into the context classifiers. At that point, every detected pin joint has three values of confidence. A naïve approach might be to define the total confidence as the summation of individual detection strengths. This may yield a poor estimate of true confidence because the SVM decision boundaries do not incorporate normalization. Instead, we propose a weighted sum of confidence, in which the classifier weights,  $w_i$ , are given by,

$$w_i = \left( \frac{\text{Pr}}{\mu} \right)_i \quad (1)$$

where Pr and  $\mu$  are the precision of the classifier on training data and the average distance of true positives to the SVM decision boundary, respectively. Sample results from this improved detection scheme are shown in Fig. 3.

### C. Greedy Foreground Extraction

The cascade of classifiers described above mainly improves the relative confidence of true positives with respect to false positives. To achieve high precision, we implement a greedy unsupervised foreground extraction method and discard any background detections. The approach is outline below.

---

**Algorithm 2 – Foreground Extraction**

---

1. Run Sobel edge detector over image.
  2. Dilate edges by an 8-pixel radius.
  3. Trace boundaries and extract connected regions.
  4. Select the region with the maximum area.
  5. Fill holes in the region.
  6. Save region as foreground.
- 

Despite being a greedy approach, it performs exceptionally well on the images used in our experiments. Accuracy for both textbook images and sketches was 99%. The effect of this algorithm enhancement is depicted in Fig. 4.

## IV. EXPERIMENTS

### A. Data

Two image datasets were utilized in the experiments described in this paper (see Fig. 5 for examples). First, we use the MECH135 dataset [7], which includes 135 images of planar mechanisms from five different textbooks [38-42]. All mechanisms are closed kinematic chains and contain only revolute joints. In addition, we asked 25 engineering students

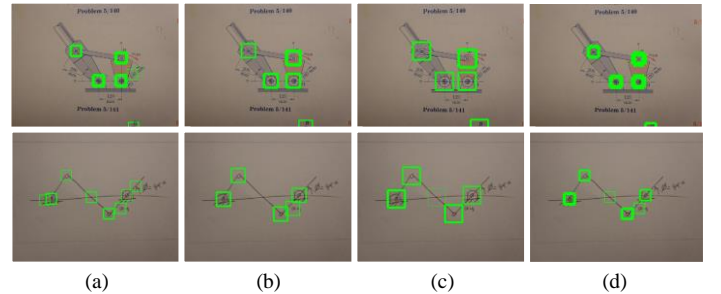


Figure 3. Using multiple context-based classifiers on sample textbook images and sketches. (a-c) Individual classifiers with increasing window size. (d) Final joint detection confidence. Line width positively correlates with confidence.



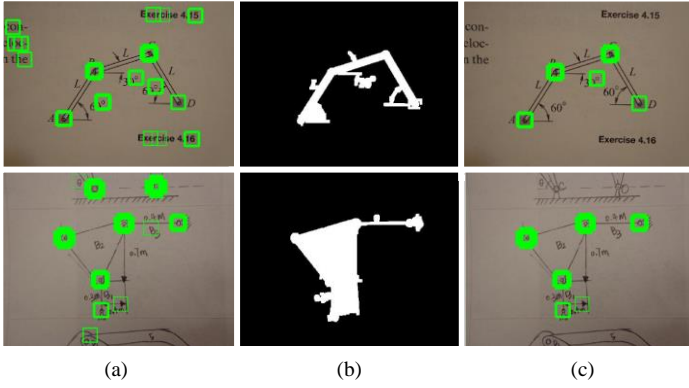


Figure 4. Greedy foreground extraction on sample textbook images and sketches. (a) Weighted sum of multiple classifiers. (b) Binary image showing foreground in white. (c) Joint detections after discarding background instances.

to sketch ten randomly selected mechanisms from the MECH135 dataset on paper. In general, we did not restrict the sketching style nor the level of abstraction implemented by the students, as long as the true underlying mechanical structure was evident in the sketch. Pictures were taken of all hand-drawn sketches, yielding a second dataset of 250 images, approximately two samples per image in the MECH135 dataset. While it is assumed the full mechanism is shown in each image, no explicit restrictions were made regarding position, scale, or orientation of the object. Also, our approach does not require pre-processing of the images (e.g. cropping, filtering), so they may contain noise, illumination changes, and extraneous information such as text, annotations, pencil markings, or partial components from other mechanisms. Ground truth information, including joint location and pairwise connections, was manually provided for all images in both datasets.

### B. Methods

The goal of the experimental studies was to assess the efficacy of our approach on various combinations of training and test images. To that end, six experiments were conducted using all permutations of textbook, sketch, or combined images for training and textbook or sketch images for testing. Each experiment comprised ten separate trials. For each trial, 100 training images and 20 test images were randomly selected, without overlap, from the appropriate datasets. The training images were used to generate a joint detector as described previously and subsequently applied to each test image to locate probable pin joints. Using this information, along with normalized geodesic time for pairwise joint connections, ten independent runs of the evolutionary algorithm were executed per test image, using the general settings listed in Table I. This amounts to 2,000 distinct instances for each experimental condition. Performance metrics of interest include accuracy and speed. Chi-square tests and one-way analysis of variance (ANOVA) were used to determine statistical significance. The implementation was developed in MATLAB [6], and all experiments were performed on an Intel(R) quad-core 3.40 GHz CPU with 8GB RAM.

TABLE I. General NSGA-II Parameters

Parameter	Symbol	Value
population size	$\mu$	$200N^a$
number of offspring	$\lambda$	$\mu$
maximum number of generations	$n$	20
crossover method	—	uniform
crossover probability	$p_c$	0.9
mutation method	—	uniform
mutation probability	$p_m$	0.1
tournament size	$k$	$0.02\mu$

<sup>a</sup>  $N$  refers to number of detected joints

### C. Results and Discussions

Example results on textbook graphics and sketches are shown in Fig. 5. The overall algorithm performance for each experimental condition is summarized in Table II, with the best values highlighted for each statistic. With regard to accuracy, several relevant metrics of success are presented that each contribute to one of two primary objectives: (i) was the true solution found by the evolutionary algorithm and (ii) if so, where in the prioritized Pareto-optimal set of solutions was it located?

An image is deemed *solvable* if the true solution is able to found based on the data input to NSGA-II. Consider the test cases shown in Fig. 6. These instances are unsolvable; the underlying mechanical structure will *never* be correctly identified by our approach because the joint detector failed to locate one or more true joints. In this way, the percentage of solvable test images for an experiment reflects the quality of the joint detector on that dataset. For textbook images, the joint detector always performed relatively well (min. solvable = 88%), and it produced the least number of images with false negative detections when combining textbook and sketches for training (solvable = 94%). This is an interesting result that suggests sketching sample mechanisms may improve detector performance on textbook images. However, this does not guarantee improved performance of the entire algorithm, as evidenced by comparison of the remaining accuracy measures in rows A and E. For sketches, the joint detector did not perform as well, particularly when trained on textbook images (solvable = 47%). This is understandable given the high degree of variance in sketched pins compared to textbook graphics. Still, it is encouraging to note that the number of unsolvable sketch images decreases by more than 50% if sketches are added to the training set and is minimized when only sketches are used for training. A chi-square test revealed that the experimental conditions had a statistically significant effect on joint detector quality and hence the proportion of solvable images in a given test sample,  $\chi^2(5, N = 1,200) = 180.75, p < .001$ .

We classify an image as *solved* if our algorithm was able to correctly identify the underlying mechanical structure in at least one independent run. For both image datasets, a larger fraction of test images was solved when the training images were drawn from the same dataset. Using combined datasets for training does not appear to improve this performance measure. Once again, it should be noted that training on textbooks and testing on sketches resulted in a low number of solved images. The relative difference of solvable and solved

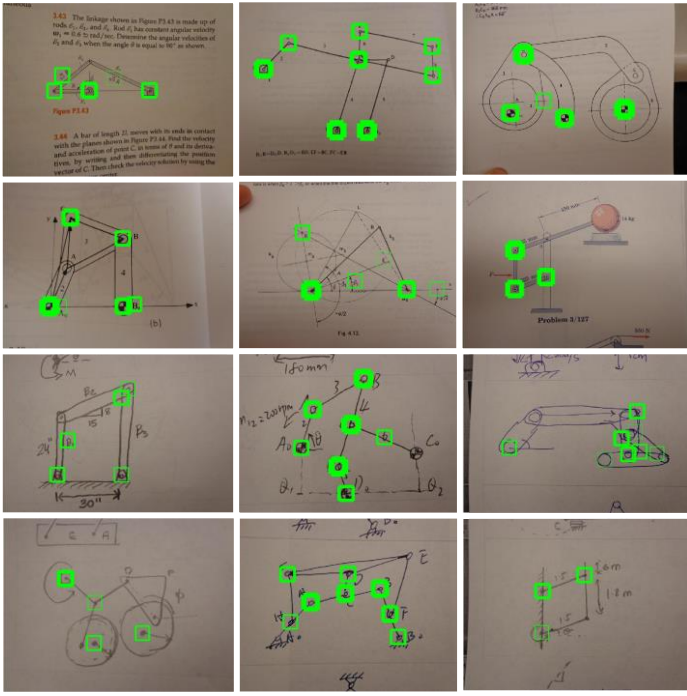


Figure 6. Unsolvable images, due to the presence of false negatives. Bounding boxes indicate pin joint detections, and line width positively correlates with confidence.

images yields the fraction of images that failed due to something other than false negative detections. The mean percentage of solvable, yet unsolved, images across all experiments is 32%. Experiment B had the best performance in this regard, with only one-fifth of solvable images failing to ever produce the correct mechanism, while experiment D was the worst-performing case, with over one-half of solvable images remaining unsolved. The observed differences in the percentage of solved images were found to be statistically significant,  $\chi^2(5, N = 1,200) = 85.70, p < .001$ . Some example failure cases and possible explanations are provided in Fig. 7.

The *overall success rate* (OSR), or the average number of runs in which the true solution was found, exhibits a similar trend to the *solved* metric. With the exception of experiment B, the true mechanism is identified in at least half of the runs. Also, recognition of textbook images is higher than sketch images in general. While overall success rate is a reasonable estimate of algorithm effectiveness for a given set of training and testing images, we suggest it is not the only meaningful metric because it is negatively skewed by unsolved images. With this in mind, we also compute the solved success rate (SSR), which characterizes the reliability of our approach for images that were correctly recognized at least once. The results are somewhat surprising; the most reliable experimental condition is the textbook/sketch case (93% SSR). In other words, if an image of a hand-drawn sketch is solvable, the algorithm presented in this paper is highly likely to correctly identify the pictured mechanism. One plausible explanation is that many sketches are less cluttered than their textbook counterparts; there is limited extraneous information that could be falsely identified by the algorithm and rigid bodies are typically drawn as simple dark lines, which is favored by our

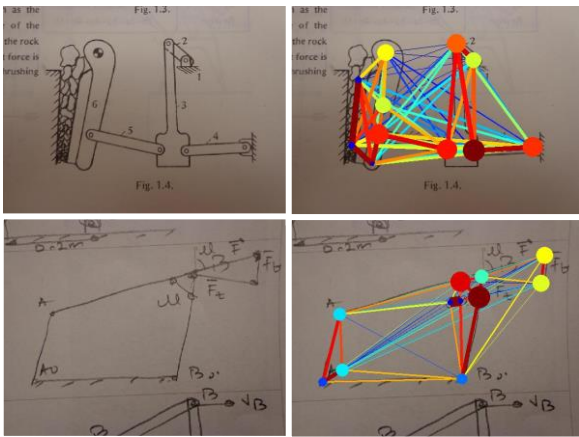


Figure 7. Example failure cases unrelated to false negative detections. (a) Raw images. (b) Visualization of data sent as input to NSGA-II. Higher confidence is indicated by darker (redder) color and thicker shapes. The first instance likely fails because the strongest joint connections are between false positives in the rocks. The primary failure in the second instance is that the strongest joint detections are false positives.

method for predicting pairwise joint connections. On a different note, experiments A, C, and E have nearly identical SSR; perhaps the modality of the training set is less critical when testing textbook images. Lastly, all experimental conditions produced a higher SSR than the previous work in [7], indicating that our modified pin detection method with foreground extraction appears to improve performance. Chi-square tests were performed and indicate there is a relationship between training and testing modalities and algorithm success rate;  $\chi^2(5, N = 12,000) = 492.26, p < .001$  and  $\chi^2(5, N = 7,210) = 39.29, p < .001$  for OSR and SSR, respectively.

While the ability of the evolutionary algorithm to find the true underlying mechanical structure is valuable, perhaps a more important performance measure is *where* the solution was found; that is, can we rank the Pareto-optimal set of solutions in such a way that the true solution has highest priority? The top- $N$  accuracy refers to the percentage of successful runs in which the true solution was at least in the top  $N$  solutions. The obvious desired result is for top-1 accuracy to be high; however, this may not be very realistic. Even one false positive joint detection with confidence higher than any of the true joints will likely allow one or more incorrect solutions to be nondominated by the correct solution. Therefore, we think it is reasonable if the true mechanism is at least in the top 5 solutions generated by NSGA-II. At that point, the best solution could be extracted from this small set either by interactive user selection or feedback from full kinematic simulations. Looking at Table II, there are a few notable highlights to mention. First, experiments with sketches as the test set generally have higher top- $N$  accuracy than similar experiments using textbook images (compare A $\leftrightarrow$ D, B $\leftrightarrow$ C, E $\leftrightarrow$ F). Second, top- $N$  accuracy on test images is lowest when training/testing modality differs and highest when modalities are the same. Finally, the best-performing case from this viewpoint is training on textbooks and testing on sketches, with over half of successful runs yielding the optimal scenario and top-5 accuracy of 84%. All observed differences with regard to



TABLE II. ALGORITHM PERFORMANCE

Experimental Setup			Accuracy (%)							Speed <sup>a</sup> (sec)	
ID	training	testing	<i>solvable</i>	<i>solved</i>	<i>success rate, overall</i>	<i>success rate, solved images</i>	<i>top-1</i>	<i>top-3</i>	<i>top-5</i>	<i>overall</i>	<i>per gene</i>
A	textbook	textbook	89.5	74.5	62.8	84.4	33.3	51.1	61.7	3.34±2.83	0.222±0.066
B	textbook	sketch	46.5	35.0	32.4	92.6	52.2	78.1	84.4	1.64±1.42	0.210±0.022
C	sketch	textbook	88.0	58.0	48.8	84.1	22.7	40.8	50.5	5.44±4.24	0.201±0.006
D	sketch	sketch	81.5	61.5	53.8	87.5	41.8	67.3	74.1	2.97±2.60	0.203±0.008
E	combined	textbook	94.0	73.5	61.9	84.1	29.3	44.9	53.6	4.25±4.61	0.201±0.011
F	combined	sketch	74.0	58.0	50.4	86.9	50.1	69.8	79.9	2.98±2.74	0.206±0.009

<sup>a</sup>. for running NSGA-II on one test image

top- $N$  accuracy were found to be statistically significant;  $\chi^2(5, N = 6,201) = 273.02$ ,  $p < .001$  for top-1 accuracy,  $\chi^2(5, N = 6,201) = 427.08$ ,  $p < .001$  for top-3 accuracy, and  $\chi^2(5, N = 6,201) = 414.30$ ,  $p < .001$  for top-5 accuracy.

An accurate, reliable recognition algorithm is less useful if computationally expensive, so speed is another important performance characteristic to consider. Table II lists the average time it takes to complete a single run of NSGA-II, both overall and per gene; the fastest and slowest experiments are B and C, respectively. A one-way ANOVA showed that the effect of experimental setup on overall runtime was significant,  $F(5, 11874) = 309.87$ ,  $p < .001$ . Post-hoc analysis using Tukey's honestly significant difference (HSD) test demonstrates all pairwise experimental conditions have significantly different mean runtimes, with the exception of D and F. As expected, the time per gene is relatively constant (~200ms), so overall runtime becomes a function of chromosome length. Recall that the number of genes is directly related to the number of detected pins in the image. Thus, comparison of overall times reflects differences in joint detector performance. For example, the optimization is slower when testing textbook images, implying those images have more falsely-detected joints on average than sketches, which is understandable given the extra information (e.g. dimensions, annotations) usually present in textbooks. The computational cost of training the joint detector is overhead and therefore neglected from this analysis. Also, scanning a test image for likely joints and computing pairwise joint connections are independent of experimental conditions, so those metrics are not listed either; total testing time remains on the order of seconds.

## V. CONCLUSIONS

In this paper, we explored the ability of an object detector trained on textbook graphics to positively contribute to the automatic recognition of kinematic mechanisms in images of hand-drawn sketches and vice versa. We improved our previous algorithm by incorporating weighted context cues from multiple classifiers and a greedy foreground extraction technique in the joint detection pipeline. Current experimental studies indicate a trade-off between solvability (whether an image can ever be solved by the evolutionary algorithm) and top- $N$  accuracy (where a solution is found on the Pareto front). Sketches appear more likely to miss true joints, but less likely to be misled by extraneous information and false positives, resulting in high top- $N$  accuracy. All test images benefited from

the inclusion of textbook graphics during training. We think this a powerful idea that could be leveraged to create an intelligent sketch recognition tool to generate kinematic simulation models in a matter of seconds without ever needing a sample sketch for learning.

## REFERENCES

- [1] J. G. De Jalon and E. Bayo. *Kinematic and dynamic simulation of multibody systems*. Springer-Verlag New York, USA, 1994.
- [2] M. Hegarty. Mechanical reasoning by mental simulation. *Trends in cognitive sciences*, 8(6):280-285, 2004.
- [3] M. Hegarty. Mental animation: inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5):1084, 1992.
- [4] M. Hegarty and V. K. Sims. Individual differences in mental animation during mechanical reasoning. *Memory & Cognition*, 22(4):411-430, 1994.
- [5] Adams. *version 2013.2*. MSC Software Corporation, Newport Beach, California, 2010.
- [6] MATLAB, *version 8.1.0.604 (R2013a)*. The MathWorks, Inc., Natick, Massachusetts, 2013.
- [7] M. Eicholtz, L. B. Kara, and J. Lohn. Recognizing planar kinematic mechanisms from a single image using evolutionary computation, *GECCO '14*, July 12-16, 2014, Vancouver, BC, Canada. in press.
- [8] K. Eissen and R. Steur, *Sketching: drawing techniques for product designers*. Bis, 2007.
- [9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition (CVPR), 2005 IEEE Conference on*, volume 1, pp. 886-893, 2005.
- [10] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: Binary robust independent elementary features. In *Computer Vision-ECCV 2010*, pp. 778-792, Springer, 2010.
- [11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis*, 60(2):91-110, 2004.
- [12] A. L. Yuille. Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59-70, 1991.
- [13] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 1385-1392, 2011.
- [14] S. Ross, D. Munoz, M. Hebert, and J. A. Bagnell. Learning message-passing inference machines for structured prediction. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 2737-2744, 2011.
- [15] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *IEEE Trans Pattern Anal Mach Intell*, 31(11):1955-1967, 2009.
- [16] B. F. Klare, Z. Li, and A. K. Jain. Matching forensic sketches to mug shot photos. *IEEE Trans Pattern Anal Mach Intell*, 33(3):639-646, 2011.
- [17] D. Lin and X. Tang. Inter-modality face recognition. In *Computer Vision-ECCV 2006*, pp. 13-26. Springer Berlin Heidelberg, 2006.

- [18] R. Davis. Magic paper: sketch-understanding research. *Computer*, 40(9):34-41, 2007.
- [19] E. J. Peterson, T. F. Stahovich, E. Doi, and C. Alvarado. Grouping strokes into shapes in hand-drawn diagrams. *Proc AAAI Conf on Artificial Intelligence*, 974-979, 2010.
- [20] C. Lee, J. Jordan, T. F. Stahovich, and J. Herold. Newtons Pen II: An intelligent, sketch-based tutoring system and its sketch processing techniques. *Eurographics Symposium on Sketch-Based Interfaces and Modeling*, pp. 9-17, 2012.
- [21] T. Hammond and R. Davis. Tahuti: a geometrical sketch recognition system for UML class diagrams. *Proceedings of the AAAI Spring Symposium*, pp. 59-66, 2002.
- [22] T. Hammond and B. Paulson. Recognizing sketched multistroke primitives. *ACM Trans Interact Intell Syst*, 1(1):1-34, 2011.
- [23] L. B. Kara, L. Gennari, and T. F. Stahovich. A sketch-based tool for analyzing vibratory mechanical systems. *Journal of Mechanical Design*, 130(10), 2008.
- [24] G. Costagliola, M. De Rosa, and V. Fuccella. Recognition and autocompletion of partially drawn symbols by using polar histograms as spatial relation descriptors. *Computers & Graphics*, 39:101-116, 2014.
- [25] L. B. Kara and T. F. Stahovich. An image-based, trainable symbol recognizer for hand-drawn sketches. *Computers & Graphics*, 29(4):501-517, 2005.
- [26] W. Lee, L. B. Kara, and T. F. Stahovich. An efficient graph-based recognizer for hand-drawn symbols. *Computers & Graphics*, 31:554-567, 2007.
- [27] L. Fu and L. B. Kara. Recognizing network-like hand-drawn sketches: a convolutional neural-network approach. *ASME International Design Engineering Technical Conference*, 2009.
- [28] L. Fu and L. B. Kara. From engineering diagrams to engineering models: visual recognition and applications. *Computer-Aided Design*, 43(3):278-292, 2011.
- [29] T. Y. Ouyang and R. Davis. A visual approach to sketched symbol recognition. *IJCAI*, 9:1463-1468, 2009.
- [30] L. Fu and L. B. Kara. (2011). Neural network-based symbol recognition using a few labeled samples. *Computers & Graphics*, Elsevier, Volume 35(5):955-966, 2011.
- [31] M. Field, S. Gordon, E. Peterson, R. Robinson, T. Stahovich, and C. Alvarado. The effect of task on classification accuracy: using gesture recognition techniques in free-sketch recognition. *Computers & Graphics*, 34:499-512, 2010.
- [32] N. Srinivas and K. Deb. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3):221-248, 1994.
- [33] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput*, 6(2):182-197, 2002.
- [34] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE Trans Evol Comput*, 3(4): 257-271, 199.
- [35] C. A. Coello Coello. Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art. *Computer methods in applied mechanics and engineering*, 191(11):1245-1287, 2002.
- [36] Y. G. Woldesenbet, G. G. Yen, and B. G. Tessema. Constraint handling in multiobjective evolutionary optimization. *IEEE Trans Evol Comput*, 13(3):514-525, 2009.
- [37] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell*, 24(5):603-619, 2002.
- [38] J. Ginsberg. *Engineering Dynamics*. Cambridge University Press, New York, New York, 2008.
- [39] D. J. McGill and W. K. King. *Engineering Mechanics: An Introduction to Dynamics*. Tichenor Publishing, Bloomington, Indiana, 2003.
- [40] J. L. Meriam and L. G. Kraige. *Engineering Mechanics, Volume 2*. John Wiley and Sons, Inc., Singapore, 1993.
- [41] J. L. Meriam and L. G. Kraige. *Engineering Mechanics: Dynamics*. John Wiley and Sons, Inc., Hoboken, New Jersey, 2007.
- [42] E. Soylemez. *Mechanisms*. Middle East Technical University, Ankara, Turkey, 1993.



# A Circular Visualization of People's Activities in Distributed Teams

Paolo Buono, Maria Francesca Costabile, Rosa Lanzilotti  
Dipartimento di Informatica, Università degli Studi di Bari Aldo Moro  
Via Orabona, 4 - 70125 Bari  
{paolo.buono, maria.costabile, rosa.lanzilotti}@uniba.it

**Abstract**—When working in distributed teams, it is very important to be aware of the activities of all members, since it provides hints about when they might be available for collaboration. We propose a novel visualization technique that combines several representations to show the daily patterns of team members' activities. It uses a 24 hours circular display to facilitate international collaboration across time zones. Current calendar information can be compared to the typical patterns and reveal likely availability. User studies evaluating the tool that implements the proposed technique are reported and discussed.

**Keywords**—Circular BoxPlot, Circular Histograms, Activity Traces.

## I. INTRODUCTION

The use of technology to allow people collaboration dates back to the sixties [1]. Today the growth of social networks and the increasing number of projects, whose development teams are distributed world-wide, have generated increasing attention to systems and tools that support remote collaboration, in order to replicate co-located collaboration conditions. However, coordination and collaboration among members in distributed teams is affected by several distance factors, which influence effectiveness and/or efficiency of the collaboration. Pallot et al. analyze 25 different distance factors and group them into four dimensions: structural, social, technical and legal. *Lack of interpersonal awareness* is identified as a social distance that influences collaboration effectiveness: each collaborator needs to know what the others are doing and the problems they are facing, in order to be able to organize his/her own contribution to better fit with others' activities [2].

Lack of interpersonal awareness is seen as one of the most problematic factors in software development activities [3]. Distributed team collaboration is mediated by software systems that collect data to keep track of team members' activities in exchanging messages, scheduling meetings, sharing and modifying documents, etc. Such data are called *activity traces*; they can provide hints for identifying when a person might be available for communication.

This paper describes a novel technique that visualizes activity traces of people collaborating in distributed teams. In the example presented, activity traces refer to sent emails, chat messages, keyboard/mouse use and workstation login/logout. Other activity traces could be considered, depending on the context of use. Even if this technique also uses some visual

representations already known, it combines them in an original visualization based on a clock metaphor, in which 24 hours are displayed in order to facilitate international collaboration across time zones. The aim is to foster interpersonal awareness by providing clues about collaborators availability through an overview of the daily activities performed by individuals in a globally distributed team. With respect to previous proposals about visualizing activity traces (e.g. see [5], [7]), a further novel contribution of our technique is that it visualizes activities of several individuals at once, quickly conveying information about their location, their availability at the current time and in the next 24 hours. The visualization thus permits understanding of time relationships among team members and the identification of patterns occurring during daily activities.

Initial ideas about this visualization technique were presented through a poster at a recent conference [4], without any focus on the motivation and on all the features of the current visualization tool. In particular, a feature not already available in the prototype described in [4] is the visualization of the team members' planned schedule in the represented 24 hours. This information is useful since it provides further clues about collaborators' availability.

This paper discusses the rationale and illustrates all the novel features of the proposed visualization technique, also reporting the performed evaluation studies. It has the following organization. Related works are reported in Section II. The metaphor and the visual representations adopted by the technique and an example of its usage are described in Section III. The user-centred design and the evaluation studies carried out with the involvement of users and domain experts, are reported in Section IV. Finally, Section V concludes the paper.

## II. RELATED WORK

The literature reports several proposals of visualizing activities of collaborative teams. We distinguish two categories of papers: those that focus on visualizing the structure of teams and those that focus on activities of a single member. Papers in the first category highlight temporal and social structures of the communications among members, while the others visualize activity traces, in order to help a member in planning his/her own activities and communication.

An example of visualization in the first category is provided by Fisher and Dourish, who represent the structure of a team through a graph, whose nodes are the members and links are

the emails exchanged among them [5]. The visualization allows users to make assumptions about the roles of the members of the team (social structure). The paper also proposes a tabular visualization, which shows the top 10 collaborators, ordered according to the number of messages that a specific member sent to the others, grouping the data by week or by month. This provides hints about the interaction frequency between the user and the team members shown in the list.

Another example of visualizing distributed teams through graphs is provided in [6]. Specifically, authors visualize directed graphs in which nodes represent users locations, indicated on a geographic map, and links represent different relationships, like contributions to the same project. They also show small multiple displays combining multiple views in order to enable visual comparisons, and matrix diagrams that should reveal specific relationships. Such visualization helps understanding the social structure of the team, but does not provide any indication about members' availability for collaboration in the near future.

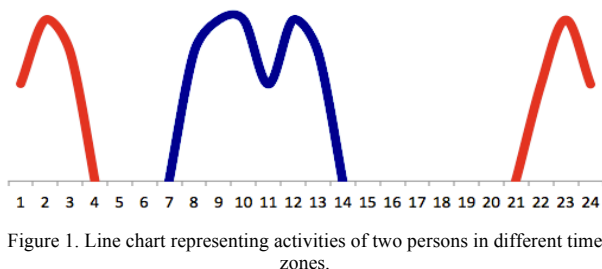


Figure 1. Line chart representing activities of two persons in different time zones.

In the second category, Begole et al. visualize activity of a single member and try to analyze some activity temporal patterns, also called rhythms [7]. They present a number of visualizations of the data that exhibit meaningful patterns in users activities. Examples of monitored activities are: computer activity, appointment span, home activity, and office activity. A line chart is used to show the average of the data about an individual's activities over a time period. It is easy to show that a linear representation is not very effective, since it fails to give the continuity of the activities along the time, especially when considering different time zones. An example, created with MS Excel, is shown in Fig. 1: the line chart represents activities of two persons (one in blue and one in red) that, starting at 7 in the morning, work 8 hours each and live in time zones that differ 11 hours. The activities of the person (red line) who works from 21 (9 PM) till 4 AM the next day is divided in two parts, which makes more difficult its interpretation. The visualization technique we propose uses a clock metaphor, which works well in visualizing daily activities of team members working at different time zones. The shape of the visualized "clock" is the one of an analogical 24 hours clock, in order to represent the whole day.

In literature, there are a few articles that somehow compare linear versus circular representations (e.g. [1], [18]). They remark that the choice between linear and circular depends very much on the user's task, which is the reason we choose the circular representation. Indeed, Kessel highlights that, even if in most cases people tend to use linear representations, they

actually prefer circular representation for cyclic events, e.g. natural processes, seasons and, in particular, human activities along the 24 hours of the day [1].

The clock metaphor was used in visualizations proposed for other application domains. An example is *spiralClock* [8], which shows events planned for a conference and displays the time along a spiral. This would not work in our case, since, in order to show activities of more people, the visualization of multiple staked spirals becomes soon crowded.

*MarkerClock* is a clock-based visualization proposed in a context of caring for older adults, which uses circular stripes [9]. It shows only 12 hours, but the very limitation is that only one data type can be visualized, while our visualization display traces of multiple types of activity.

Our technique uses circular visualization in order to make the continuity and periodicity of time intuitive. A survey on circular methods for information visualization is provided in [10]. However, among the many circular visualizations proposed so far in different domains and for different goals, very few are oriented to the visualization of daily activities that explicitly visualize the time of the day. In particular, Zhao et al. propose *pieTime*, a pie visualization that shows activities like email and phone calls, visualized on different time scales, in order to reveal temporal activities in user behavior [11].

With the purpose of exchanging information about people's availability and increasing their interpersonal awareness, shared online calendars, such as Google calendar, could be used, since they permit management and coordination of events with other people. However, it has been shown that they fail in this purpose, because often people's calendars are just time-planning and the planned schedule might substantially differ from what actually happens. Lovett et al. remarked that calendars do not represent reality well, since genuine events are hidden by a multitude of reminders and placeholders [12]. In their research, they found that most calendar events (49%) were used as personal reminders (e.g. a reminder to backup files), and another relevant quantity of events (32%) were used as placeholders, (e.g. recurring daily meetings), with no evidence that certain events actually occurred. They also suggested that the reliability of calendar representation of events can be improved by taking into account other types of data, e.g. data about event locations or data coming from other sources, like data from social networks. Other authors employ predictive user models of event attendance and text mining of event descriptions to improve the reliability of shared events [13]. For the purpose of the visualization, we are mainly considering how to represent this information in order to provide clues about collaborators' availability.

In order to contact a person, it is useful to know his/her current status, which is common in today's most well-known computer mediated communication systems. For example, Skype indicates four alternatives: Online, Busy, Away, Offline. This indication is unreliable since people often forget to change their status. Various solutions try to overcome this drawback. The one proposed by MyVine is to model availability for communication by using laptop microphones to sense nearby speech, combined with location sensor data, computer and

calendar information [15]. Our tool displays both status and activities, as described in the next section.

### III. THE CLOCKBOXPLOT (CBP) TECHNIQUE

The goal of the visualization technique described in this paper is to provide an overview of typical activities performed by members in a team, in order to give information about their availability. To this aim, two principal types of information are visualized: a) daily activity traces; b) planned activities in the next 24 hours according to shared calendars.

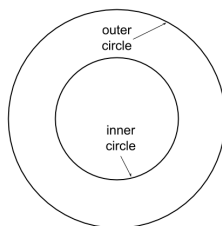


Figure 2. A stripe.

The technique uses a basic visual representation called *circular stripe* (or simply *stripe*). As shown in Fig. 2, it is inspired by the mathematical concept of *annulus*, i.e. it is the area between two concentric circumferences. It is also called *doughnut* or *ring*. A stripe is the area in which activity traces related to an individual are displayed, as described below.

The technique is called ClockBoxPlot (CBP in the rest of the paper), since it exploits a clock metaphor and represents, in a compact way, activity traces through the *boxplot* representation.

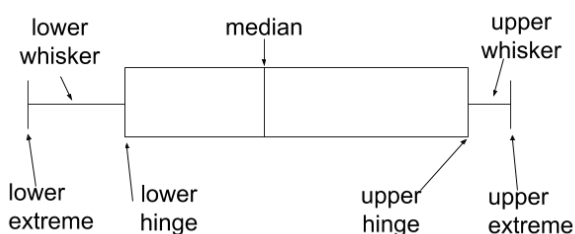


Figure 3. Tukey's boxplot.

John Tukey introduced the boxplot, also called box and whiskers plot, as part of his toolkit for Exploratory Data Analysis [16]. A boxplot provides a summary of the distribution of data, ordered by their frequency. The main component is the *box*, whose size represents data between the lower and upper quantile, called *lower hinge* and *upper hinge* respectively (see Fig. 3). Two *wiskers* connect hinges to extremes that are out of the box: they represent the values with the lowest and the highest frequencies, respectively (lower extreme and upper extreme). The length of the lower whisker represents the distance between the value with the lowest frequency and the lowest quantile. The *median* of the data is

indicated in the box. In addition, potential *outliers* are represented outside the boxplot (not shown in Fig. 3).

Boxplots provide compact representations of data by displaying less details than histograms, but taking less space. They are useful for comparing distributions across different groups of data. The boxplot used in CBP does not show the median and the outliers. Moreover, because it is visualized within a stripe, a circular boxplot is used, as shown in Fig. 4.

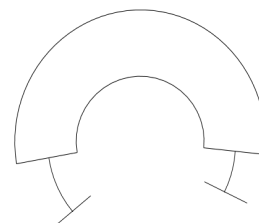


Figure 4. Circular boxplot.

Histograms have been added to CBP in order to visualize, in more details with respect to boxplot, data referring to a specific activity of a person. The histograms are drawn inside the stripe, thus they are visualized as circular histograms, whose bins are displayed along a circumference. Since a stripe is delimited by two concentric circumferences, each stripe can contain two histograms, one with bars placed outside the inner circle and one with bars placed inside the outer circle. Fig. 5 shows the two histograms drawn in a stripe.

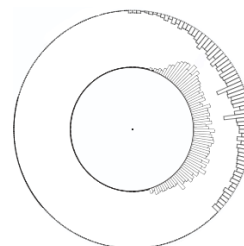


Figure 5. Histograms within a stripe.

Fig. 6 is a screenshot of the ClockBoxPlot tool that implements the CBP visualization technique. The external circle has 24 hours indicated on it. In this example, six stripes are visualized. The innermost circle shows the current time (16:24) and user location (Bari), also indicating the time zone expressed with reference to the Greenwich Meridian Time (GMT +1). Each stripe refers to a team member and, by using two histograms (one in purple and one in blue) and one boxplot (in yellow), it represents activities of that member in the 24 hours after the current time (16:24). The boxplot represents different data about a person's activities on a computer, such as keyboard stroke frequency and login/logout timestamps. It thus indicates the time intervals in which a person is typically active on his/her computer. By taking into account the structure of the boxplot, in the time intervals corresponding to whiskers, the probability to find the person available is lower.



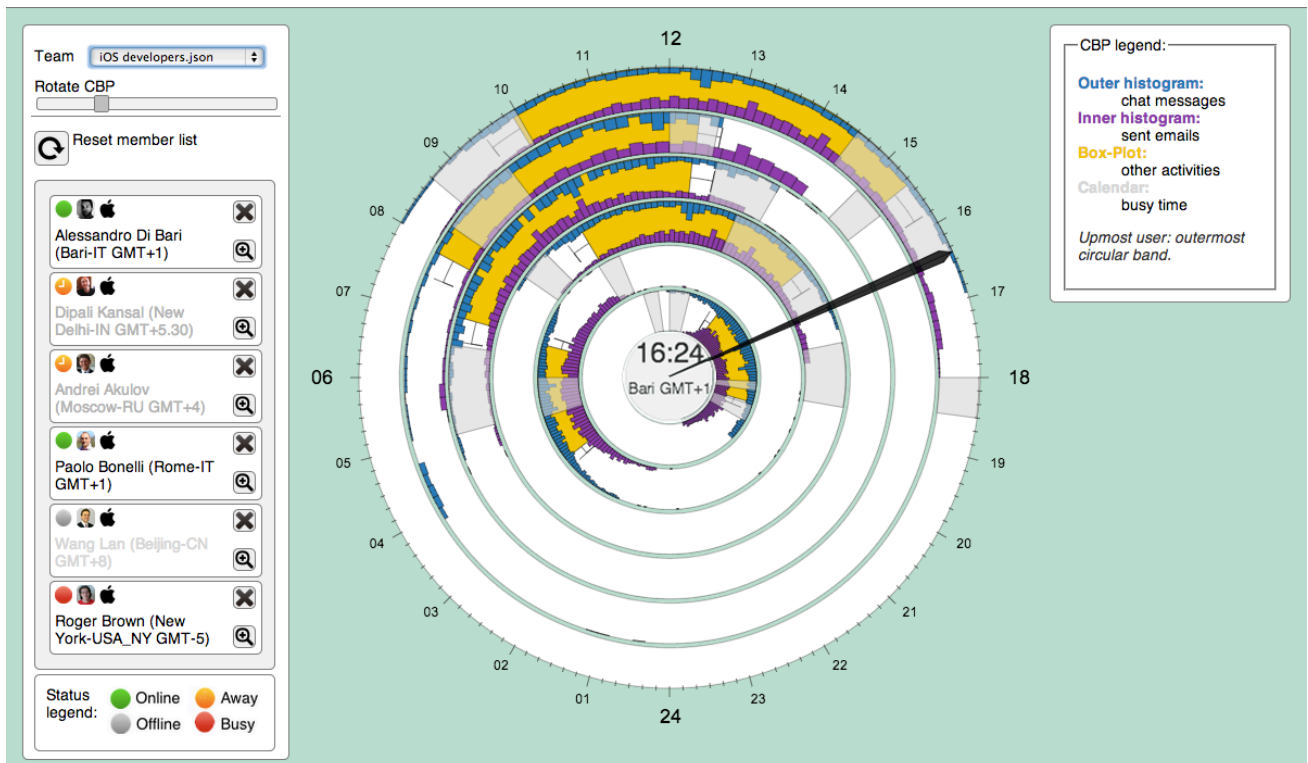


Figure 6. A screenshot of the CBP tool. Noon (12) is at the top and midnight (24) is at the bottom. The activities of iOS developers in a team of six are visualized, starting from the time shown at the center (16:24) up to the next 24 hours. The activities of the developer at the top of the list are represented in the outmost stripe, the others are represented in the more internal stripes according to the list order. In the member list, the names of three developers currently away from the computer or offline are grayed. Gray zones in a stripe show person unavailability according to his/her event calendar.

In the example in Fig. 6, the blue histogram shows frequencies of chat messages that the specific member has sent to other team members, while the purple histogram shows frequencies of email messages. Other activity traces could be available; in this example, they are system login/logout and key press frequencies. The CBP tool allows the user to select those activities that he/she wants to visualize in more details, which will be shown by the two histograms. The remaining activity traces are visualized with the boxplot, which thus represents the probability that the collaborator will be available for communication on the basis of the other collaboration traces. This probability is computed by using algorithms like those in [7].

Since the objective of the visualization is to give information about people's availability, it is useful to add data coming from shared calendars, in order to indicate time intervals in which a person would likely be busy. For privacy reasons, it is more likely that a person can make available calendar events without specifying the nature of such events but only their time intervals. Such intervals in which the person will be busy are grayed in the stripe.

Even if the time at the user location is indicated at the center, the CBP also uses a black hand pointing at the user's time on the clock, in order to make easier to analyze the activities visualized in the different stripes starting from that time. The activities are actually visualized from the time

indicated in the center (i.e. starting from the black hand) up to the next 24 hours.

The example in Fig. 6 shows six stripes, each one corresponding to one member listed in the panel on the left of the screen, where other information is provided. Specifically, one of them lives in Bari, one in New Delhi, one in Moscow, one in Rome, one in Beijing, one in New York. It is also indicated the time zone with reference to GMT. The order of the names in the list corresponds to the stripes going from the outmost to the inmost stripe. Thus, the outmost stripe represents the activities of "Alessandro Di Bari". The yellow boxplot spans from 9:55 to 15:25; one whisker is set at time 9:35 and the other one at time 15:50. This means that the probability that this specific member will be available is higher between 9:55 and 15:25, it is lower between 9:35 and 9:55 and between 15:25 and 15:50, and almost zero in the remaining time. Furthermore, "Alessandro Di Bari" has some events scheduled between 8:40 and 10:00 and between 14:30 and 16:24 of the next day. Actually, we cannot know if he will be available after 16:24, because the shown time interval is 24 hours. In the shown example, it is evident that the four most outside members can easily find time to work together, but for Lan and Brown (the two inmost stripes) it is harder to meet.

Often a user needs to compare the activities of two collaborators. This is much easier if the corresponding stripes are visualized one next to the other. The user can change the

stripe order by a simple drag and drop. Several users that tried CBP found this feature very useful, not only for comparison purposes, but also to better observe activities of a member, because more external stripes are bigger than more internal ones. In addition, a stripe can be zoomed by clicking the lens icon of the corresponding member in the users panel. For example, Fig. 7 shows the same CBP of Fig. 6 with the outermost stripe zoomed.

Looking at Fig. 6, one may wonder about scalability issues. Our experience shows that even if many people participate in teams, the actual collaboration is often limited to a small number of them, not greater than a dozen. When the list of members does not fit in a screen, a scrollbar appears at the left side of the users panel. Moreover, the user not interested in analyzing certain members can temporarily remove them from the visualization by clicking on the X icon in the member panel (see Fig. 6). The user can come back to the complete member list by clicking on the 'reset' icon on top of the member list.

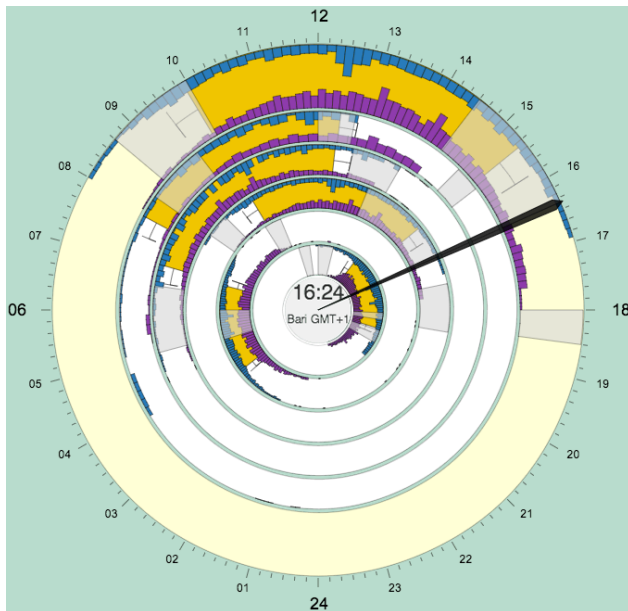


Figure 7. Outermost stripe zoomed in order to better observe it. A click on the stripe or on a member in the list highlights both the member and the associated stripe.

#### IV. USER-CENTRED DESIGN OF CBP

The tool has been developed by following a User-Centred Design approach: different prototypes have been created and analyzed in formative evaluation sessions, whose results were instrumental to improve the quality of the successive prototypes. Different qualitative studies, involving users and experts in HCI and application domain, were performed, as summarized in this section.

##### A. Informal usability evaluation

The first CBP prototypes were analyzed through informal usability evaluations. Specifically, 5 Computer Science PhD students (3 males), who had attended an InfoVis course, analyzed independently the first paper prototype. Their comments were discussed with HCI experts and taken into account for implementing an early version of the CBP tool. Two successive running prototypes were analyzed in two different sessions; the first involved two researchers working on data analysis, while three experts of collaborative software development evaluated the second prototype. Each participant was individually asked to explore the CBP prototype. An HCI expert observed the interaction, acting as a facilitator whenever some difficulties arose and taking notes of the interface aspects that appeared more problematic. The results suggested to perform changes in order to provide appropriate feedback to the user, addressing, e.g., use of colors, legend terms and layout, status visualization and stripe visualization.

##### B. User test

A user test was performed on the running version of the CBP described in this paper. It was an observational study aimed at receiving user feedback on its design and usability.

A total of 6 senior students (4 male), aged 21-23 years old, participated in the study as their assignment for the InfoVis module in the HCI course. The current use of CBP is for distributed software design teams, thus these students represented a proper users sample, since they have experience in software design and development and use version control systems. An HCI researcher acted as a facilitator during each individual test session in a university laboratory. Another HCI researcher took notes. Each session consisted of three phases:

1) *Understanding*. 5 minutes were given to the participants to observe and provide their interpretation of the visualization. They did not interact with the tool but only commented on the different visualization elements, explaining what they understood about their meaning and their possible functionality. The facilitator intervened to prompt the participant to provide his/her interpretation.

2) *Get familiar*. Participants were asked to use the tool, performing 6 simple tasks, in order to get familiar with the interactions they could perform with the various visualization elements. Specifically, the first task asked to select, among the available teams, the "iOS developer.json" team. The second task required to zoom on a stripe associated to a member. Third and fourth tasks required to switch the position of two different pair of stripes. The fifth task asked to remove from the visualization a member and the associated stripe. The last task required to reset the visualization to the initial state.

3) *Execution*. Participants were asked to perform 6 typical tasks a user would carry out with CBP. These tasks were presented as questions, whose answers required an interpretation of some of the visualized data. However, the interpretation is facilitated if the user properly manipulates the visualization. The questions were about identifying availability of team members at specific times, which is the main purpose of CBP. For example a question was: "It's 7 PM. Could Roger Brown receive an immediate reply from John Smith in case he

sends him an email now?”. To provide the answer, it is useful to have the stripes of the two members adjacent one to the other and, possibly, zoom on them. Thus, the user should interact with the tool accordingly. Other questions were similar, e.g. another asked the user to identify the members that, at a certain time, could reply to a chat message. Again in this case, the user has to interact with the visualization, since it is easier to compare different stripes if they are placed each adjacent to the other.

The second and third phases were conducted using the thinking aloud method, which is well known and appreciated for providing very useful results without requiring many resources [17]. A laptop was used, and all actions performed by the participants were recorded. Even if times to complete the tasks were measured, our main goal was to observe user behavior and detect possible difficulties in understanding and using the tool. Participant interactions were video-taped and later transcribed.

#### *Test results and discussion*

At the beginning of the *Understanding* phase, the very first look at the visualization seemed to disorient the participants. Soon, they started to understand the user interface elements and what they could perform with them. All participants immediately realized that it is a time-oriented visualization, due to the clock shape. Four of them asked if the visualization could initially be shown with a different hour display, because they said that they would prefer to have at the top a different time than 12 (see Fig. 6).

The tasks performed by the participants in the *Get familiar* phase required simple actions (e.g. a gesture, two mouse clicks). Out of 6 users, only one user required assistance. Specifically, that user had difficulties in 3 tasks (second, third and fourth task) that required to associate a team member to his/her stripe. The HCI researcher provided him some clarifications. It appeared that he understood, since he was able to perform in a time below the average the successive two tasks, which had different objectives. However, he also had difficulties in the next phase. The other five users, who did not have difficulties in interacting with the user interface widgets, took on average 62 seconds to complete all the 6 tasks, which corresponds to an average of 10 seconds for each task. This time was measured from the moment in which the participant finished reading the task aloud until the goal was reached.

The tasks performed by participants in the *Execution* phase required more time than those in the *Get familiar* phase, also because they did not list the sequence of actions to perform, but they were formulated as questions to whom an answer should be provided. The user who had difficulties in the previous phase was not able to complete successfully the 6 tasks. Again, he showed problems in associating members and stripes; in some tasks, even if he correctly manipulated the visualization, he provided wrong answers about times. The remaining five participants successfully completed 28 out of 30 tasks. The failure of two participants for a same task was partial, since the sequence in the performed actions was correct but the final answer was wrong because the two participants misinterpreted the time in different time zones. The average time to complete all 30 tasks (of all five participants) was 3 minutes.

The observation of the participants during the *Execution* phase revealed that two different strategies were used to compare two stripes. Specifically, some participants zoomed the more internal stripe in order to enlarge the circular histograms and the boxplot, improving their readability. In this way it would be easier to compare with the other more external stripe, in which histograms and boxplot are bigger. The other strategy consisted in putting two stripes close each other by modifying the order of the members in the member list.

Participants immediately got familiar with the clock metaphor. All of them well understood that the visualization represented activities of the whole day. Only two of them did not immediately realize that the visualized period started at the moment of the use of the tool and ended the day after at the same time. Once this was explained to them, they did not have further difficulties.

#### *C. Focus group with domain experts*

Besides usability of the tool, we were very much interested to get more information about the utility of CBP for expert users and about possible uses in other contexts. Thus, we recently carried out a focus group that involved four experts of distributed software development, who are colleagues at our University. They were different from the three colleagues that analyzed one of the first prototypes, as reported in Section IV.A. Before the focus group, a session was planned to let each participant interact with CBP in a quite research laboratory. This session had the following organization. Initially, the participants watched a video of about two minutes that gives a demonstration of the CBP functionality. Then, the participants had a few minutes to practice with the tool. This practice was stopped once the participant said that he/she got familiar with the tool and understood the tool functionality. During this session, an HCI researcher acted as facilitator. He was answering possible questions but also prompted the participant to perform tasks similar to those performed in the *Execution* phase of the user test. Another HCI researcher took notes about the participant's interaction. Again, we were not interested in quantitative measures about the executed tasks, since such measures have a value only when considering much larger user samples. We wanted that the participants could understand the possibilities offered by the tool, in order to later discuss in the focus group about its utility in their work.

After each participant individually practiced with the tool, the focus group was held in a meeting room at our laboratory, with the two HCI researchers acting as moderator and observer, respectively. The discussion started by asking their impressions on CBP (Was it difficult to understand? Easy to use? etc.), but soon focused on their opinion about the utility of CBP, the existence of similar tools, its possible integration in collaboration tools they currently use. Then, participants were asked to discuss possible uses in different contexts in which people collaboration is important. Participants' interaction with the tool was video-recorded, while the focus group was audio-recorded. Later, video and audio were transcribed for the analysis.

All participants didn't know any similar tool, and they agreed on the utility of CBP as an effective tool for presence awareness. They appreciated very much the visualization at a

glance of the activities of team members that CBP provides. One participant mentioned the possible use of CBP in collaborative development environments, such as Microsoft Team Foundation Server or IBM Rational Team Concert. The other participants agreed. About other possible uses, CBP was judged useful for working contexts but not, for example, in social networks, which are mainly used for contacting friends and for entertainment. A participant explained that it is really difficult to know the schedule of a friend, so it is easier to setup a doodle when planning a meeting with friends.

A participant remarked that CBP could be useful also in code versioning systems, in which a team of developers is involved. It is very important to know data related to specific actions of each developer, e.g. visualizing data about commits, pulls and pushes, ticket requests, branches and forks, etc. For example, agile development is characterized by frequent commits. Another participant proposed himself to be part of a longitudinal study, in which the tool could be integrated as a plugin in Firefox. He was interested in visualizing primarily email messages, but other data could be also shown.

In general, participants appreciated the tool. Once they practiced a bit with the tool, they did not have problems in interacting with it. Two participants remarked that the demo presented in the video is too short. Actually, they could stop the video or rewind it but they did not. They also said that this was not a big problem since the HCI researcher supplemented the video with some comments when requested. However, this is suggesting us to revise the video in order to provide a better description.

## V. CONCLUSIONS AND FUTURE WORK

A novel visualization technique and a tool that implements it have been presented in this article. The technique uses circular representations to help people understand the activities performed by other team members during the day, in order to provide information about their availability.

The tool can be generalized to visualize various activity traces in different contexts, such as sensor data, work shifts and other activities going on during 24 hours. In particular, the focus group indicated possible use of CBP in collaborative development environments currently on the market, as well as in code versioning systems.

## VI. ACKNOWLEDGMENTS

We wish to thank all users that participated in our studies. We are grateful to G. Desolda for contributing to the first prototype of the tool and A. Di Bari for his help in the development of the current version of the tool. This work is partially supported by the Italian Ministry of University and Research (MIUR) under grant PON 02\_00563 3470993 "VINCENTE" and by the Italian Ministry of Economic Development (MISE) under grant PON Industria 2015 MI01\_00294 "LOGIN".

## REFERENCES

- [1] D. C. Engelbart. The mother of all demos. In Fall Joint Computer Conference, San Francisco, USA, 1968.
- [2] M. Pallot, U. Bergmann, H. Kühnle, K. S. Pawar, and J. CKH Riedel. Collaborative working environments: Distance factors affecting collaboration. In 16th International Conference on Concurrent Enterprising, ICE'2010, pages 1-32, PA, USA, 2010. IGI Publishing Hershey.
- [3] F. Lanubile, F. Calefato, and C. Ebert. Group awareness in global software engineering. *IEEE Software*, vol. 30, no. 2, pp. 18-23, 2013.
- [4] P. Buono and G. Desolda. Visualizing collaborative traces in distributed teams. In *Advanced Visual Interfaces, AVI '14*, pages 343-344, New York, NY, USA, 2014. ACM.
- [5] D. Fisher and P. Dourish. Social and temporal structures in everyday collaboration. In *SIGCHI Conference on Human Factors in Computing Systems, CHI '04*, pages 551-558, New York, NY, USA, 2004. ACM.
- [6] B. Heller, E. Marschner, E. Rosenfeld, and J. Heer. Visualizing collaboration and influence in the open-source software community. In *Mining Software Repositories*, pages 223-226, 2011.
- [7] J. "Bo" Begole, J. C. Tang, R. B. Smith, and N. Yankelovich. Work rhythms: Analyzing visualizations of awareness histories of distributed groups. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work, CSCW '02*, pages 334-343, New York, NY, USA, 2002. ACM.
- [8] P. Dragicevic and S. Huot. Spiraclock: A continuous and non-intrusive display for upcoming events. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems, CHI EA '02*, pages 604-605, New York, NY, USA, 2002. ACM.
- [9] Y. Riche and W. Mackay. Markerlock: A communicating augmented clock for elderly. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-computer Interaction Volume Part II, INTERACT'07*, pages 408-411, Berlin, Heidelberg, 2007. Springer-Verlag.
- [10] G. M. Draper, Y. Livnat, and R. F. Riesenfeld. A survey of radial methods for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(5):759-776, September 2009.
- [11] O. J. Zhao, T. Ng, and D. Cosley. No forests without trees: Particulars and patterns in visualizing personal communication. In *2012 iConference, iConference '12*, pages 25-32, New York, NY, USA, 2012. ACM.
- [12] T. Lovett, E. O'Neill, J. Irwin, and D. Pollington. The calendar as a sensor: Analysis and improvement using data fusion with social networks and location. In *Ubicomp '10*, pages 3-12, New York, NY, USA, 2010. ACM.
- [13] J. Tullio, J. Goecks, E. D. Mynatt, and D. H. Nguyen. Augmenting shared personal calendars. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology, UIST '02*, pages 11-20, New York, NY, USA, 2002. ACM.
- [14] J. Hong, E. Suh, and S. Kim. Context-aware systems: A literature review and classification. *Expert Syst. Appl.*, 36(4):8509-8522, May 2009.
- [15] J. Fogarty, J. Lai, and J. Christensen. Presence versus availability: The design and evaluation of a context-aware communication client. *Int. Journal Human-Computing Studies*, 61(3):299-317, September 2004.
- [16] J. Wilder T. *Exploratory Data Analysis*, Addison-Wesley, 1977.
- [17] J. Nielsen. *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- [18] C. Plaisant and B. Shneiderman. Scheduling on-off home control devices. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*, pages 459-460, New York, NY, USA, 1991. ACM.
- [19] J. Sauro, *Measuring Usability with the System Usability Scale (SUS)*, 2011, available at: [www.measuringusability.com/sus.php](http://www.measuringusability.com/sus.php), Last access: July 2014
- [20] A. M., Kessell, *Cognitive methods for information visualization: Linear and cyclical events*, Ph.D. thesis, Stanford University, 2008

# A Tabu Search Based Approach for Graph Layout

Fadi K. Dib

Computer Science Department  
Gulf University for Science and Technology  
Kuwait, Kuwait  
deeb.f@gust.edu.kw

Peter Rodgers

School of Computing  
University of Kent  
Canterbury, UK  
P.J.Rodgers@kent.ac.uk

**Abstract**—This paper describes an automated tabu search based method for drawing general simple graph layouts with straight lines. To our knowledge, this is the first time tabu methods have been applied to graph drawing. We formulated the task as a multi-criteria optimization problem with a number of metrics which are used in a weighted fitness function to measure the aesthetic quality of the graph layout. The main goal of this work is speeding up the graph layout process without sacrificing the layout quality. To achieve this we used a tabu search based method that goes through a predefined number of iterations to minimize the value of the fitness function. Tabu search always chooses the best solution in the neighborhood. This may lead to cycling, so a tabu list is used to store moves that are not permitted so that the algorithm does not choose previous solutions for a set period of time. We give experimental results applied on random graphs and we provide statistical evidence that our method outperforms one of the fast search-based drawing methods (hill climbing) in execution time while it produces comparably good graph layouts. We also demonstrate the method on real world graph datasets.

**Keywords**—*Information visualization; graph drawing; graph layout; tabu search*

## I. INTRODUCTION

In this work we address the research area of graph drawing. Here, the goal is to lay out a network diagram so that it can be analyzed and examined by users. There are several multi-criteria approaches to graph drawing which are based on explicit cost functions that combine several metrics of graph layout quality. This approach has the advantage of allowing explicit, tunable combinations of metrics to meet user preferences. However, such methods work slowly, typically taking a considerable time to lay out the graph. The contribution of this paper is to improve the performance of such systems by introducing tabu search features. To our knowledge, this is the first time tabu methods have been applied to drawing general simple graph layouts with straight lines.

Search based methods typically measure a number of metrics and combine them to form a fitness measure. When a new solution is found (perhaps by moving a node) the metrics are calculated again and a new fitness measure is found. This process happens a large number of times during the search, and so the process of finding a good layout is slow. Many drawers in the literature used search based methods, such as simulated annealing [2, 3], genetic algorithms [4, 5, 6, 7] and hill

climbing [8, 9]. These produce good layouts, but they have great potential for improvement. For example, simulated annealing adds an element of non-determinism in order to escape from local minima in the search space. This slows down the performance of the algorithm since this stochastic behavior means that a larger number of iterations would be necessary to reach a minimum in the search space. Genetic algorithms, on the other hand, typically have an even slower rate of convergence compared to simulated annealing and hill climbing as it makes a wider search of the problem space. The main problem with hill climbing is that it gets trapped in local optima.

Our main goal in this work is concerned with improving the drawer's efficiency by speeding up the drawing process, using a search based method known as tabu search, without sacrificing the layout quality. We are not looking for the global optimum solution, but aim to obtain a good optimal solution quickly. Therefore, we compare our approach with hill climbing. In addition to its simple implementation, hill climbing has proven its efficiency in graph drawing applications [8, 9]. The main disadvantage of hill climbing is the likelihood of finding a sub-optimal local minimum in the search space. However, as the method is completely deterministic, comparison against hill climbing is more reliable than against the non-deterministic approaches of simulated annealing and genetic algorithms.

Tabu search is a general search based technique proposed by Glover [10, 11, 12] for finding good solutions to combinatorial optimization problems. It is considered to be a neighborhood search method (like simulated annealing) but it takes a more aggressive approach. It proceeds on the assumption that there is little benefit in choosing an inferior solution unless it is necessary, as in the case of escaping from a local optimum [13]. In other words, tabu search improves the efficiency of exploration process by keeping track of local information (like the current value of the objective function) along with information related to the exploration process. This systematic use of memory is an essential property of this searching technique. Tabu search keeps information on the itinerary through the last solutions visited. The role of this is to restrict the choice of some subsets in the neighborhood by forbidding moves to some neighbor solutions that have already been visited [14]. This constrains the direction of the search process by preventing the algorithm from going back to a previously reached state. At each iteration of the exploration process, it selects the best solution in the neighborhood. This is

unlike hill-climbing as it might make a down-hill move. Therefore, this technique does not run out of choices for the next move. However, this might lead to cycling by trapping the algorithm at locally optimal solutions. This problem can be resolved by introducing two structures called *tabu lists* and *aspiration functions* which are used to keep information about past moves in order to respectively constrain and diversify the search for good solutions [13].

Tabu search has shown good results and comparably fast solutions for some graph theory applications such as graph partitioning [13], graph coloring [15] and straight line crossing minimization [16, 17, 18]. It has also been used to solve different multiple objective optimization problems. The algorithm was used to solve four different applications in different areas [19]. In every application, the solutions were at least as good as, if not better than, the reported results using different search based techniques. Tabu search has been applied to the problem of routing school buses [20]. This algorithm was shown to be competitive in a set of problem instances for which a scatter search method was applied. Other researchers [21] proposed a tabu search algorithm as meta-heuristic method for network reconfiguration of multiple objectives problems in a radial distribution system. The work concluded that tabu search can quite easily handle the complicated constraints that are typically found in real-life applications. However, it failed in some circumstances for the following two reasons: an insufficient understanding of fundamental concepts of the tabu search method; and a lack of understanding of the problem in hand.

Our paper describes an approach for drawing general simple graphs with straight lines. This is achieved with a tabu search based method which draws general graphs with multiple aesthetic criteria that include node-node occlusion, edges length, edge crossings, and angular resolution. These criteria are used in a weighted fitness function to measure the quality of the graph layout. Whilst there have been empirical studies of what may be the most effective criteria for layout [1], we are not overly concerned with the particular criteria or their weights: our method would work effectively with other criteria or weightings.

The method goes through predefined number of iterations to minimize the value of the fitness function and it uses a tabu list to store tabu moves in order to prevent the algorithm from choosing previously reached moves for particular nodes for a period of time. We have tested our method on random graphs of different sizes and we describe the experiments and statistical analysis that brought us to the conclusion that our tabu search based method produces graph layouts as good as, if not better than, layouts drawn with a hill climbing method with a clear improvement in the time spent to draw the graph. We have also recognized improvements in both quality and time over hill climbing when the method is applied to real world graphs.

The rest of this paper is organized as follows: Section II describes some background in using search based techniques in graph layout; Section III describes our approach; Section IV describes experimental results on random graphs; Section V describes the results of applying our approach to real world

graph datasets; finally, Section VI gives our conclusions and suggests future work.

## II. RELATED WORK IN GRAPH LAYOUT

Multi-criteria graph layout can be modelled as a multiple objective optimization problem. When an algorithm attempts to draw a graph layout according to several graph aesthetic criteria, some of these criteria might conflict with each other. Hence, a fitness function that linearly combines all criteria is formulated. The optimizer attempts to minimize this function.

The problem with using general fitness functions is that it is usually computationally expensive to find a minimum fitness value. Since the overall fitness function might include both continuous and discrete measures, general search based approaches, such as simulated annealing, genetic algorithms, and hill climbing, have been used in order to find a minimum fitness value [35, 36].

Simulated annealing was the first general search method to be applied to the graph layout problem [2]. It was used to draw undirected graphs with straight line edges, taking into account several drawing aesthetics: distributing nodes evenly, making edge lengths uniform, minimizing edge crossings, and placing nodes not too close to edges. All these criteria were combined into a function that could be subject to a general optimization fitness function. This search based approach models the physical process of heating a material and then slowly cooling the temperature to decrease defects, so minimizing the system energy. It is often used for large-scale combinatorial optimization problems and implemented in a way that tries to escape from local minimum to global minimum by applying uphill moves (moves that worsen, rather than improve, the temporary solution). This allows the approach to escape from some local minimal solutions but with no guarantee that a global minimum can be reached eventually. The algorithm produces nice graph layouts for small sized graphs. However, it does not perform well for larger graphs.

An adjustment to simulated annealing approach was made in the algorithm proposed in [3]. Here, the fitness function is minimized using gradient descent. The gradient vector of the fitness function represents the direction in which the node should move to increase the value of the fitness function. Thus, this algorithm will move the node to the opposite direction to minimize the value of the fitness function. But this method is still slow when being applied on large graphs and it has some challenges. For example, the fitness function needs to be expressed explicitly in terms of coordinates as its derivative must be found. Some criteria, such as minimizing edge crossings, are discontinuous and not differentiable.

Genetic algorithms have also been applied to the graph layout problem. A genetic algorithm approach for drawing graphs under a number of visual constraints was proposed in [4, 5]. The proposed algorithm produces graphs with good quality in addition to its flexibility. It can be easily adapted to take new layout aesthetics into account. However, the major problem in this algorithm is its slow rate of convergence. It initially makes rapid progress towards a solution, but then it



converges very slowly to a global optimum, or at least to a good local one.

A genetic algorithm with local fine tuning based on the spring algorithm for the drawing of undirected graphs with straight-line edges has been proposed in [6]. According to some preliminary results, the algorithm produces layouts with a minimal number of edge crossings on all test graphs. The algorithm benefits from the combination of the genetic algorithm and the spring algorithm to produce good layouts for a large class of graphs with implicit symmetry, similar spring lengths, and even distribution of nodes. Although the layouts found by this algorithm have good general structures, some fine tuning might still be needed. Moreover, the comparatively long running time of the algorithm is a key disadvantage. One reason for the high time complexity of the algorithm comes from the chosen crossover operator to solve the competing conventions problem which states that a recombination of two good parents may yield a very poor offspring.

Similar work was introduced in [7]. This proposed a genetic algorithm that nicely draws undirected graphs of moderate size. But the algorithm still suffers from the lack of proper crossover operation which would speed up the computations by decreasing the number of needed generations.

Hill climbing is another search based approach that has been used in the field of graph drawing. It is one of the simplest search based algorithms used in the field of artificial intelligence. It is good for finding local optimum but it is not guaranteed to find the global optimum out of all possible solutions. It works by iteratively improving a given solution, which is often selected in a random way, by applying a transformation in the current solution or picking any solution in its neighborhood. Then, the new solution is compared to the old one. If the new solution is better than the old one, the new solution substitutes the old one. This process is repeated until a maximum number of repetitions is reached.

Hill climbing has been used to minimize number of edge crossings [8]. The experiments conducted on random graphs of different sizes showed that stochastic hill climbing outperforms efficient and popular search based techniques such as evolution strategies and genetic algorithms.

A hill climbing approach has been used to implement an automatic mechanism for drawing metro maps [9]. Metro map drawing is a specialized form of graph drawing. A good metro map layout has evenly spaced stations, lines at regular angles (typically multiples of 45 degrees) and labels placed in unambiguous locations. This work applied multi-criteria optimization using a fitness function consisting of five different metrics in a weighted sum, along with some rules that prevented some bad moves for each station (e.g. a station that was north of another station could not be moved south of it). A hill climbing algorithm was used to reduce the fitness function and find improved map layouts. Since hill climbing does not guarantee finding the global minimum, a clustering technique was applied to the map. The hill climber moves both stations and clusters when finding improved layouts. The mechanism produces good map layouts and in some cases better than both published and distorted layouts. However, the performance of the algorithm is slow.

Search based methods used to solve graph layout problem are generally successful in producing graphs with nice layouts but just for small or mid-sized graphs. In addition, the execution time of these methods is very slow.

### III. OUR APPROACH

This section describes the basic concepts of our approach. We detail the algorithm of tabu search drawer, the criteria measured, and the method for combining criteria to produce a fitness value.

In outline, as described in Algorithm 1, our tabu search method operates in the following manner: first, we find a random initial graph layout such that no two nodes have the same position. We compute the fitness value of the initial layout. Then the following steps are performed for a predefined number of iterations: for each node, we search the points around a square centered on the node at a given distance, as shown in Fig. 1. Eight points around the square are checked (above, below, left, right, and the four corners). The ratio of the current solution's fitness function value with the previous solution's fitness function value is computed at each point around the square. Solutions with fitness function ratios above or equal to a predefined threshold value (`tabuCutOff`), are considered as tabu moves and will be stored in a tabu list. We then move the node to a non-tabu point where the value of the fitness function is a minimum compared to all the points of the square even if the new point does not improve the current value of the fitness function (this is why tabu search does not run out of solutions) and the previous solution becomes a tabu solution. After an arbitrary chosen number of iterations, as a cooling down process, the square size centered around the node is reduced and the `tabuCutOff` value is decreased to intensify the searching process. Finally, the tabu list is updated by removing old solutions from the list after a number of iterations in which a move should remain in tabu list before it can be released (`tabuDuration`) in order to diversify the searching space. Fig. 2 presents two examples of graph layouts drawn by our approach.

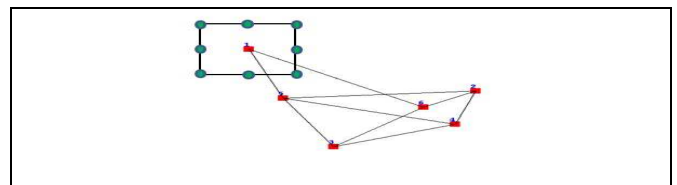


Fig. 1. Example of the points around the square checked by our algorithm on each node

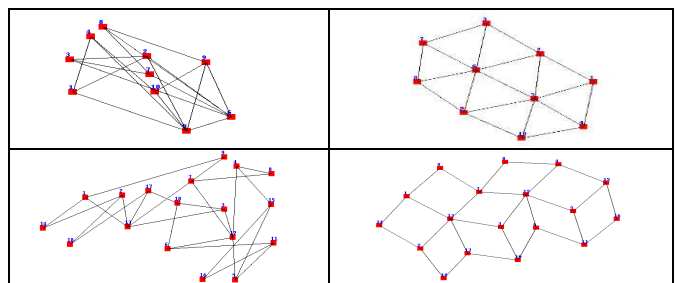


Fig. 2. Examples of layouts before (left) & after (right) applying our approach

### Algorithm 1. Tabu Search Drawer

#### Given:

Connected Graph  $G(V, E)$ :  $V$  is a set of nodes and  $E=(V \times V)$  is a set of edges.

`max_iterations`: predefined maximum number of iterations of the drawer.

`coolDown_iterations`: predefined number of iterations in which the process starts cooling down

`tabuDuration`: predefined number of iterations in which a move should remain in the tabu list.

`tabuCutOff`: predefined minimum value that determines whether a move is tabu or not.

#### Algorithm:

```
1: InitializeTabuList()
2:  $G_{Layout} = \text{RandomizeLayout}(G)$ 
3: iterations = 0
4: while (iterations < max_iterations)
5:   for  $v \in V$  do
6:     for  $\text{squarePos} \in \text{allNonTabuSquarePositions}$  do
7:        $\text{currentPos} = v_{\text{pos}}$ 
8:        $\text{fitness}_{\text{Current}} = \text{Fitness}(G_{Layout})$ 
9:       Update  $G_{Layout}$  s.t.  $v_{\text{pos}} = \text{squarePos}$ 
10:       $\text{fitness}_{\text{New}} = \text{Fitness}(G_{Layout})$ 
11:      if ( $\text{fitness}_{\text{New}} / \text{fitness}_{\text{Current}} > \text{tabuCutOff}$ )
12:        addToTabu( $v, \text{squarePos}$ )
13:      end if
14:    end for
15:     $v_{\text{pos}} = \text{Min\_NonTabuPosition}()$ 
16:    addToTabu( $v, \text{currentPos}$ )
17:  end for
18:  if (iterations mod coolDown_iterations == 0)
19:    ReduceSquareSize()
20:    Decrease(tabuCutOff)
21:  end if
22:  if (iterations  $\geq \text{tabuDuration}$ )
23:    RemoveTabuSolutions(iterations - tabuDuration)
24:  end if
25:  iterations = iterations + 1
26: end while
```

Our fitness function follows the standard approach for search based graph drawing methods. We implemented four metrics for measuring the quality of the graph [2, 9]. These represented the aesthetics of: distributing nodes evenly, making uniform edge lengths, minimizing edge crossings, and improving angular resolution. All these metrics contribute in the graph quality fitness function which is computed as follows:

$$\text{fitness} = w_1 * m_1 + w_2 * m_2 + w_3 * m_3 + w_4 * m_4 \quad (1)$$

where  $w_i$  and  $m_i$  are the weight and the measure for criteria  $i$  respectively. The problem in a multiple objective optimization function is that the value of a specific measure may dominate the others. Therefore, we applied a normalization process to ensure that the value of each measure is between 0 and 1.

We cannot determine unified weights that work properly for any graph. Therefore, the weights should be assigned by decision makers according to their preferences on which measure they prefer to dominate. We assigned the value 1 to all

weights in order to avoid domination of one measure over another.

We realized that re-computing the fitness function at each point is a time consuming process. Therefore, we modified the way of computing the value of the function by caching the results such that the old value of the function is used to compute the new value. We just compute the change made in the function when a node is moved. When a node moves to a new position, the amount of change in the fitness function value, between the previous position of the node and its new position, is computed. If the change in the value improves the fitness function, we subtract the amount of the change made by the previous position of the node, and we add the amount of the change made by the new position of the node to get the new value of the fitness function. This process increased the speed of our method.

To tune the two parameters `tabuDuration` and `tabuCutOff`, we performed several experiments. We generated 100 random graphs (different to those shown in Table I and II). These were divided into 5 sets such that each set had different number of nodes and edges. Hence, each set consisted of 20 test cases with the same number of nodes and edges. The characteristics of the five sets were exactly the same of the first five groups of the graphs in the second category as will be described in the next section. We tested the drawer on the 100 test cases for six different values of `tabuDuration` {5, 6, 7, 8, 9, and 10} and for six different values of `tabuCutOff` {50, 60, 70, 80, 90, and 100}. In most of the cases, the best fitness function values and the shortest execution time were generated when the values of the two parameters were: `tabuDuration` = 7 and `tabuCutOff` = 80.

## IV. EXPERIMENTAL RESULTS

The programming language used in our implementation is Java (version 1.7.0; Java HotSpot™ 64-Bit Server VM 21.0-b17 on Windows 7). We have tested our approach on different random graphs of different sizes. The experiments were performed using Lenovo Thinkpad T430, Intel® Core™ i7-3520M CPU processor with frequency of 2.90 GHz and 8 GB RAM.

We generated random graph datasets in two categories. The graphs of the first category have the same number of nodes but with different densities (i.e. different number of edges), whereas the graphs of the second category have different number of nodes with varying values of densities.

Our random graph generator generated random connected graphs. The parameters to it were the number of nodes and the density of the graph. It generated random locations for the nodes based on the size of the window where the graph will be displayed. Then, the generator chose random nodes as end points of edges. Self-sourcing edges and multiple edges between the same pair of nodes were not allowed. Finally, the generator tested the connectivity of the generated graph by running a breadth first search algorithm. Only connected graphs were accepted.

There were 200 random graphs in the first category split into 10 groups of 20 test cases each. All the graphs in this category had 150 nodes, randomly positioned. However each group had a differing number of edges than the other groups so that the density varied. However, the graphs in each group had same number of edges but with different initial layouts. See TABLE I for characteristics of the graphs in the first category.

The second category also had 200 random graphs, again split into 10 groups. Number of nodes in each group was increased by 50. The value of the density was chosen for each group to avoid too dense graphs. A similar random process used to generate graphs in the first category was applied to this category. See TABLE II for characteristics of the graphs in the second category.

TABLE I  
Characteristics of the graphs in the 1st category

Group	Nodes	Edges	Density
1	150	558	0.05
2	150	1117	0.1
3	150	1676	0.15
4	150	2235	0.2
5	150	2793	0.25
6	150	3352	0.3
7	150	3911	0.35
8	150	4470	0.4
9	150	5028	0.45
10	150	5587	0.5

TABLE II  
Characteristics of the graphs in the 2nd category

Group	Nodes	Edges	Density
1	50	153	0.125
2	100	544	0.11
3	150	1173	0.105
4	200	1890	0.095
5	250	2645	0.085
6	300	3363	0.075
7	350	3969	0.065
8	400	4788	0.06
9	450	5556	0.055
10	500	6237	0.05

We have applied our tabu search based approach and the hill climbing approach to the randomly generated graphs. All the weights of the metrics in the fitness function were equal. The metrics were normalized and therefore, equalizing the weights would equalize the effect of each metric on the value of the fitness function.

We note that the hill climbing approach used the same optimized fitness function calculations that the tabu search applied – only changes to the fitness function from moved nodes were recalculated.

To make a comprehensive comparison between tabu and hill climbing, we divided our experiments into three phases. In phase I, we applied both methods on the graphs of the two

categories. The methods executed on the 20 test cases in each group of the two categories, and then the average execution time and the average fitness function value were computed for each group. The hill climbing approach was executed until it found the best solution that can be reached by the approach (i.e. a solution that cannot be of further improvement). On the other hand, our tabu search based approach ran for 50 iterations. A cut-off point had to be chosen because tabu search always moves to the point with the best fitness value of all the eight points around the square on each node even if the new point does not improve the current value of the fitness function and hence it would not run out of solutions.

Fig. 3 and Fig. 4 show bar charts of the results obtained from phase I. The charts clearly show the difference between the two methods in terms of the quality of the produced layouts and the execution time. The figures show that both methods give similar values for the fitness function with a slight advantage to our method. However, the execution time of our approach clearly outperforms the execution time of the hill climbing approach.

In phase II, we investigated the performance of approaches rather than the quality of the produced layouts. Therefore, the following process was performed to test which method has faster execution time when they reach similar values for the fitness function:

1. We ran the hill climbing method on the graphs until no improvements could be made on the value of the fitness function.
2. We ran our tabu search method until it reached an equal or better fitness function value compared to the one found by the hill climbing drawer.
3. We measured the execution time of the methods.

Fig. 5 shows bar charts of the results obtained from phase II. The columns obviously show that our drawer always finishes faster than the hill climbing drawer. These results indicate that excluding previously visited solutions from further investigation for a specific period of time is clearly an effective property in tabu search.

Finally, in phase III, we investigated the quality of the layout produced by the drawers rather than the performance. The following process was performed to test which method produces graph layouts with smaller values of fitness function when both drawers execute for the same period of time:

1. We ran the tabu search method on the graphs for 50 iterations. The execution time is computed and saved.
2. We ran the hill climbing method for the same period of time spent by the tabu search method.
3. We measured the value of the fitness function produced by the drawers in each of the above steps.

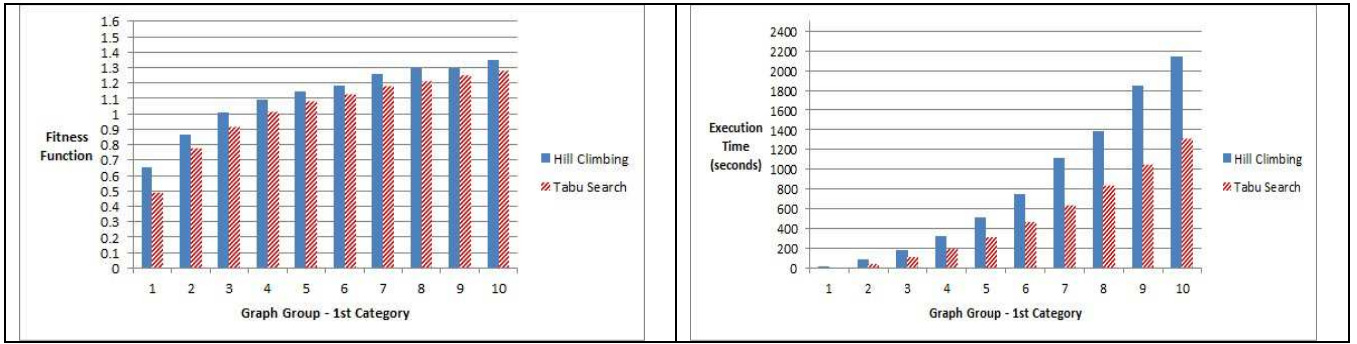


Fig. 3. Bar charts of the fitness function and execution time (in seconds) obtained by both methods when applied on the graphs of 1st category (Phase I)



Fig. 4. Bar charts of the fitness function and execution time (in seconds) obtained by both methods when applied on the graphs of 2nd category (Phase I)

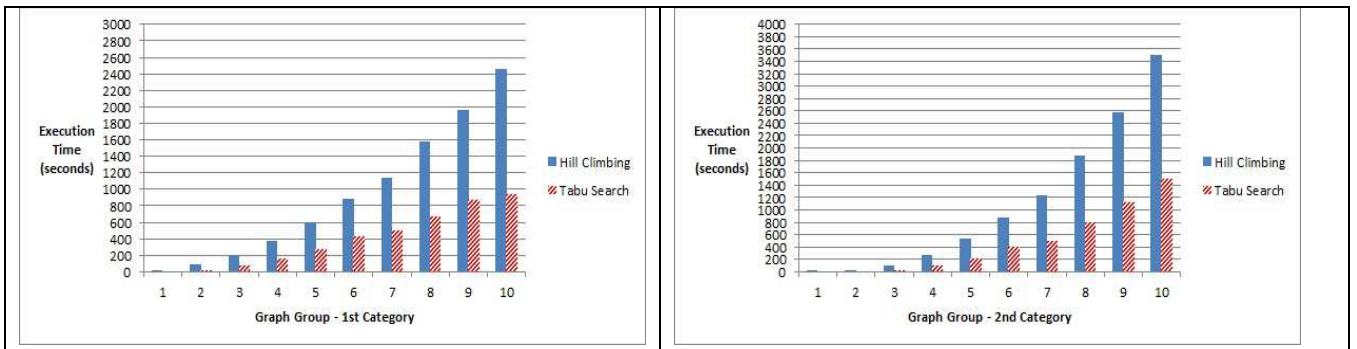


Fig. 5. Bar charts of the average execution time (in seconds) when the methods are applied on the graphs of the 1st and the 2nd categories (Phase II)

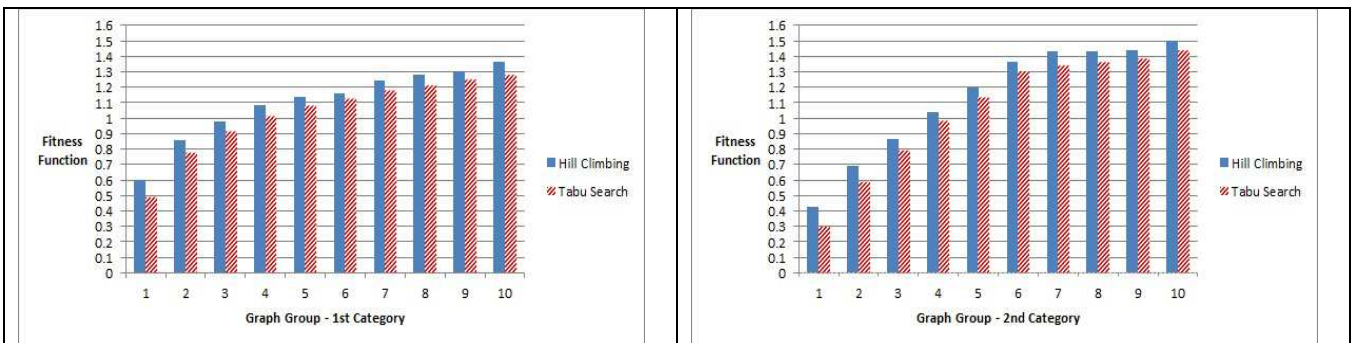


Fig. 6. Bar charts of the average values of the fitness function when the methods are applied on the graphs of the 1st and the 2nd categories (Phase III)

Fig. 6 shows bar charts of the results obtained from phase III. The columns look similar with a slight advantage to our tabu search based drawer. Therefore, we can conclude that our approach produces better graph layouts compared to hill

climbing or similar layouts in the worst case when both drawers run for the same period of time.

Fig. 7 shows three different examples of graphs drawn by hill climbing approach and our approach.

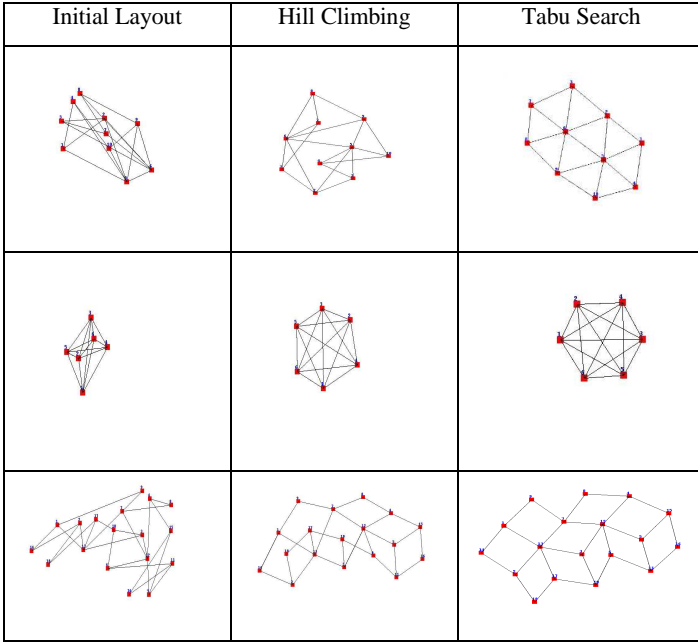


Fig. 7. Examples of graphs drawn by hill climbing and tabu search approaches

In terms of threats to validity, the algorithms used were deterministic and both used the same starting layout. The main internal threat seems to be in the implementation of the algorithms. Both methods were implemented by the same coder, and were run on the same machine. There is the possibility that one of the hill climber or tabu search was implemented in a more efficient way, however, the methods are sufficiently similar, sharing key code, so permitting some confidence that neither was particularly disadvantaged. In terms of external threats – that is to the generalizability of the results. We tested a number of randomly generated graphs that prevents selection bias (except in the parameters of the generation algorithm, such as number of nodes and edges). However, randomly generated graphs generally do not have the same characteristics as real world graphs, and so in the next section we explore the method applied to real world datasets sourced from the internet, although further real world testing is required to fully explore the generality of the results.

TABLE III  
Real world graph datasets characteristics and sources

Graph	Nodes	Edges	Density	Source
1	34	78	0.139	[22]
2	62	159	0.084	[23]
3	105	441	0.081	[24]
4	112	425	0.068	[25]
5	115	613	0.094	[26]
6	198	2742	0.141	[27]
7	277	1918	0.050	[28]
8	297	2148	0.049	[29]
9	453	2025	0.020	[30]
10	500	13038	0.104	[31]
11	332	2126	0.039	[32]
12	415	7519	0.088	[33]
13	128	2075	0.255	[34]

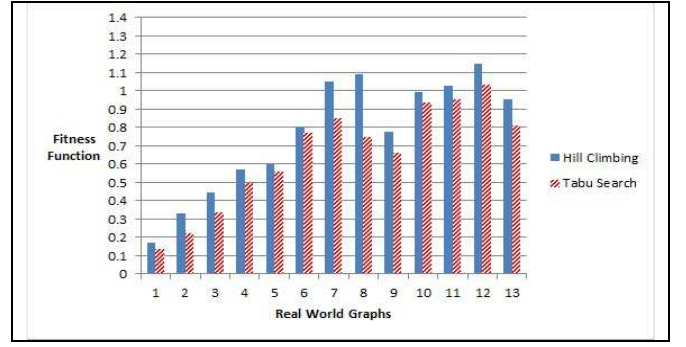


Fig. 8. Bar chart for the values of the fitness function of the two methods when applied on graph datasets in TABLE III

## V. REAL WORLD GRAPH DATASETS

After performing several experiments on random graphs, we tested our system on real world graph datasets. We selected 13 different datasets from different sources as shown in TABLE III that also shows number of nodes, number of edges, and density in each graph. The graphs have different sizes with different densities. The initial layout of the nodes in each graph was generated randomly.

The results of the experiments are shown in Fig. 8 and TABLE IV. In the first two datasets only, the execution time of hill climbing is slightly faster than our tabu search based approach. This is due to the small size of those graphs. The results in the table demonstrate that our approach outperforms hill climbing approach in terms of execution time while the size of graphs increases. We also note from the figure that the values of the fitness function are always better in our approach regardless of the size of the graph.

Fig. 9 is an example of the layout produced by our drawer when applied on the first graph dataset in the list of real world datasets described in TABLE III.

TABLE IV  
Execution time (in seconds) for both methods on graph datasets in TABLE III

Graph	Execution Time (seconds)	
	Hill Climbing	Tabu Search
1	0.699	0.931
2	1.405	1.826
3	11.822	9.421
4	16.234	9.671
5	18.192	15.543
6	468.838	298.526
7	258.042	149.060
8	323.139	185.277
9	347.227	193.338
10	14107.292	5968.619
11	384.990	210.799
12	4458.639	2190.103
13	257.392	178.600

## VI. CONCLUSIONS AND FUTURE WORK

We have described an automated tabu search based approach for drawing general simple graphs with straight lines based on multi-criteria optimization. The method searches for



the best positions for the nodes that minimizes the value of the fitness function and draws a nice graph layout accordingly. Forbidding reverse moves, and the ability to escape local optima are two features that make tabu search a more effective layout method than hill climbing. Our experimental results on random graphs and real world graphs show that tabu search approach is faster than hill climbing, and in some cases the time is almost half, regardless of the size of the graph in terms of number of nodes and edges. On the other hand, both approaches produce layouts with good quality and in most cases tabu search approach slightly outperforms hill climbing.

In terms of future work, the definition of a tabu move might change. Instead of using absolute node position to determine whether a move is tabu or not, we might use its relative position. For instance, when a node moves to the left direction, its adjacent nodes should not move in the same direction, because moving them to the same direction is like shifting the whole sub-graph in one direction. Also, a graph clustering method can be used to divide the graph into sub-graphs where

the nodes in each sub-graph would move according to their relative positions in their own cluster and each sub-graph would move according to its absolute position.

Also, it may be possible to develop a systematic way for choosing the values of the parameters used by our method. Several tests have been made with different values for `tabuDuration` and `tabuCutOff` parameters to come up with values that speed up the performance and produce nice graph layouts at the same time. More tests on different sets of graphs with different characteristics might lead to a clear process for choosing the values of the parameters.

Finally, the performance of our method may be further improved by implementing a hybrid of tabu search and other search based methods such as scatter search and path relinking. An interesting aspect of scatter search is that the approach performs a deterministic search instead of a random one. Path relinking, on the other hand, has the advantage of using previously encountered good solutions to obtain diversification and intensification in the search.

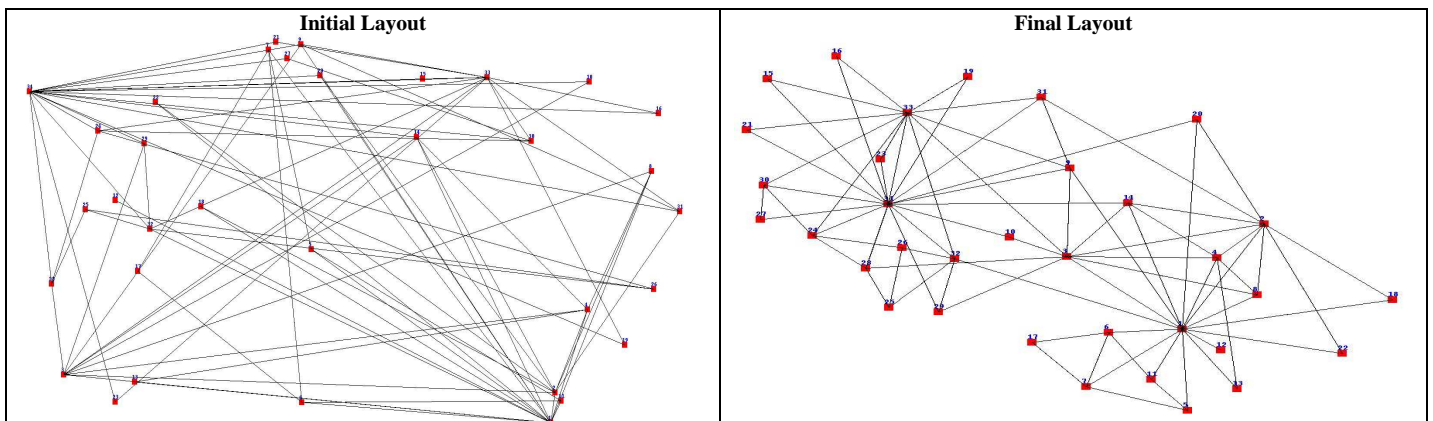


Fig. 9. Layout of graph dataset 1 (listed in TABLE III) produced by our method

## REFERENCES

- [1] H. C. Purchase, "Metrics for graph drawing aesthetics," *Journal of Visual Languages and Computing*, vol. 13, no. 5, pp. 501-516, 2002.
- [2] R. Davidson and D. Harel, "Drawing graphs nicely using simulated annealing," *ACM Transactions on Graphics*, vol. 15, no. 4, pp. 301-331, 1996.
- [3] J. Brank, "Drawing graphs using simulated annealing and gradient descent," *Zbornik C 7. mednarodne multikonference Informacijska družba*, pp. 67-70, 2004.
- [4] C. Kosak, J. Marks and S. Shieber, "A parallel genetic algorithm for network diagram layout," San Diego, CA, USA, 1991.
- [5] C. Kosak, J. Marks and S. Shieber, "Automating the layout of network diagrams with specified visual organization," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 3, pp. 440-454, 1994.
- [6] J. Branke, F. Bucher and H. Schneck, "Using genetic algorithms for drawing undirected graphs," *Proceedings of Third Nordic Workshop on Genetic Algorithms and their Applications*, pp.193-206, 1996.
- [7] T. Eloranta and M. Erkki, "TimGA: A genetic algorithm for drawing undirected graphs," *Divulgaciones Matematicas*, vol. 9, no. 2, pp. 155-171, 2001.
- [8] A. Rosete-Suárez, A. Ochoa-Rodriguez and M. Sebag, "Automatic graph drawing and stochastic hill climbing," In *Proceedings of the Genetic and Evolutionary Computation Conference*, vol. 2, pp. 1699-1706, 1999.
- [9] J. Stott, P. Rodgers, J. C. Martinez-Ovando and S. G. Walker, "Automatic metro map layout using multicriteria optimization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 1, pp. 101-114, 2011.
- [10] F. Glover, "Future paths for integer programming and links to artificial intelligence," *Computer and Operations Research*, vol. 13, no. 5, pp. 533-549, 1986.
- [11] F. Glover and H. J. Greenberg, "New approaches for heuristic search: A bilateral linkage with artificial intelligence," *European Journal of Operational Research*, vol. 39, pp. 119-130, 1989.
- [12] F. Glover, "Tabu search - part I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190-206, 1989.
- [13] A. Lim and Y. M. Chee, "Graph partitioning using tabu search," *IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 1164-1167, 1991.
- [14] A. Hertz, E. Taillard and D. De Werra, "A tutorial on tabu search," In *Proceedings of Giornate di Lavoro AIRO*, vol. 95, pp. 13-24, 1995.
- [15] A. Hertz and D. De Werra, "Using tabu search techniques for graph coloring," *Computing*, vol. 39, no. 4, pp. 345-351, 1989.
- [16] M. Laguna, R. Marti and V. Valls, "Arc crossing minimization in hierarchical digraphs with tabu search," *Computers and Operations Research*, vol. 24, no. 12, pp. 1175-1186, 1997.
- [17] R. Marti, "A tabu search algorithm for the bipartite drawing problem,"



- European Journal of Operational Research, vol. 106, no. 2, pp. 558-569, 1998.
- [18] R. Martí and M. Laguna, "Heuristics and meta-heuristics for 2-layer straight line crossing minimization," *Discrete Applied Mathematics*, vol. 127, no. 3, pp. 665-678, 2003.
- [19] A. BAYKASOGLU, S. OWEN and N. GINDY, "A taboo search based approach to find the pareto optimal set in multiple objective optimization," *Engineering Optimization*, vol. 31, no. 6, pp. 731-748, 1999.
- [20] J. Pacheco and R. Martí, "Tabu search for a multi-objective routing problem," *Journal of the Operational Research Society*, vol. 57, no. 1, pp. 29-37, 2005.
- [21] T. Thakur and J. Dhiman, "A tabu search algorithm for multi-objective purpose of feeder reconfiguration," *Journal of Electrical and Electronics Engineering Research*, vol. 3, no. 4, pp. 71-79, 2011.
- [22] W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452-473, 1977.
- [23] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Sloaten and S. M. Dawson, "The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations," *Behavioral Ecology and Sociobiology*, vol. 54, no. 4, pp. 396-405, 2003.
- [24] V. Krebs, "http://www.orgnet.com," unpublished. Retrieved March 1, 2014.
- [25] M. E. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical review E*, vol. 74, no. 3, p. 036104, 2006.
- [26] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821-7826, 2002.
- [27] P. M. Gleiser and L. Danon, "Community structure in jazz," *Advances in complex systems*, vol. 6, no. 04, pp. 565-573, 2003.
- [28] Y. Choe, B. H. McCormick and W. Koh, "Network connectivity analysis on the temporally augmented *C. elegans* web: A pilot study," In *Soc Neurosci Abstr*, vol. 30, no. 921.9, 2004.
- [29] J. G. White, E. Southgate, J. N. Thomson and S. Brenner, "The structure of the nervous system of the nematode *Caenorhabditis elegans*," *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, vol. 314, no. 1165, pp. 1-340, 1986.
- [30] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization," *Physical review E*, vol. 72, no. 2, p. 027104, 2005.
- [31] J. Marcelino and M. Kaiser, "Critical paths in a metapopulation model of H1N1: Efficiently delaying influenza spreading through flight cancellation," *PLoS currents*, vol. 4, 2012.
- [32] V. Batagelj and A. Mrvar, "Pajek datasets. Web page <http://vlado.fmf.uni-lj.si/pub/networks/data>," 2006.
- [33] P. J. Taylor, *World city network: a global urban analysis*, Routledge, 2003.
- [34] C. J. Melián and J. Bascompte, "Food web cohesion," *Ecology*, vol. 85, no. 2, pp. 352-358, 2004.
- [35] M. G. Resende and C. C. Ribeiro, "GRASP: greedy randomized adaptive search procedures," in *Search Methodologies*, Springer US, pp. 287-312, 2014.
- [36] R. Tamassia, *Handbook of graph drawing and visualization*, AMC, 10, 12, 2011.

# eulerForce: Force-directed Layout for Euler Diagrams

Luana Micallef

School of Computing, University of Kent, Canterbury, UK  
L.Micallef@kent.ac.uk

Peter Rodgers

School of Computing, University of Kent, UK  
P.J.Rodgers@kent.ac.uk

**Abstract**—Euler diagrams use closed curves to represent sets and their relationships. They facilitate set analysis, as humans tend to perceive distinct regions when closed curves are drawn on a plane. However, current automatic methods often produce diagrams with irregular, non-smooth curves that are not easily distinguishable. Other methods restrict the shape of the curve to for instance a circle, but such methods cannot draw an Euler diagram with exactly the required curve intersections for any set relations. In this paper, we present *eulerForce*, as the first method to adopt a force-directed approach to improve the layout and the curves of Euler diagrams generated by current methods. The layouts are improved in quick time. Our evaluation of *eulerForce* indicates the benefits of a force-directed approach to generate comprehensible Euler diagrams for any set relations in relatively fast time.

**Index Terms**—Euler diagram, Venn diagram, force-directed.

## I. INTRODUCTION

Euler diagrams can represent containment, exclusion and intersection among data sets using closed curves [10]. They are widely used in various areas (e.g., genetics [20]; ontologies [15]), and automatic diagram drawing techniques have been devised (e.g., [27; 30]). A number of visual languages use Euler diagrams as a basis (e.g., Euler/Venn diagrams [32]; Venn-II diagrams [28]; constraint diagrams [18]; see survey [29]).

The closed curves facilitate reasoning about sets as they have a strong perceptual organizational effect on humans in dividing the space into regions and communicating memberships [23]. However, the curves have to be smooth and not too close to one another [2], highly symmetrical, and when possible, circles [3]. An Euler diagram should be *well-matched* [4], such that the regions in the diagram correspond exactly to the required set relations. If possible, an Euler diagram should also be *well-formed* [26], such that: each set is depicted by exactly one curve; each set relation is depicted by exactly one region; the curves are simple, non-concurrent and cross when they meet; and no point is on more than two curves. Nonetheless, generating an Euler diagram that satisfies all of these criteria is not always possible [24].

The well-matched diagrams produced by current methods (e.g., [27]) often have non-smooth, non-symmetric curves that are not easily distinguishable, as in Fig. 1. Other methods use circles to ensure curve smoothness and symmetry (e.g., [30]), but the generated diagrams are not well-matched and some of the regions might not correspond to any of the required set relations. Alternatively, some methods draw only well-formed Euler diagrams (e.g., [11]), but the curves are often non-

smooth and a diagram cannot be drawn for all data. Also, the importance of different aesthetic criteria varies by context and data.

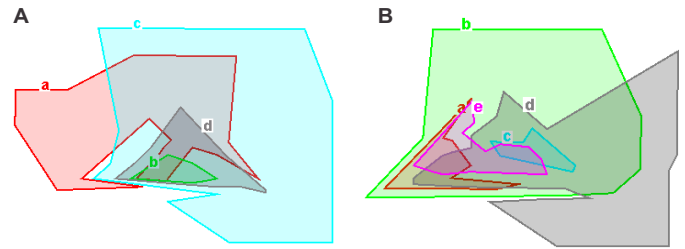


Fig. 1. Well-matched Euler diagrams generated by a drawing method [27].

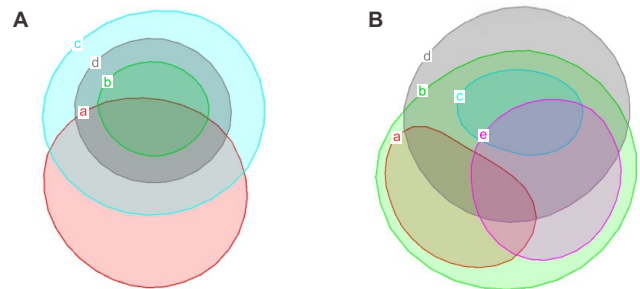


Fig. 2. The improved layouts generated by our force-directed method, *eulerForce*, for the Euler diagrams in Fig. 1.

Using a layout method, the diagram is transformed into another that depicts the same set relations, but optimizes specific aesthetic criteria. Two such methods, one by Rodgers et al. [25] and another by Flower et al. [14], have been proposed, but both are computationally expensive.

Rodgers et al. defined (but did not implement) a method that uses graph transformations to generate a layout that satisfies a particular well-formedness property [25]. However, this method does not take into account important curve aesthetics such as regularity, smoothness and symmetry and so, it cannot improve the layout of diagrams like those in Fig. 1, which are already well-formed. Graph transformations could also be computationally expensive [9].

Flower et al. implemented a method that uses a multi-criteria optimization technique to improve curve aesthetics [14]. They defined metrics to handle curve roundness, smoothness, closeness and size uniformity, and combined them in a fitness function. Thus, this method could improve the

layout of diagrams like Fig. 1A, but not Fig. 1B as their method handles diagrams with up to four curves. The effectiveness and correctness of these aesthetic metrics were not evaluated, and it is still unclear how the different metrics interact. The method uses a hill-climbing heuristic and thus, it is likely to encounter local minima and provide a local rather than a global best-optimized solution. The method is slow, as multi-criteria optimizations are more computationally expensive than single-criteria ones [21]. Assigning appropriate weights to the various criteria is difficult [21] and expecting users to assign these weights makes the method unusable.

In graph drawing, force-directed methods have been widely used and evaluated to produce layouts with desired aesthetic features with relatively good performance [5; 19]. The physical analogy used by such methods is that of a system of physical structures (the vertices of the graph) that exert a force over others in the system, such that these structures move according to the force applied to them. The system is brought to a halt when the algorithm positions the structures appropriately so that the forces are in equilibrium. One of the simplest force-directed methods is the spring embedder [6]. In such methods, the forces result from electrically charged particles (the vertices) that repel one another, so that the vertices are not too close to each other, and springs (the edges between vertices) that attract connected particles, so that the length of the edges is approximately uniform.

A closed curve represented as a polygon is like a graph with a set of vertices and edges, so the repulsive and attractive forces used in a spring embedder for graph drawing would transform a closed curve into a smooth regular circle. Thus, if such forces are applied to all the curves in a diagram and other new forces are applied to ensure that the required curve intersections are maintained, the diagrams in Fig. 1 would be converted to those in Fig. 2, so the curves are smooth, more regular and evenly distributed. The diagram layouts in Fig. 2 were generated by our method *eulerForce*, which is the first to use a force-directed approach to improve the curve aesthetics and layout of Euler diagrams.

In this paper, we describe *eulerForce*, the force model and algorithm it uses to improve the diagram layouts, and our evaluation of the method. The implementation of *eulerForce* is available at <http://www.eulardiagrams.org/eulerForce>.

## II. THE FORCE MODEL AND ALGORITHM

The main challenge was to devise an appropriate force model that acts on the vertices, edges and curves in the diagram to improve the layout of Euler diagrams while still depicting the same set relations. This is the first force model for Euler diagrams, so we opted for a simple algorithm to equilibrate the forces. This facilitates understanding of the different forces and how they interact with one another to allow for further refinement of the force model.

### A. Force Model

Our physical system is similar to that of the simple spring embedder (Section I), in that the vertices act like electrically charged particles and the edges like springs. The force model consists of *repulsive* and *attractive* forces between different

structures in the layout, including (i) vertices, (ii) edges and (iii) entire polygons. Thus, the forces in our system differ from those used in simple graph drawing methods by systematically moving any of these structures rather than just the vertices.

Similar to the typical spring embedder in graph drawing, our repulsive forces follow the inverse square law and our attractive forces follow the Hooke's law [5]. Thus, given  $d$  is the Euclidean distance between two structures  $s_1$  and  $s_2$  in the diagram, these forces are defined as follows: *repulsive forces* – inversely proportional to the squared distance between structures  $s_1$  and  $s_2$ , so the repulsive force between  $s_1$  and  $s_2$ , that is the repulsive force exerted on  $s_2$  by  $s_1$  and on  $s_1$  by  $s_2$ , is  $f_r = c_r/d^2$  where  $c_r$  is a constant that determines the strength of the force; *attractive forces* – directly proportional to the distance between structures  $s_1$  and  $s_2$  so the attractive force exerted between  $s_1$  and  $s_2$ , that is the attractive force exerted on  $s_1$  and  $s_2$  by the spring between  $s_1$  and  $s_2$ , is  $f_a = c_a d$  where  $c_a$  is the stiffness of the spring that determines the strength of the force and the natural length of the spring is zero. The constants  $c_r$  and  $c_a$  vary depending on the objective and the required strength of the force. In specific cases, the definition of the repulsive or attractive force could defer from those above, yet the direction remains unchanged.

Our repulsive forces are the same as those used in Eades' spring embedder [6]. Our attractive forces are different from those of Eades, as Eades uses logarithmic rather than linear (Hooke's law) springs stating that the latter could be too strong. However, Di Battista et al. argue that, "it is difficult to justify the extra computational effort by the quality of the resulting drawings" [5]. Since our attractive forces assume linear, Hooke's law springs with natural length zero, they are the same as those used in Tutte's force-directed barycentre method [33]. We opted for such attractive forces as these forces are namely used to smooth the curves and to regain regions that are lost during the layout improvement process. Thus, while in the former the edges should be as short as possible to produce smooth curves, in the latter the force of the spring should be strong enough to attract structures and regain the lost regions.

We now discuss how such repulsive and attractive forces between vertices, edges and polygons are used in our force model to generate layouts that meet our objectives (in bold).

### Obtaining regular, smooth, similarly shaped convex curves

We use typical forces for a simple spring embedder [5].

**(F1)** Repulsion for vertices not to be too close to one another: for every polygon  $p$  in the current layout and for every pair of distinct vertices  $v_1$  and  $v_2$  of  $p$ , a repulsive force is exerted between  $v_1$  and  $v_2$ , so  $v_1$  and  $v_2$  move away from one another.

**(F2)** Attraction for approximately uniform edge lengths: for every polygon  $p$  in the current layout and for every pair of distinct vertices  $v_1$  and  $v_2$  of  $p$  that are connected by an edge, an attractive force is exerted between  $v_1$  and  $v_2$ , so  $v_1$  and  $v_2$  move closer to one another.

### Maintaining the same set of regions as that in the initial diagram layout

We devised a set of forces for each different type of curve relation to ensure that: (a) the current improved layout maintains the regions in the initial layout; (b) if the current layout has new regions or is missing any of the regions

in the initial layout, forces correct the layout accordingly. We opted to use forces to correct layouts that depict the incorrect set of regions rather than to disallow such layouts altogether, to avoid local minima. So for every pair of distinct polygons in the initial layout, the following forces are applied.

**(F3)** If the two polygons in the initial layout *do not intersect*, and in the current layout they still *do not intersect*, if  $p_1$  and  $p_2$  are these two polygons in the current layout, for every vertex  $v_1$  of  $p_1$  and for every vertex  $v_2$  of  $p_2$ , a repulsive force is exerted between  $v_1$  and  $v_2$ , so these vertices move accordingly and the required disjointness of  $p_1$  and  $p_2$  is reinforced.

**(F4)** If the two polygons in the initial layout *do not intersect*, but in the current layout they *do intersect*, if  $p_1$  and  $p_2$  are these two polygons in the current layout, for every vertex  $v_1$  of  $p_1$  and vertex  $v_2$  of  $p_2$ : if  $v_1$  is inside or on an edge of  $p_2$  and  $v_2$  is inside or on an edge of  $p_1$ , an attractive force is exerted between  $v_1$  and  $v_2$ ; if  $v_2$  is not inside or on an edge of  $p_1$ , a *repulsive force* is exerted on  $v_1$  by  $v_2$ ; if  $v_1$  is not inside or on an edge of  $p_2$ , a *repulsive force* is exerted on  $v_2$  by  $v_1$ . As these vertices move accordingly, the required disjointness of  $p_1$  and  $p_2$  is regained.

**(F5)** If the two polygons in the initial layout *intersect*, and in the current layout they still *intersect*, if  $p_1$  and  $p_2$  are these two polygons in the current layout, for every vertex  $v_1$  of  $p_1$  and for every vertex  $v_2$  of  $p_2$ : if both  $v_1$  and  $v_2$  are on the boundary of the overlapping region, that is  $v_1$  is inside  $p_2$  and  $v_2$  is inside  $p_1$ , a repulsive force is exerted between  $v_2$  and  $v_1$ , so these vertices move accordingly and the required intersection of  $p_1$  and  $p_2$  is reinforced; if  $v_1$  is not inside  $p_2$  and  $v_2$  is inside or on an edge of  $p_1$ , a repulsive force is exerted on  $v_1$  by  $v_2$ , so these vertices move accordingly and  $p_1$  and  $p_2$  are not too close to one another; if  $v_2$  is not inside  $p_1$  and  $v_1$  is inside or on an edge of  $p_2$ , a repulsive force is exerted on  $v_2$  by  $v_1$ , so these vertices move and  $p_1$  and  $p_2$  are not too close to one another.

**(F6)** If the two polygons in the initial layout *intersect*, but in the current layout they *do not intersect*, if  $p_1$  and  $p_2$  are these two polygons in the current layout, for every vertex  $v_1$  of  $p_1$  and vertex  $v_2$  of  $p_2$ , a special attractive force defined as  $f = c/d^2$  where  $c$  is a constant determining the strength of the force and  $d$  is the Euclidean distance between  $v_1$  and  $v_2$  is exerted between  $v_1$  and  $v_2$ , so these vertices move accordingly and the required intersection of  $p_1$  and  $p_2$  is regained.

**(F7)** If in the initial layout one of the polygons *contains* the other and in the current layout the polygons still *depict* the required *containment*: if  $p_1$  and  $p_2$  are these two polygons in the current layout and  $p_2$  is contained in  $p_1$ , for every vertex  $v_1$  of  $p_1$  and for every vertex  $v_2$  of  $p_2$ , a repulsive force is exerted between  $v_1$  and  $v_2$ , so these vertices move accordingly and the required containment of  $p_2$  in  $p_1$  is reinforced.

**(F8)** If, in the initial layout, one of the polygons *contains* the other, but in the current layout, the polygons *do not depict* the required *containment*, if  $p_1$  and  $p_2$  are these two polygons in the current layout and according to the initial layout,  $p_2$  should be contained in  $p_1$ , for every vertex  $v_1$  of  $p_1$  and vertex  $v_2$  of  $p_2$ : if  $v_1$  is inside or on an edge of  $p_2$  and  $v_2$  is not inside or on an edge of  $p_1$ , an attractive force is exerted between  $v_2$  and  $v_1$ ; if  $v_2$  is inside or on an edge of  $p_1$ , a repulsive force is exerted on  $v_1$  by  $v_2$ ; if  $v_1$  is not inside or on an edge of  $p_2$ , an attractive force

is exerted on  $v_2$  from  $v_1$ . As these vertices move accordingly, the required containment of  $p_2$  in  $p_1$  is regained.

F3-F8 are applied between vertices of polygons to (a) maintain the regions of the initial layout and (b) correct layouts that are not depicting the same set of regions as that of the initial layout. However, to ensure (a) and reduce the need for (b), if a vertex  $v_1$  of polygon  $p_1$  is closer to a point  $x$  on an edge  $e = (v_2, v_{2b})$  of a polygon  $p_2$  than vertex  $v_2$  of  $p_2$ , F3-F8 are also applied between  $v_1$  and  $e$ , such that  $e$  is moved based on the forces exerted on it about  $x$ .

**Depicting each set relation by exactly one region** As the vertices are moved during the layout improvement process, a region depicting a set relation could be split up into more than one component, making the diagram difficult to comprehend as one of the most important well-formedness properties is not met [26]. Thus, for every pair of distinct polygons,  $p_1$  and  $p_2$ , in the current layout and for every region  $r$  in any or both of  $p_1$  and  $p_2$ : **(F9)** while  $r$  is made up of more than one component, if  $k$  is the smallest component of  $r$ , for every vertex  $v_1$  of  $p_1$  and vertex  $v_2$  of  $p_2$ , if  $v_1$  is inside or on an edge of  $k$  and  $v_2$  is not inside or on an edge of  $k$ , an attractive force is exerted between  $v_1$  and  $v_2$ , so these vertices move accordingly and a component of  $r$  is discarded.

**Ensuring the curves are not close to one another** Layouts with curves close to one another are difficult to comprehend [2] and could break the important wellformedness property of non-concurrent curves [26]. The repulsive forces in our model keep the vertices apart and thus aid to achieve this objective.

**Centring contained curves in their containing curve or region** Sometimes a curve is contained in another curve or a region. The repulsive forces in the model would ensure that this contained polygon remains inside the containing polygon or region. However, centring this contained polygon in its containing polygon or region, so that its boundary is equidistant from that of the containing structure, could improve the layout and its symmetry. Thus, **(F10)** when a polygon is contained in another polygon or region, if  $c_1$  is the centroid of the contained polygon and  $c_2$  is the centroid of the containing polygon or region, an attractive force is exerted on  $c_1$  from  $c_2$ , so that the entire contained polygon is moved closer to  $c_2$  and centred in its containing polygon or region.

**Attaining adequately sized curves and regions** If the size of the regions is inadequate, the layout could be difficult to understand, particularly when regions are not easily visible and their area is disproportional to that of other regions [2]. Thus, a set of forces is required to adjust the size of the polygons and to move these polygons closer or further away from one another, so the required adequate region areas are obtained.

An adequate region area could be one that is similar to the area of other regions in the layout, so that the total area of the diagram is evenly distributed among its regions [2]. However, to facilitate the identification of the number of curves in which a region is located, an adequate region area could be one that is inversely proportional to the number of curves in which it resides, in that the greater the number of curves it is located in, the smaller the region area. So, if a  $k$ -curve region is a region

located in  $k$  curves in a diagram with  $n$  curves, the area of the region is assigned a weight  $w=n/k$ . Thus, if for instance a diagram has three curves ( $n=3$ ), a 1-curve region ( $k=1$ ,  $w=3$ ) will be twice as large as a 2-curve region ( $k=2$ ,  $w=3/2$ ) and three times as large as a 3-curve region ( $k=3$ ,  $w=1$ ).

The size of the polygons are adjusted accordingly by progressively increasing or decreasing the strength of the repulsive force  $F_1$  that ensures that the vertices of polygons are not too close to one another. The greater the repulsive force, the further away neighbouring vertices of a polygon are from one another, thus enlarging the size of the polygon. The polygons are then moved using the following forces to adjust the region areas. **(F11)** To increase a region area: if  $r$  is the region whose area should be increased and  $c_1$  is the centroid of  $r$ , for every polygon  $p$  that contains  $r$ , if  $c_2$  is the centroid of  $p$ , an attractive force is exerted on  $c_2$  from  $c_1$ , so that the entire polygon  $p$  is moved closer to  $c_1$ , thus increasing its size. **(F12)** To decrease a region area: if  $r$  is the region whose area should be decreased and  $c_1$  is the centroid of  $r$ , for every polygon  $p$  that contains  $r$ , if  $c_2$  is the centroid of  $p$ , a repulsive force is exerted on  $c_2$  from  $c_1$ , so that the entire polygon  $p$  is moved further away from  $c_1$ , thus decreasing the size of  $r$ .

Similar to F3-F8, other forces have been included to correct any generated layouts whose regions differs from those in the initial layout, either because new regions are displayed or required regions are missing. We could have disallowed these incorrect layouts from the layout improvement process altogether, but we opted to accept them and correct them using the following forces, to reduce the chances of reaching a local minimum. Thus, if while increasing or decreasing region area, **(F13)** the current layout has a region that is not depicted in the initial layout: if  $r$  is the region that is in the current but not the initial layout and  $c_1$  is the centroid of  $r$ , for every polygon  $p$  that contains  $r$  in the current but not in the initial layout, if  $c_2$  is the centroid of  $p$ , a repulsive force is exerted on  $c_2$  from  $c_1$ , so the entire polygon  $p$  is moved further away from  $c_1$ , thus reducing the size of  $r$  and its appearance in the layout until it is no longer visible. If alternatively **(F14)** the current layout does not have a region that is depicted in the initial layout: if  $r$  is the region that is in the initial but not the current layout, for every pair of distinct polygons  $p_1$  and  $p_2$  that should contain  $r$ , if  $c_1$  is the centroid of  $p_1$  and  $c_2$  is the centroid of  $p_2$ , an attractive force is exerted between  $c_1$  and  $c_2$ , so the polygons that should contain  $r$  get closer and the missing region is regained.

### B. Algorithm

Our algorithm is similar to that used by Eades [6] to balance out the forces in the system. Given some set relations, an Euler diagram is generated by a current automatic drawing method and used as the initial layout. The algorithm then goes through the system in discrete time steps, so that at every step, the resultant force exerted on each of the vertices, edges and entire polygons in the layout is calculated and the vertices, edges and entire polygons are moved accordingly based on the magnitude and the direction of the resultant force. This new layout is then used as the starting layout for the next discrete

time step. After a number of steps, the magnitude of the resultant force exerted on each of the vertices, edges and entire polygons is reduced to zero and the algorithm stops as the forces in the system equilibrate and no further changes in the layout are carried out.

Since most of the forces in the system are exerted on and relocate the vertices of the polygons in the layout, polygons with fewer vertices are subject to fewer changes than those with more vertices. Thus, before the algorithm goes through the system in discrete time steps, the number of vertices on each of the polygons in the layout is equalized. For instance, if a layout has two polygons  $p_1$  and  $p_2$ , and  $p_1$  has 10 vertices and  $p_2$  has 12 vertices, two vertices are added to  $p_1$ . This is done by first adding a vertex  $x$  between two vertices  $v_1$  and  $v_2$  of the polygon that are connected by an edge  $(v_1, v_2)$  and then, removing  $(v_1, v_2)$  and adding two new edges  $(v_1, x)$  and  $(x, v_2)$  between  $v_1$  and  $x$  and  $x$  and  $v_2$  respectively. Since the forces in the system can enlarge the size of the polygons, at the end of every discrete time step, the length of the edges of each polygon is checked and vertices are added to make the edges smaller and the polygons smoother.

Due to the various forces in the system, a limit is set on the magnitude of the resultant force exerted on a structure. This limit is inversely proportional to the number of discrete time steps the algorithm has already gone through in the system, so major changes are only carried out at the initial steps when a more extensive search for an appropriate layout is required. During the final steps, minor changes are carried out to refine the layout and ensure the algorithm converges to a solution.

The transition from the initial to the final layout is animated, thus facilitating understanding of how the forces in the system aid in improving the layout and how they interact with one another [5]. This method was thus helpful to understand and appropriately define the required forces to lay out Euler diagrams and to devise the first force model to improve the layout of such diagrams. Moreover, such a simple algorithm could possibly aid in preserving the mental map of the layout [7] from the initial to the final improved layout.

Eades's simple spring embedder [6] was aimed for non-dense graphs with few vertices. Poor layouts by this embedder are reported for graphs with hundreds of vertices [19], as in such cases a local minimum is more likely to be reached. As discussed earlier, we mitigate this issue by using specific forces that correct generated layouts that depict different regions than those in the initial layout. Even so, Euler diagram layouts typically have fewer than hundreds of vertices as often these diagrams have few curves. Later on, further sophisticated techniques can be adopted to handle more specific aesthetic criteria and to improve the efficiency and performance of our force-directed algorithm.

## III. EVALUATION

To evaluate our method *eulerForce*, we used its software implementation to improve the layouts of Euler diagrams generated by a current drawing method [27], and we compared *eulerForce*'s layouts with those generated by the only other implemented layout method for Euler diagrams [14]. All the

experiments were run on an Intel Core 2 Duo CPU E7200 @2.53GHz with 3.23GB RAM, 32-bit Microsoft Windows XP Professional SP1, SP2 and SP3 and Java Platform 1.6.0.14.

#### A. Accuracy, Time and Aesthetics

We tested *eulerForce* on diagrams automatically generated by Rodgers et al.'s method [27], to evaluate its effectiveness in generating improved layouts that satisfy our objectives. Rodgers et al.'s method was chosen, as it is the only method that draws a diagram for set relations for which a well-matched, well-formed Euler diagram can be drawn. Thus, if an improved layout generated by *eulerForce* did not satisfy our objective of depicting each set relation by exactly one region or our objective of ensuring the curves are not close to one another, the diagram layout was not well-formed and a limitation in our method was evident, as a well-formed diagram for those set relations is known to exist (i.e., the initial diagram generated by Rodgers et al.'s method).

A library of Euler diagrams generated by Rodgers et al.'s method for all the set relations for which a well-formed Euler diagram with three, four and five curves can be drawn was assembled. This library included: nine Euler diagrams with three curves, 114 Euler diagrams with four curves, and 342 Euler diagrams with five curves.

Our method *eulerForce* was then used to improve the layout of the diagrams in this library. Fig. 3–Fig. 5 illustrate a few of: (i) the diagrams in the library (also Fig. 1), and (ii) the corresponding layout generated by *eulerForce* (also Fig. 2). The layouts (ii) in Fig. 3 and Fig. 4 depict precisely the same set of regions as those in the initial library layout (i) (also Fig. 1 and Fig. 2), but those in Fig. 5 do not and are thus examples of cases where *eulerForce* fails to produce an appropriate layout. We now discuss these layouts and the results obtained.

**Accuracy** The improved layouts for all the nine and 114 diagrams with respectively three and four curves had the same regions as those of the initial incomprehensible layouts, and thus satisfied our objective of maintaining the same set of regions as that in the initial diagram layout. For the 342 diagrams with five curves, only 209 of the improved layouts (61%) satisfied our objective of maintaining the same set of regions as that in the initial diagram layout. The latter result could be due to the increased number of vertices that are unmanageable with a simple spring embedder [6; 19], particularly when the diagram has various regions.

Fig. 5A(ii) generated by *eulerForce* for the diagram and initial layout Fig. 5A(i) has two new unwanted regions, *abcd* and *abcde*, that are not depicted in the initial layout. Thus, curves *a* and *b* should be disjoint. Curve *a* in the final layout generated by *eulerForce* in Fig. 5A(ii) is not completed smooth as the forces that were specifically devised to correct layouts depicting regions that are different from the initial are striving to regain the disjointness between curves *a* and *b*. However, these forces seem to be weaker than other interacting forces in the system and thus, an inappropriate layout is generated. This also indicates the limitations of a simple spring embedder in managing various interacting forces in the system.

Fig. 5B(ii) generated by *eulerForce* for the diagram and initial layout Fig. 5B(i) has two missing required regions, *ad* and *be*, that are depicted in the initial layout and one new unwanted regions, *abcde*, that is not depicted in the initial layout. All the curves in the final layout generated by *eulerForce* in Fig. 5B(ii) are smooth and regular. However, the layout is not well-formed as there is a point on the three curves *a*, *b* and *e*. This example indicates the limitations of a simple spring embedder when a diagram has various regions. For various curve overlaps to be displayed, the curve will likely have to attain a less regular shape and thus, the strength of the forces, particularly those that aim at generating regular, smooth and similarly shaped convex curves, might have to be dynamically tuned using more sophisticated techniques. In fact, for region *abcde* not to be depicted in the diagram and for the diagram to be well-formed in that no point is on more than two curves, curves *b*, *c* and *e* should attain a more elongated shape rather than a circular shape, as in Fig. 5B(ii).

Thus, more sophisticated force-directed techniques such as those used for laying out large graphs (e.g., [16]) should be adopted for the algorithm to overcome local minima and to handle Euler diagrams with thousands of vertices and with various curves and regions.

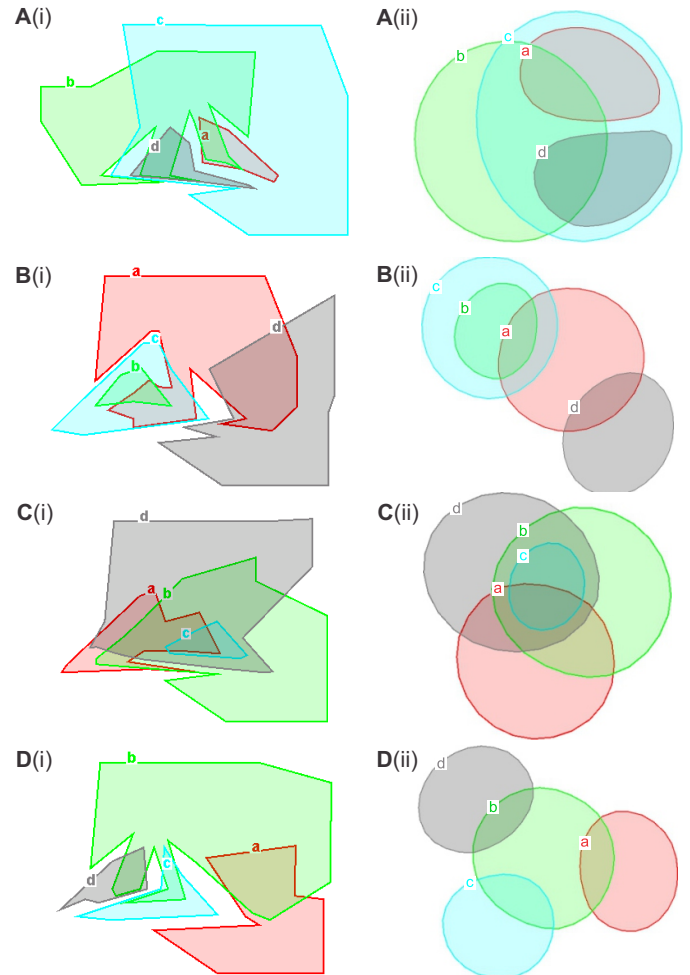


Fig. 3. Examples of (i) diagrams with four curves by Rodgers et al.'s method [27] in our library and (ii) the correct layouts by *eulerForce*.



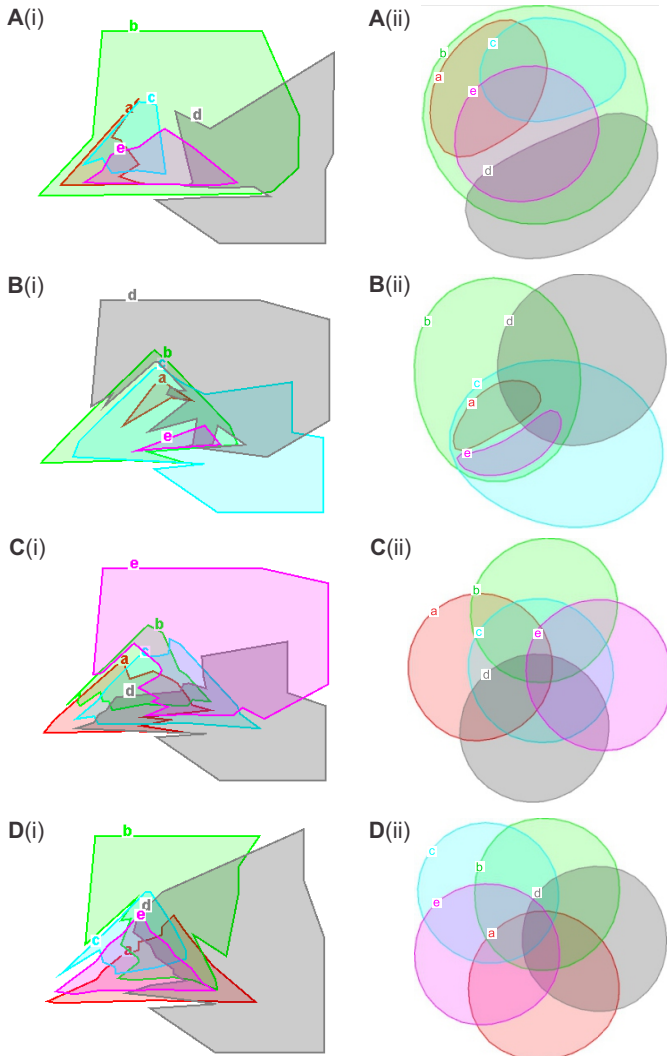


Fig. 4. Examples of (i) diagrams with five curves by Rodger et al.'s method [27] in our library and (ii) the correct layouts by *eulerForce*.

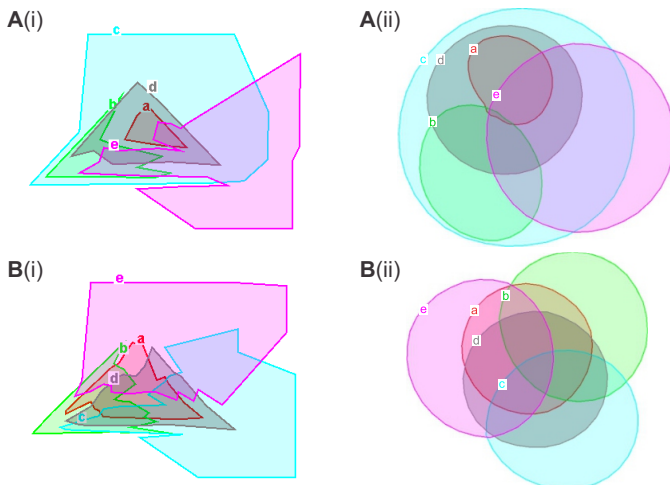


Fig. 5. Examples of (i) diagrams with five curves by Rodger et al.'s method [27] in our library and (ii) the incorrect layouts by *eulerForce*.

**Time** On average, the final improved layout for diagrams with three curves was generated by *eulerForce* in 7 seconds, those with four curves in 26 seconds, and others with five curves in 77 seconds. Thus, though our current method uses a simple algorithm, which is not as efficient as other more sophisticated alternatives, improved layouts are still generated in relatively fast time. This is comparable to force-directed approaches for graphs, which typically produce layouts in around a minute [19]. Also, a response time of 10 seconds or less ensures the users' attention is maintained [22]. However, better-optimized algorithms should be considered in future force-directed approaches for Euler diagram layouts.

**Aesthetics** As illustrated in the examples in Fig. 2–Fig. 4, the curves of all the generated layouts depicting the correct set of the regions were smooth. Also, whenever possible, the curves were regular, similarly shaped and convex, all of which facilitate understanding [2]. So *eulerForce* satisfies our objective of obtaining regular, smooth, similarly shaped convex curve. Similarly, the curves of all the generated layouts depicting the correct set of the regions were well-formed and satisfied the most important well-formedness properties of regions made up of at most one component and non-concurrent curves, as in Fig. 2–Fig. 4. Even in diagrams with various curves contained in other curves or regions, as in Fig. 2, Fig. 3A–C and Fig. 4A–B, none of the curves are too close to one another. This could have been further facilitated by the forces that centre contained curves in their containing curve or region.

Layouts generated by a spring embedder are likely to be symmetric [8], as shown by most layouts in Fig. 2–Fig. 4. However, besides the basic forces that are typical for a spring embedder in graph drawing, other forces that we devised for Euler diagrams are likely to aid in generating symmetric layouts. In particular, the forces that centre contained curves in their containing curve or region aid in generating highly symmetric layouts, as Fig. 2, Fig. 3A–C and Fig. 4A–B.

Having adequately sized regions and curves also aid diagram comprehend [2]. The area of the diagram could be evenly distributed among its regions, but in our case we opted for an adequate region area that is inversely proportional to the number of curves in which it resides. The generated layouts including Fig. 2–Fig. 4 indicate that this approach is effective as it ensures that: curves contained in other curves or regions are not too large for them to fit appropriately in the containing curve or region with possibly other regions, as in Fig. 2, Fig. 3A–C and Fig. 4A–B, and without breaking well-formedness; the number of curves in which a region is located is easier to identify.

For the layouts to be effectively evaluated, formalized aesthetic metrics and cognitive measures are required. Very few studies have investigated the aesthetics of such diagrams (Section I), but no criteria have been formalized.

#### B. *eulerForce* versus Previous Methods

The only previous layout method that has been implemented is Flower et al.'s multi-criteria optimization method [14]. We compared the diagram layouts generated by Flower et al.'s method with those generated by *eulerForce*.

As initial layouts, Flower et al. used diagrams generated by techniques [12; 13] available at the time. Fig. 6A(i) and Fig. 6B(i) illustrate diagrams generated by these techniques. The technique we used to generate the initial layouts for *eulerForce* [27] is more recent, but yet a variant of those used by Flower et al. for their method.

Given sets  $a, b, c, d$  and the set relations  $\{\emptyset, a, c, ac, cd, acd, bcd, abcd\}$ , Flower et al.'s initial layout is Fig. 6A(i) and the generated improved layout is Fig. 6A(ii), while *eulerForce*'s initial layout is Fig. 1A and the generated improved layout is Fig. 2A. Flower et al.'s initial and final layout look similar as the position and the orientation of the curves is barely changed, indicating that the method is limited to a minimal local search leading to a layout whose aesthetics could be improved further. For instance, the layout generated by *eulerForce* has regular, similarly shaped, circular curves. The containing and contained curves  $c, d$  and  $b$  are centre aligned and the distance between curve  $c$  and  $d$  is the same as the distance between curve  $d$  and  $b$ . All of these features further aid in indicating subsets in the data depicted by the diagram, thus facilitating data analysis. So, in contrast to Flower et al.'s layout, *eulerForce*'s layout is symmetric, compact, easy to understand and remember.

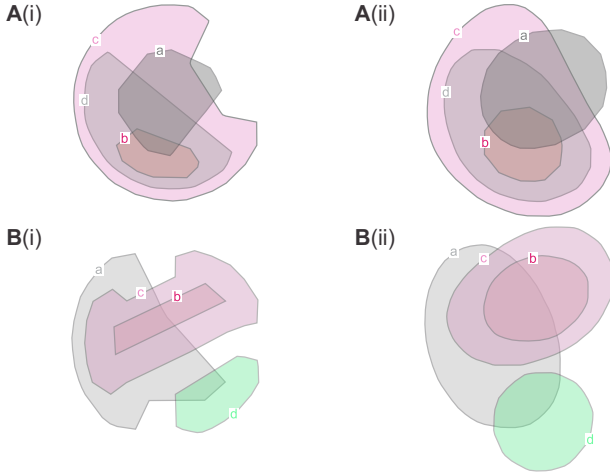


Fig. 6. The improved layouts (ii) generated by Flower et al.'s method [14] for the diagrams and initial layouts (i).

Similar observations are evident for the layouts depicting set relations  $\{\emptyset, a, c, d, ac, ad, bc, abc\}$  where Flower et al.'s initial layout is Fig. 6B(i) and the generated improved layout is Fig. 6B(ii), while *eulerForce*'s initial layout is Fig. 3B(i) and the generated improved layout is Fig. 3B(ii). Flower et al.'s final layout, Fig. 6B(ii), was generated after 80 iterations and after the line segments in the diagram were converted to Bézier curves. The final layout of *eulerForce*, Fig. 6B(ii), was generated in 17 seconds. So a layout improvement method using a force-directed approach as *eulerForce* could be faster than ones using multi-criteria optimization like Flower et al.'s method. After all, multi-criteria optimization is known to be computationally expensive [21]. In contrast to *eulerForce*, Flower et al.'s method is limited to diagrams with up to four

curves and thus, no layouts with more than four curves could be included in our comparative analysis.

Though the initial layouts used by *eulerForce* in our evaluation are less comprehensible than those used by Flower et al.'s method, the final improved layouts generated by *eulerForce* are more aesthetically desirable and easier to use than those generated by Flower et al.'s method. The effectiveness of the layouts should be evaluated using formalized aesthetic metrics and cognitive measures. However, none are available for Euler diagrams and so, our comparative analysis and evaluation of the layouts was limited to a visual comparison of the layouts and based on the findings of the very few studies on Euler diagram aesthetics [2; 3; 26]. Even though Flower et al. defined a few aesthetic metrics to devise their layout method [14], these metrics were not evaluated.

#### IV. CONCLUSION

In this paper, we have described our layout method, *eulerForce*, the first method that uses a force-directed approach to improve the layout of Euler diagrams. Our evaluation indicates great potential for using force-directed techniques to improve Euler diagram layouts in quick time and to generate comprehensible diagrams given the required set relations.

It would be interesting to evaluate the layouts generated by *eulerForce* for initial layouts that are not well-formed and for set relations for which a well-formed Euler diagram cannot be drawn. Until now, *eulerForce* has been evaluated for initial layouts that are well-formed and for set relations for which a well-formed diagram can be drawn. This was intentional to evaluate the effectiveness of the forces that we specifically devised to ensure that there is only one region for each set relation and that the curves are not too close to one another. However, the effectiveness of these forces in handling not well-formed diagrams should be evaluated, so that if necessary, the force model is adapted to handle such diagrams.

We adopted a simple spring embedder algorithm to facilitate understanding and evaluation of our force model, which is the first for Euler diagrams. However, this algorithm is not as efficient as other force-directed algorithms and is unable to handle hundreds of vertices [19]. Such limitations are evident in our *eulerForce* evaluation for Euler diagram layout with five curves, as discussed in Section III. Until now, our focus was on the force model rather than the algorithm. In the future, sophisticated force-directed algorithms such as those used for laying out large graphs [17] can be adopted and investigated in the context of Euler diagrams.

For instance, a multilevel approach such as that used in graph drawing [34] can be adopted to overcome local minima and to efficiently handle layouts with thousands of vertices and thus, with various curves and regions like those in Fig. 5. As an example, Hu's method [16] uses this approach to lay out graphs with over 10,000 vertices in less than a minute.

The Barnes-Hut algorithm [1] can be used to efficiently and dynamically compute the appropriate forces at every step of the layout improvement process. This method has already been successfully used in graph drawing (e.g., [16]) and could aid in cases such as those in Fig. 5. Force-directed techniques in

graph drawing have also demonstrated that adding magnetic fields to the system and its springs could aid in satisfying various aesthetic criteria [31] and should thus be considered for Euler diagram layouts.

Other future work includes gathering more empirical evidence to assess Euler diagram aesthetics and to formalize metrics that evaluate the effectiveness of Euler diagram layouts.

## REFERENCES

- [1] J. Barnes, and P. Hut, "A hierarchical  $O(N \log N)$  force-calculation algorithm," *Nature*, 324, 1986, Nature Publishing Group, pp. 446-449.
- [2] F. Benoy, and P. Rodgers, "Evaluating the comprehension of Euler diagrams," *Proceedings of the 11th International Conference on Information Visualization (IV)*, 2007, IEEE Computer Society, pp. 771-780.
- [3] A. Blake et al., "The Impact of Shape on the Perception of Euler Diagrams," *Proceedings of the 8th International Conference on the Diagrammatic Representation and Inference (Diagrams)*, in press, 2014, Springer.
- [4] P. Chapman et al., "Visualizing Sets: An Empirical Comparison of Diagram Types," *Proceedings of the 8th International Conference on the Diagrammatic Representation and Inference (Diagrams)*, *Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, in press, 2014, Springer.
- [5] G. Di Battista et al., "Force-Directed Methods," in *Graph drawing: algorithms for the visualization of graphs*, Prentice-Hall, Upper Saddle River, NJ, USA, 1999, pp.303-325.
- [6] P. Eades, "A Heuristic for Graph Drawing," *Congressus Numerantium*, 42, 1984, pp. 149-160.
- [7] P. Eades et al., *Preserving the Mental Map of a Diagram*, International Institute for Advanced Study of Social Information Science, Fujitsu Limited, Numazu-shi, Shizuoka, Japan, 1991.
- [8] P. Eades, and X. Lin, "Spring Algorithms and Symmetry," *Theoretical Computer Science*, 240, 2000, Elsevier, pp. 379-405.
- [9] H. Ehrig et al., *Handbook of Graph Grammars and Computing by Graph Transformation: Applications, Languages and Tools*, 2, World Scientific Publishing Co, River Edge, NJ, USA, 1999.
- [10] L. Euler, *Lettres à une Princesse d'Allemagne sur divers sujets de physique et de philosophie*, vol. 2, *Lettres 102-108*, L'Académie Impériale des Sciences de Saint-Petersbourg, St Petersburg, Russia, 1768.
- [11] J. Flower, A. Fish, and J. Howse, "Euler Diagram Generation," *Journal of Visual Languages & Computing*, 19, 2008, Elsevier, pp. 675-694.
- [12] J. Flower, and J. Howse, "Generating Euler Diagrams," *Proceedings of the 2nd International Conference on the Diagrammatic Representation and Inference (Diagrams)*, *Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)* 2317, 2317, 2002, Springer, pp. 285-285.
- [13] J. Flower, J. Howse, and J. Taylor, "Nesting in Euler diagrams," *Proceedings of the 1st International Workshop on Graph Transformation and Visual Modeling Techniques (GT-VMT 2002)*, *Electronic Notes in Theoretical Computer Science* vol. 72 no. 3, 72(3), 2003, Elsevier, pp. 93-102.
- [14] J. Flower, P. Rodgers, and P. Mutton, "Layout Metrics for Euler Diagrams," *Proceedings of the 7th International Conference on Information Visualization (IV)*, 2003, IEEE Computer Society, pp. 272-280.
- [15] J. Howse et al., "Visualizing ontologies: A case study," *Proceedings of the 10th International Semantic Web Conference (ISWC)*, 2011, Springer, pp. 257-272.
- [16] Y. Hu, "Efficient and High-Quality Force-Directed Graph Drawing," *Mathematica Journal*, 10, 2005, pp. 37-71.
- [17] Y. Hu, "Algorithms for Visualizing Large Networks," *Combinatorial Scientific Computing*, 2011,
- [18] S. Kent, "Constraint diagrams: visualizing invariants in object-oriented models," *Proceedings of the 12th ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA)*, 32(10), 1997, ACM, pp. 327-341.
- [19] S. G. Kobourov, "Spring Embedders and Force Directed Graph Drawing Algorithms," *Computing Research Repository (CoRR)*, abs/1201.3011, 2012,
- [20] L. P. Lim et al., "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs," *Nature*, 433, 2005, Nature Publishing Group, pp. 769-773.
- [21] R. T. Marler, and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and Multidisciplinary Optimization*, 26, 2004, Springer, pp. 369-395.
- [22] R. B. Miller, "Response time in man-computer conversational transactions," *Proceedings of the December 9-11, 1968 (AFIPS) fall joint computer conference, part I*, 1968, ACM, pp. 267-277.
- [23] S. E. Palmer, "Common region: A new principle of perceptual grouping," *Cognitive Psychology*, 24, 1992, Elsevier, pp. 436-447.
- [24] P. Rodgers, "A Survey of Euler Diagrams," *Journal of Visual Languages and Computing, Special Issue on Visualization and Reasoning using Euler Diagrams*, 25, 2014, Elsevier,
- [25] P. Rodgers et al., "Euler Graph Transformations for Euler Diagram Layout," *Proceedings of the 27th IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 2010, IEEE, pp. 111-118.
- [26] P. Rodgers, L. Zhang, and H. Purchase, "Wellformedness Properties in Euler Diagrams: Which Should Be Used?," *IEEE Transactions on Visualization and Computer Graphics*, 18, 2012, IEEE, pp. 1089-1100.
- [27] P. Rodgers et al., "Embedding Wellformed Euler diagrams," *Proceedings of the 12th International Conference on Information Visualization (IV)*, 2008, IEEE Computer Society, pp. 585-593.
- [28] S.-J. Shin, *The logical status of diagrams*, Cambridge University Press, New York, NY, USA, 1994.
- [29] G. Stapleton, "A survey of reasoning systems based on Euler diagrams," *Electronic Notes in Theoretical Computer Science*, 134, 2005, Elsevier, pp. 127-151.
- [30] G. Stapleton et al., "Automatically drawing Euler diagrams with circles," *Journal of Visual Languages & Computing*, 23, 2012, Elsevier, pp. 163-193.
- [31] K. Sugiyama, and K. Misue, "Graph Drawing by the Magnetic Spring Model," *Journal of Visual Languages and Computing*, 6, 1995, Elsevier, pp. 217-231.
- [32] N. Swoboda, "Implementing Euler/Venn reasoning systems," in *Diagrammatic Representation and Reasoning*, Springer, 2002, pp.371-386.
- [33] W. T. Tutte, "How to draw a graph," *Proceedings of the London Mathematical Society*, 13, 1963, Citeseer, pp. 743-768.
- [34] C. Walshaw, "A multilevel algorithm for force-directed graph drawing," *Proceedings of the 8th International Symposium on Graph Drawing (GD 2000)*, *Lecture Notes in Computer Science* 1984, 2001, Springer, pp. 171-182.

# A Normal Form for Spider Diagrams of Order

Aidan Delaney, Gem Stapleton, John Taylor  
University of Brighton, Brighton, BN2 4GJ, United Kingdom

Simon Thompson  
School of Computing, University of Kent, CT2 7NF, United Kingdom

## Abstract

We develop a reasoning system for an Euler diagram based visual logic, called spider diagrams of order. We define a normal form for spider diagrams of order and provide an algorithm, based on the reasoning system, for producing diagrams in our normal form. Normal forms for visual logics have been shown to assist in proving completeness of associated reasoning systems. We wish to use the reasoning system to allow future direct comparison of spider diagrams of order and linear temporal logic.

## 1 Introduction

Shin's rebirth of Peirce's  $\alpha$  and  $\beta$  systems for reasoning [14] has produced a variety of Euler diagram based visual logics, for example [6, 4, 16, 17]. Euler diagram based visual logics allow reasoning about sets, their elements and their relationships. Associated with visual logics are reasoning systems that embody equivalence between diagrams [2, 9, 18]. Spider Diagrams of Order (SDoO) and Second-Order Spider Diagrams [3] differ from the main body of work on Euler diagram based logics as elements of their token syntax were designed to be as expressive as star-free regular languages and regular languages respectively.

Weakly expressive language classes, such as regular languages and star-free regular languages, are used to formalise real-world temporal specifications [5, 11]. Due to the real-world application there has been recent interest in incorporating temporal semantics in these diagrammatic logics [1, 13]. In this paper we address the problem of adding temporal semantics to Euler diagrams by adding a syntax and semantics for specifying order of the elements. Furthermore, we develop the first reasoning system for an Euler diagram based logic that includes an order relation. In demonstrating our reasoning system for spider diagrams of order we produce both a normal form and an algorithm to produce the normal form. Our algorithm also contributes to

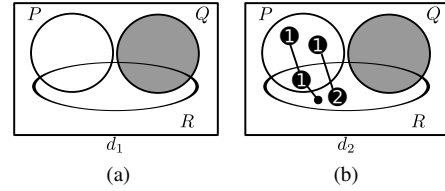


Figure 1: An Euler and a spider diagram.

the recent interest in normal forms for Euler diagram based logics [8].

In section 2 we define the syntax and semantics of spider diagrams of order. In section 3 we present each of our reasoning rules. Thereafter, in section 4 we present our normalisation algorithm by example. An implementation of our algorithm is available under an open-source license at <https://github.com/AidanDelaney/SpiderReasoning>.

## 2 Spider Diagrams of Order

The Euler diagram in Fig. 1(a) contains three labelled contours and six zones. A **zone** is defined to be a pair,  $(in, out)$ , of disjoint subsets of the set of contour labels. The set  $in$  contains the labels of the contours that the zone is inside whereas  $out$  contains the labels of the contours that the zone is outside. The set of all zones is denoted  $\mathcal{Z}$ . A **region** is a set of zones. As an example, there exists a zone inside the contour  $P$  but outside both contours  $Q$  and  $R$  denoted  $(\{P\}, \{Q, R\})$ . The zone inside the bounding box and outside all contours can be described by being inside  $\emptyset$  and outside  $\{P, Q, R\}$ . We note that there is no zone in the diagram corresponding to  $(\{P, Q\}, \{R\})$  i.e. there is no zone inside contours labelled  $P$  and  $Q$  but outside the contour labelled  $R$ . Euler diagrams may be conjoined using the symbol  $\wedge$ , disjoined using the symbol  $\vee$  or negated using the symbol  $\neg$ .

A **Spider Diagram of Order** is an Euler diagram containing one or more graphs. The vertices of a graph are labelled with ‘•’ or an integer. A graph is restricted such that it is acyclic and may not have more than one vertex of a given label in a given zone. To maintain consistency with the literature we call graphs of this form **spiders** and term a vertex within a graph to be a **spider foot**. The diagram in Fig 1(b) contains two spiders, one spider consisting of three feet labelled ‘1’, ‘1’ and ‘•’ the other spider contains two feet labelled ‘1’ and ‘2’. In the following definition, as throughout the paper, we use  $\cup$  to mean set union,  $\cap$  to mean set intersection and  $A - B$  denotes the set difference between  $A$  and  $B$ .

**Definition 1** A **spider foot** is an element of the set  $(\mathbb{Z}^+ \cup \{\bullet\}) \times \mathcal{Z}$  and the set of all feet is denoted  $\mathcal{F}$ . A spider foot  $(k, z) \in \mathcal{F}$  where  $k \in \mathbb{Z}^+ \cup \{\bullet\}$  has **rank**  $k$ . The rank of a spider foot induces a relation  $<$  on the feet, defined by  $(k_1, z_1) < (k_2, z_2)$  if both  $k_1, k_2 \in \mathbb{Z}^+$  and  $k_1 < k_2$  hold or  $k_1 = \bullet$  or  $k_2 = \bullet$ .

Whilst it may seem strange that  $<$  as just defined is not a strict ordering (because  $\bullet$  is both less than and greater than all other feet) this choice of  $<$  simplifies many definitions.

**Definition 2** A **spider**,  $s$ , is a non-empty set of feet together with a positive natural number, that is  $s \in \mathbb{Z}^+ \times (\mathbb{P}\mathcal{F} - \{\emptyset\})$ , and the set of all spiders is denoted  $\mathcal{S}$ . The set  $p$  is the **foot set** of spider  $s = (n, p)$ . The **habitat** of a spider  $s = (n, p)$  is the region  $\eta(s) = \{z : (k, z) \in p\}$ .

Formally, the set of all contour labels is denoted  $\mathcal{C}$ .

**Definition 3** A **unitary spider diagram of order**,  $d$ , is a quadruple  $\langle C, Z, ShZ, SI \rangle$  where:

$C = C(d) \subseteq \mathcal{C}$  is a finite set of contour labels,

$Z = Z(d) \subseteq \{(in, C - in) : in \subseteq C\}$  is a set of zones,

$ShZ = ShZ(d) \subseteq Z(d)$  is a set of shaded zones,

$SI = SI(d) \subsetneq \mathcal{S}$  is a finite set, called the **spider identifiers**, such that for all spiders  $(n_1, p_1), (n_2, p_2) \in SI(d)$  if  $p_1 = p_2$  then  $n_1 = n_2$ .

The set of **spiders** in  $d$  is defined to be

$$S(d) = \{(i, p) : (n, p) \in SI(d) \wedge 1 \leq i \leq n\}.$$

The symbol  $\perp$  is also a unitary spider diagram. We define  $C(\perp) = Z(\perp) = ShZ(\perp) = SI(\perp) = \emptyset$ .

Spider diagrams of order may also be combined using the Boolean operations  $\wedge, \vee$  and  $\neg$ . In addition we allow the binary connective  $\triangleleft$ . A spider diagram of order that contains one of the  $\wedge, \vee, \neg$  or  $\triangleleft$  connectives is a **compound**

diagram. Furthermore, a spider diagram of order containing either no spiders or containing spiders consisting of only single feet is an  $\alpha$ -**diagram**. A zone can be considered to be *missing* from a spider diagram as presented in [9].

**Definition 4** Given an Euler diagram,  $d$ , a zone  $(in, out)$  is said to be **missing** if it is in the set  $\{(in, C(d) - in) : in \subseteq C(d)\} - Z(d)$  with the set of such zones denoted  $MZ(d)$ . If  $d$  has no missing zones then  $d$  is in **Venn form**.

Spider diagrams of order have a model based semantics.

**Definition 5** An **interpretation** is a triple  $(U, \prec, \Psi)$  where  $U$  is a universal set and  $\Psi : \mathcal{C} \rightarrow \mathbb{P}U$  is a function that assigns a subset of  $U$  to each contour label and  $\prec$  is a strict total order on  $U$ . The function  $\Psi$  can be extended to interpret zones and sets of regions as follows:

1. each zone,  $(in, out) \in \mathcal{Z}$ , represents the set

$$\Psi(z) = \bigcap_{c \in in} \Psi(c) \cap \bigcap_{c \in out} (U - \Psi(c)) \text{ and}$$

2. each region,  $r \in \mathbb{P}\mathcal{Z}$ , represents the set which is the union of the sets represented by  $r$ 's constituent zones, that is

$$\Psi(r) = \bigcup_{z \in r} \Psi(z).$$

**Definition 6** Let  $I = (U, \prec, \Psi)$  be an interpretation and let  $d (\neq \perp)$  be a unitary spider diagram. Then  $I$  is a **model** for  $d$ , denoted  $I \models d$ , if and only if the following conditions hold.

1. **The missing zones condition** All of the missing zones represent the empty set, that is,  $\bigcup_{z \in MZ(d)} \Psi(z) = \emptyset$ .
2. **The spider mapping condition** There exists an injective function,  $\varphi : S(d) \rightarrow U$  and a function  $f : S(d) \rightarrow \mathcal{F}$ , called a **valid pair**, such that the following conditions hold:

- (a) **The selected foot condition** Each spider  $s$  must map, under  $f$ , to a spider foot in its foot set:

$$\forall (n, p) \in S(d) \ f(n, p) \in p.$$

- (b) **The spiders' location condition** All spiders represent elements in the sets represented by the zone in which the selected foot, under  $f$ , is placed:

$$\forall s \in S(d) \ (f(s) = (k, z) \Rightarrow \varphi(s) \in \Psi(z)).$$

- (c) **The shading condition** Shaded regions represent a subset of elements denoted by spiders:

$$\forall z \in ShZ(d) \ \Psi(z) \subseteq im(\varphi).$$

- (d) **The order condition** The ordering information provided by the selected spider feet agrees with that provided by the strict order relation.

That is,

$$\forall s_1, s_2 \in S(d) (\varphi(s_1) \prec \varphi(s_2) \Rightarrow f(s_1) < f(s_2)).$$

If  $d = \perp$  then no interpretation is a model for  $d$ .

The conjunction of conditions 1 and 2 above is the **semantics predicate** for spider diagrams of order. The semantics of the connectives  $\wedge$ ,  $\vee$  and  $\neg$  extend in the obvious manner, however the semantics of  $\triangleleft$  requires some explanation.

The  $\triangleleft$  operation allows the specification of an order between unitary diagrams. In order to define the semantics of compound spider diagrams of order involving  $\triangleleft$  we present the definition of ordered sum of interpretations [7].

**Definition 7** The **ordered sum** of two interpretations  $m_1 = (U_1, \prec_1, \Psi_1)$  and  $m_2 = (U_2, \prec_2, \Psi_2)$ , denoted  $m_1 + m_2$ , where  $U_1$  and  $U_2$  are disjoint, is the interpretation  $m = (U, \prec, \Psi)$  such that

- $U = U_1 \cup U_2$ ,
- $\prec = \prec_1 \cup \prec_2 \cup \{(a, b) : a \in U_1 \wedge b \in U_2\}$ ,
- $\Psi(c) = \Psi_1(c) \cup \Psi_2(c)$  for all  $c \in \mathcal{C}$ .

Given an interpretation,  $I$ , and a diagram,  $D_1 \triangleleft D_2$ ,  $I$  models  $D_1 \triangleleft D_2$  if there exist models  $m_1$  and  $m_2$  for  $D_1$  and  $D_2$  respectively and  $I = m_1 + m_2$ . We now define when two diagrams are semantically equivalent.

**Definition 8** Let  $D_1$  and  $D_2$  be spider diagrams of order. If the model set for  $D_1$  is exactly that of  $D_2$  then  $D_1$  and  $D_2$  are **semantically equivalent**, denoted  $D_1 \equiv D_2$ .

Having defined the syntax and semantics of spider diagrams of order, we now define the rules of our sound reasoning system.

### 3 Reasoning Rules

We introduce seven reasoning rules for spider diagrams of order. These rules are subsequently used to produce diagrams in normal form; providing the basis of our reasoning system. The rules are:

1. introduction of a contour label,
2. introduction of a missing zone,
3. splitting spiders,
4. separate rank and bounds,
5. factor lowest spiders,

6. drop spider-foot rank, and

7. rule of replacement.

The rules of replacement, introduction of a contour label, introduction of a missing zone and splitting spiders rule are generalised from [18], whereas the other three rules are completely new. For each rule we present a statement of the rule, a formal definition of the rule and an example of the use of the rule. All of the reasoning rules presented in this section produce semantically equivalent diagrams. Therefore, each of the rules defines its own inverse and if  $D_2$  is the consequence of applying a rule to  $D_1$  then  $D_1$  is the consequence of applying the inverse of the rule to  $D_2$ .

The following rule describes how to introduce a missing contour into a diagram producing a semantically equivalent diagram.

**Rule 1 (Introduction of a Contour Label)** Let  $d$  be a unitary spider diagram of order and let  $d'$  be a unitary spider diagram of order obtained from  $d$  by introduction of a contour label as follows.

- The new contour has a label that is not present in  $d$ .
- The contour introduced in  $d'$  splits each zone  $z$  of  $d$  into two zones  $z_1$  and  $z_2$  and both  $z_1$  and  $z_2$  are shaded where  $z$  is shaded.
- Each unordered foot of a spider in zone  $z$  of  $d$  is replaced in  $d'$  by a pair of unordered spider feet in  $z_1$  and  $z_2$ .
- Each ordered spider foot in zone  $z$  is similarly replaced in  $d'$  by a pair of ranked feet of the same rank in  $z_1$  and  $z_2$ .

Then  $d$  may be replaced by  $d'$  and vice-versa.

**Example 1** Let  $d$  be the diagram in Fig. 2(a). Let  $d'$  be the diagram in Fig. 2(b) where each zone in  $d$  has been split by the introduced contour  $R$  in  $d'$ . Each spider foot of  $d$  has been replaced by a pair of spider feet in  $d'$  such that one foot of the pair is the original foot and the other foot is extended into new zone created by the partition of the original zone containing the original foot.

**Formal Description of Rule 1** Let  $d$  be a unitary spider diagram of order such that  $d \neq \perp$ . Let  $l \in \mathcal{C} - C(d)$  and



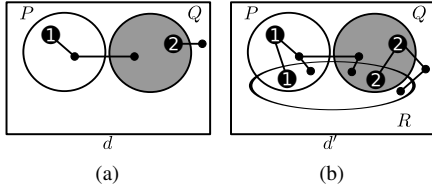


Figure 2: The introduction of a contour label rule

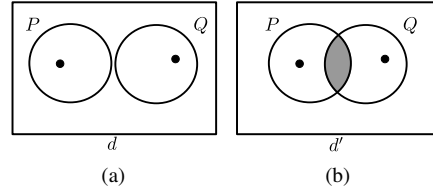


Figure 3: The introduction of a missing zone rule.

let  $d'$  be the diagram where

$$\begin{aligned}
 C(d') &= C(d) \cup \{l\}, \\
 Z(d') &= \{(in \cup \{l\}, out) : (in, out) \in Z(d)\} \cup \\
 &\quad \{(in, out \cup \{l\}) : (in, out) \in Z(d)\}, \\
 ShZ(d') &= \{(in \cup \{l\}, out) : (in, out) \in ShZ(d)\} \cup \\
 &\quad \{(in, out \cup \{l\}) : (in, out) \in ShZ(d)\}, \\
 SI(d') &= \{(n, p') : \exists (n, p) \in SI(d) \wedge p' = \\
 &\quad \{(k, (in \cup \{l\}, out)) : (k, (in, out)) \in p\} \\
 &\quad \cup \{(k, (in, out \cup \{l\})) : (k, (in, out)) \in p\}\}.
 \end{aligned}$$

Then  $d$  may be replaced by  $d'$  and vice-versa.

The add contour rule is sound as the resultant diagram is semantically equivalent to the original, as we now state.

**Theorem 1** Let  $d$  be a unitary spider diagram of order such that  $d \neq \perp$ . Let  $l \in \mathcal{C} - C(d)$ . Let  $d'$  be a spider diagram of order such that  $l$  is introduced to  $d$  resulting in  $d'$  by rule 1, introduction of a contour. Then  $d \equiv_{\models} d'$ .

Given an arbitrary diagram  $D$  and the introduction of a contour label rule we may introduce all contours in  $\mathcal{C}$  producing a diagram containing all contours. The introduction of a missing zone rule, when coupled with the introduction of a contour label rule, allows us to produce diagrams in Venn-form containing all contours.

**Rule 2 (Introduction of a Missing Zone)** Let  $d$  be a unitary spider diagram of order with missing zone  $z$  and let  $d'$  be a copy of  $d$  where  $z$  is added to  $d'$  and  $z$  is shaded. Then  $d$  can be replaced by  $d'$  and vice-versa.

**Example 2** Let  $d$  be the unitary diagram in Fig. 3(a). The zone  $z = (\{P, Q\}, \{\})$  is missing from  $d$ . Let  $d'$  be the diagram in Fig. 3(b). The zone  $z$  has been added as a shaded zone to  $d'$ .

**Formal Description of Rule 2** Let  $d (\neq \perp)$  be a unitary spider diagram of order. Let  $z \in MZ(d)$ . Then  $d'$  is a

unitary spider diagram of order where

$$\begin{aligned}
 C(d') &= C(d), \\
 Z(d') &= Z(d) \cup \{z\}, \\
 ShZ(d') &= ShZ(d) \cup \{z\}, \\
 SI(d') &= SI(d).
 \end{aligned}$$

Then  $d$  can be replaced by  $d'$  and vice-versa.

The following establishes the soundness of the introduction of a missing zone rule.

**Theorem 2** Let  $d (\neq \perp)$  be a unitary spider diagram of order. Let  $z \in MZ(d)$ . Let  $d'$  be the diagram obtained by applying rule 2 introduction of a missing zone to  $d$ . Then  $d \equiv_{\models} d'$ .

The *splitting spiders* rule allows us to represent the disjunctive information held within a unitary diagram as a disjunction of unitary diagrams. By repeated application we generate a disjunction of  $\alpha$ -diagrams.

**Rule 3 (Splitting spiders)** Let  $d (\neq \perp)$  be a unitary spider diagram of order containing a spider  $s$  with foot set  $p$  where  $|p| > 1$ . Let  $d_1$  and  $d_2$  be copies of  $d$  and let  $\{p_1, p_2\}$  be a partition of  $p$ . Then  $s$  is replaced in  $d_1$  with  $s_1$  where the foot set of  $s_1$  is  $p_1$ . Similarly,  $s$  is replaced in  $d_2$  with  $s_2$  where the foot set of  $s_2$  is  $p_2$ . Then  $d$  can be replaced by the diagram  $d_1 \vee d_2$  and vice-versa.

**Example 3** Let  $d_1$  be the diagram in Fig. 4(a). A single application of the splitting spiders rule may result in the diagram  $d_2 \vee d_3$  in Fig. 4(b). A further application of the splitting spiders rule to  $d_3$  produces a disjunction of  $\alpha$ -spider diagrams of order.

In order to formally describe the split spiders rule we require the following definition which allows us to remove of spiders from, and add spiders to, a unitary diagram.

**Definition 9** Let  $d$  be a unitary spider diagram of order. Let  $p$  be a foot set such that  $\{z : (k, z) \in p\} \subseteq Z(d)$ . Let  $d'$  be a unitary spider diagram of order that contains the same set of contours, set of zones and set of shaded zones as  $d$ .

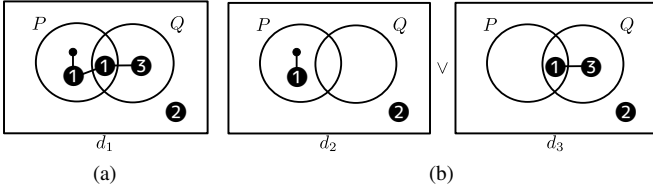


Figure 4: The *splitting spiders* rule.

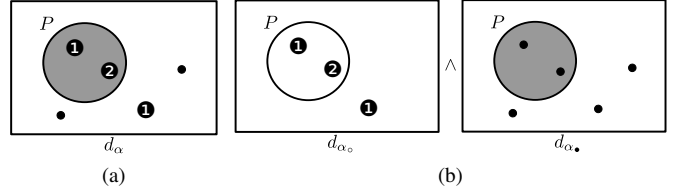


Figure 5: The *separate rank and bounds* rule.

We may **remove a spider** with foot set  $p$  from  $d$ , denoted  $S(d) \ominus p$  to give  $d'$  such that,  $d'$  is identical to  $d$  except that:

$$S(d') = S(d) - \{(n, p) : (n, p) \in SI(d)\}.$$

Alternatively, we may **add a spider** with foot set  $p$  to  $d$ , denoted  $S(d) \oplus p$  to give  $d'$  such that,  $d'$  is identical to  $d$  except that:

$$S(d') = S(d) \cup \{(n+1, p) : (n, p) \in SI(d)\} \cup \{(1, p)\}.$$

**Formal Description of Rule 3** Let  $d$  be a spider diagram of order containing a spider  $s = (n, p)$  with  $|p| > 1$  and let  $\{p_1, p_2\}$  be a partition of  $p$ . Let  $d_1$  and  $d_2$  be unitary diagrams such that:

- $d_1 = (d \ominus p) \oplus p_1$ , and
- $d_2 = (d \ominus p) \oplus p_2$ .

Then  $d$  can be replaced by the diagram  $d_1 \vee d_2$  and vice versa.

**Theorem 3** Let  $d$  be a unitary spider diagram of order and let  $d_1 \vee d_2$  be the result of the application of rule 3 splitting spiders to  $d$ . Then  $d \equiv d_1 \vee d_2$ .

When given an arbitrary diagram  $D$ , we may use the introduction of a contour label, introduction of a missing zone and *splitting spiders* rules, to produce a diagram  $D_\alpha$  where each unitary component is in Venn-form, contains all contours in  $\mathcal{C}$  and is an  $\alpha$ -diagram. We now introduce a series of three rules which, when given  $D_\alpha$  produce a diagram that contains no ranked feet. The first of our three rules isolates the order information from the bounds information provided by unranked spiders and shading.

**Rule 4 (Separate rank and bounds)** Let  $d$  be a unitary  $\alpha$ -spider diagram of order. The diagram  $d$  can be decomposed into the conjunction of  $d_1$  and  $d_2$  where  $d_1$  contains the ordered spider feet from  $d$  and no shading. Furthermore,  $d_2$  contains an unranked spider foot for each spider foot in  $d$  and each shaded zone in  $d$  is also shaded in  $d_2$ . Then  $d$  can be replaced by diagram  $d_1 \wedge d_2$  and vice-versa.

**Example 4** Let  $d$  be the diagram in Fig. 5(a). A single application of rule 4 separate rank and bounds produces  $d_{\alpha_o} \wedge d_{\alpha_\bullet}$  in Fig. 5(b). The ordered spider feet are separated from the bounds information as the diagram  $d_1$  contains only the order information, provided by ordered spider feet in  $d$ . The diagram  $d_2$  contains the bounds information provided by both the shading and spiders in  $d$ . For each of the spiders in  $d$  there exists an unranked spider in  $d_2$  with the same habitat.

**Formal Description of Rule 4** Let  $d (\neq \perp)$  be a unitary  $\alpha$ -diagram of order in Venn-form containing all contours in  $\mathcal{C}$ . Let  $d_1$  and  $d_2$  be diagrams such that

$$\begin{aligned} C(d_1) &= C(d_2) = C(d), \\ Z(d_1) &= Z(d_2) = Z(d), \\ ShZ(d_1) &= \emptyset, \\ ShZ(d_2) &= ShZ(d), \end{aligned}$$

and

$$\begin{aligned} SI(d_1) &= \{(n, \{(k, z)\}) : (n, \{(k, z)\}) \in SI(d) \wedge k \neq \bullet\}, \\ SI(d_2) &= \{(n, \{(\bullet, z)\}) : n > 0 \wedge n = \sum_{(m, \{(k, z)\}) \in SI(d)} m\}. \end{aligned}$$

Then  $d$  can be replaced by  $d_1 \wedge d_2$  and vice-versa.

Our proof of the soundness of rule 4 proceeds by first presenting a series of lemmas. Each lemma corresponds to a step demonstrated in Fig. 6. The diagram  $d$  is semantically equivalent to the diagram  $d \wedge d_2$  by lemma 1 below. Lemma 2 will show that  $d \wedge d_2$  is equivalent to  $d_3 \wedge d_2$  i.e. we may drop shading from  $d$  without changing the meaning of the diagram  $d \wedge d_2$ . Finally, lemma 3 will show that

$$d_3 \wedge d_2 \equiv d_4 \wedge d_2 \equiv d_5 \wedge d_2$$

i.e. we may remove all unranked spiders from  $d_3$ , one at a time, without changing the meaning of the diagram. We now show that the first step in this process holds.

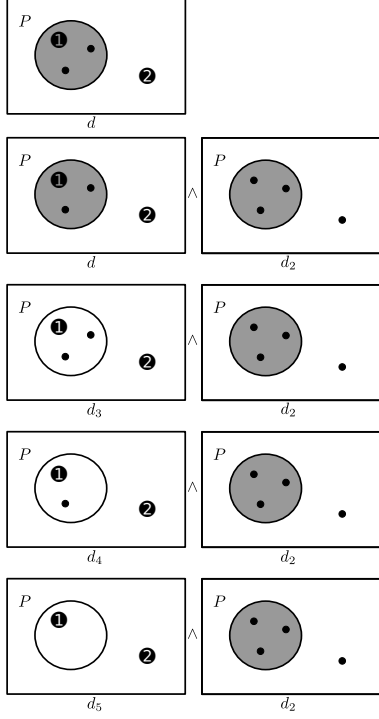


Figure 6: The separating the rank and bounds rule.

**Lemma 1** Let  $d$  be a unitary  $\alpha$ -diagram of order in Venn-form containing all contours in  $\mathcal{C}$ . Let  $d_2$  be a diagram where

$$\begin{aligned} C(d_2) &= C(d) \\ Z(d_2) &= C(d) \\ ShZ(d_2) &= ShZ(d) \end{aligned}$$

and

$$SI(d_2) = \{(n, \{(\bullet, z)\}) : n > 0 \wedge n = \sum_{(m, \{(k, z)\}) \in SI(d)} m\}.$$

Then  $d \equiv d \wedge d_2$ .

Returning to Fig. 6, lemma 1 shows that  $d$  is semantically equivalent to  $d \wedge d_2$ . We now show that  $d \wedge d_2$  is semantically equivalent to  $d_3 \wedge d_2$ . Here,  $d_3$  is obtained from  $d$  by removing the shading.

**Lemma 2** Let  $d \wedge d_2$  where  $d \neq \perp$  and  $d_2 \neq \perp$  be a spider diagram of order where  $d$  is a unitary  $\alpha$ -diagram of order in Venn-form containing all contours in  $\mathcal{C}$  and where

$$\begin{aligned} C(d_2) &= C(d) \\ Z(d_2) &= C(d) \\ ShZ(d_2) &= ShZ(d) \end{aligned}$$

and

$$SI(d_2) = \{(n, \{(\bullet, z)\}) : n > 0 \wedge n = \sum_{(m, \{(k, z)\}) \in SI(d)} m\}.$$

Let  $d_3$  be a copy of  $d$  where  $ShZ(d_3) = \{\}$ . Then  $d \wedge d_2 \equiv d_3 \wedge d_2$ .

We now show that where rank information has been separated from bounds information we can remove unranked spiders. This is illustrated in Fig. 6, where  $d_3 \wedge d_2$  becomes  $d_4 \wedge d_2$ , by removing an unranked spider.

**Lemma 3** Let  $d_3 \wedge d_2$  be a spider diagram of order where

- $d_3$  is a unitary  $\alpha$ -spider diagram of order in Venn-form containing all contours from  $\mathcal{C}$  and  $ShZ(d_3) = \{\}$ ,
- $d_2$  is a unitary  $\alpha$ -spider diagram of order in Venn-form containing all contours from  $\mathcal{C}$  containing only unranked spider feet, and
- there exists a habitat preserving injective function  $\pi : S(d_3) \rightarrow S(d_2)$ .

Let  $d_4$  be a copy of  $d_3$  where one of the unranked spiders  $s = (n, \{(\bullet, z)\})$  in  $d_4$  is removed i.e.  $S(d_4) = S(d_3) \ominus \{(\bullet, z)\}$ . Then  $d_3 \wedge d_2 \equiv d_4 \wedge d_2$ .

Having demonstrated that we can remove a single unranked spider, we can repeatedly remove such spiders from diagrams like  $d_3$  in Fig. 6 until no unranked spiders remain. We use this observation in the proof of the next theorem.

**Theorem 4** Let  $d$  be a unitary  $\alpha$ -diagram of order where  $d_1 \wedge d_2$  is the result of applying rule 4 separate rank and bounds rule to  $d$ . Then  $d \equiv d_1 \wedge d_2$ .

Given a diagram  $d$  that contains only rank information (such a diagram is generated by an application of *separate rank and bounds*) the factor lowest spiders rule allows us to factor the different ranks out into a product of diagrams, where each unitary component of the product contains spiders of the same rank. We will subsequently show that ranked spider feet may be substituted by unranked spider feet given a unitary diagram containing only spiders of the same rank.

**Rule 5 (Factor lowest spiders)** Let  $d$  be a unitary  $\alpha$ -diagram containing only spiders whose feet are ranked and containing no shaded zones. Then  $d$  may be replaced by  $d_k \triangleleft d'$  where  $d_k$  contains those spider feet of lowest rank in  $d$  and  $d'$  contains all other spider feet.

**Example 5** Let  $d_{\alpha_0}$  be the diagram in Fig. 7(a) and  $d_{\alpha_0^1} \triangleleft d_{\alpha_0^2}$  be the diagram in Fig. 7(b). We factor  $d_{\alpha_0}$  into  $d_{\alpha_0^1}$  and  $d_{\alpha_0^2}$  where  $d_{\alpha_0^1}$  contains all the lowest ranked spider feet from  $d_{\alpha_0}$  i.e. those spider feet labelled 1. The diagram  $d_{\alpha_0^1} \triangleleft d_{\alpha_0^2} \equiv d_{\alpha_0}$ .

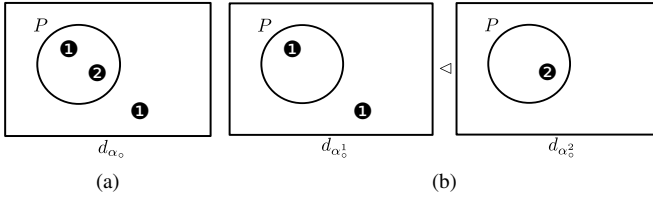


Figure 7: The factor lowest spiders rule

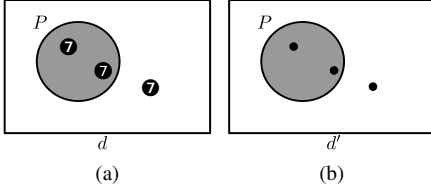


Figure 8: The drop spider foot rank rule.

**Formal Description of Rule 5** Let  $d$  be a unitary  $\alpha$ -diagram containing only spiders whose feet are ranked and containing no shaded zones. Considering the ranked spider feet in  $d$ , let  $k$  be the lowest rank of these feet. Let  $d_k$  and  $d'$  be diagrams such that

$$\begin{aligned} C(d_k) &= C(d') = C(d) \\ Z(d_k) &= Z(d') = Z(d) \\ ShZ(d_k) &= ShZ(d') = ShZ(d) = \{\} \end{aligned}$$

and

$$\begin{aligned} SI(d_k) &= \{(n, (k, z)) : (n, (k, z)) \in SI(d)\} \\ SI(d') &= SI(d) - SI(d_k). \end{aligned}$$

Then  $d$  may be replaced by  $d_k \triangleleft d'$  and vice versa.

**Theorem 5** Let  $d (\neq \perp)$  be a unitary  $\alpha$ -diagram containing only spiders whose feet are ranked and containing no shaded zones. Let  $d \equiv_{\models} d_k \triangleleft d'$  be the result of applying rule 5 factor lowest spiders to  $d$ . Then  $d \equiv_{\models} d_k \triangleleft d'$ .

**Rule 6 (drop spider-foot rank)** Let  $d$  be a unitary  $\alpha$ -diagram such that each foot of each spider in  $d$  is of rank  $k \in \mathbb{Z}^+$ . Then  $d$  may be replaced by a diagram  $d'$  where  $d'$  is a copy of  $d$  and each foot of each spider foot in  $d'$  is unranked.

**Example 6** Let  $d$  be the diagram in Fig. 8(a) is a spider diagram of order in which all spiders contain the same rank spider feet. Then  $d$  may be replaced by  $d'$  in Fig. 8(b).

**Formal Description of Rule 6** Let  $d$  be a unitary  $\alpha$ -diagram and  $k \in \mathbb{Z}^+$  where

$$(n, \{(j, z)\}) \in SI(d) \Rightarrow j = k.$$

The diagram  $d$  may be replaced by  $d'$  where

$$\begin{aligned} C(d') &= C(d), \\ Z(d') &= Z(d), \\ ShZ(d') &= ShZ(d), \end{aligned}$$

and

$$SI(d') = \{(n, \{(\bullet, z)\}) : (n, \{(k, z)\}) \in SI(d)\}.$$

**Theorem 6** Let  $d (\neq \perp)$  be a unitary  $\alpha$ -diagram where all spiders contain only feet of rank  $k$ . Let  $d'$  be the diagram produced by application of rule 6 to  $d$ . Then  $d \equiv_{\models} d'$ .

Our final rule, the rule of replacement, allows us to replace any spider diagram of order which is a sub-diagram in a compound expression with a semantically equivalent diagram. The purpose of this rule is to allow a sub-diagram in a compound expression to be replaced by the result of application of a reasoning rules to that sub-diagram. We first define a sub-diagram. We observe that the syntax of a spider diagram of order is defined by the following grammar in Backus-Naur form:

$$\begin{aligned} \text{diagram} &::= \langle \text{unitary\_diagram} \rangle | \text{conjunction} \\ &\quad | \text{disjunction} | \text{negation} | \text{product}; \\ \text{conjunction} &::= \text{diagram} \wedge \text{diagram}; \\ \text{disjunction} &::= \text{diagram} \vee \text{diagram}; \\ \text{negation} &::= \neg \text{diagram}; \\ \text{product} &::= \text{diagram} \triangleleft \text{diagram}; \end{aligned}$$

Given any spider diagram of order we may now construct its abstract syntax tree. Each tree contains unitary spider diagrams of order at leaf nodes and compound operators at non-leaf nodes.

The set of all abstract syntax trees is  $T$  and the set of all spider diagrams of order is  $\Delta$ .

**Theorem 7** Let  $D$  be a spider diagram of order. There exists a unique abstract syntax tree  $t$  and bijective function  $\delta : \Delta \rightarrow T$  such that  $\delta(D) = t$ .

Let  $D$  be a well-formed spider diagram of order with abstract syntax tree  $t$  where  $t$  contains a non-leaf node  $r$ . The tree  $t_r$  with root node  $r$  is a **sub-tree** of  $t$ . Furthermore,  $\delta^{-1}(t_r)$  is a **sub-diagram** of  $D$ .

Not only do we need to define what a sub-tree is, but we also need to know when two diagrams are syntactically equivalent. The following two definitions define syntactic equivalence

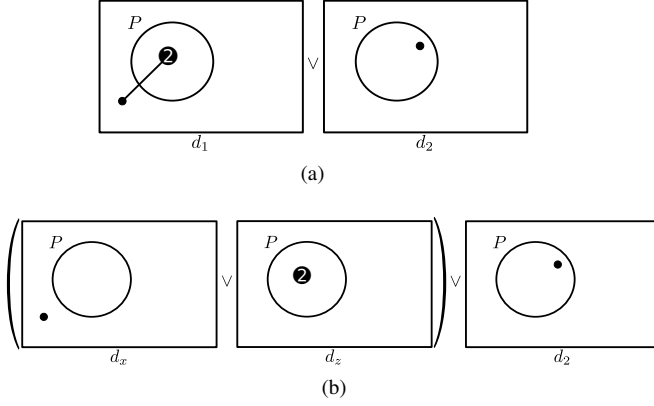


Figure 9: The rule of replacement.

**Definition 10 (Adapted from [12] and [15])** Let  $D_1$  and  $D_2$  be spider diagrams of order. Then  $D_1 \Vdash D_2$  if and only if  $D_1$  can be transformed into  $D_2$  by applying one of the reasoning rules given in this section. We say that  $D_2$  is **obtainable** from  $D_1$ , denoted  $D_1 \vdash D_2$ , if and only if there is a sequence of diagrams  $\langle D^1, D^2, \dots, D^m \rangle$  such that  $D^1 = D_1, D^m = D_2$  and  $D^{k-1} \vdash D^k$  for each  $1 \leq k \leq m$ .

**Definition 11 (Adapted from [12] and [15])** Let  $D_1$  and  $D_2$  be spider diagrams of order. If  $D_1 \vdash D_2$  and  $D_2 \vdash D_1$  then  $D_1$  and  $D_2$  are **syntactically equivalent** denoted,  $D_1 \equiv_{\vdash} D_2$ .

We may now define the rule of replacement.

**Rule 7 (Rule of Replacement)** Let  $D$  and  $D'$  be spider diagrams of order. Let  $D_r$  be a sub-diagram of  $D$ , where  $D_r$  is syntactically equivalent to  $D'$ . Then an instance of  $D_r$  in  $D$  may be replaced by  $D'$ .

**Example 7** In the diagram  $d_1 \vee d_2$  in Fig. 9(a) we apply a rule to  $d_1$  such that  $d_1 \vdash d_x \vee d_y$  (specifically, the split spiders rule). Then we may replace  $d_1$  by  $d_x \vee d_y$  as seen in Fig. 9(b).

**Formal Description of Rule 7** Let  $D_1$  and  $D'$  be spider diagrams of order where  $D_r$  is a sub-diagram of  $D_1$ . If  $D_r \equiv_{\vdash} D'$  then any occurrence of the subtree  $\delta(D_r)$  in  $\delta(D_1)$  may be replaced by  $\delta(D')$  to produce  $D_2$ . Then  $D_1$  can be replaced by  $D_2$  and vice-versa.

The soundness of the rule of replacement is given in the following theorem:

**Theorem 8** Let  $D_1$  and  $D_2$  be spider diagrams of order where  $D_2$  is obtained from  $D_1$  by application of rule 7 rule of replacement. Then  $D_1$  is semantically equivalent to  $D_2$ .

Each of our seven reasoning rules is sound. However, our reasoning system is incomplete. Previous approaches to showing completeness of spider diagram based reasoning systems do not readily generalise to spider diagrams of order. As a first step in developing a complete reasoning system we produce a normal form for spider diagrams of order. In the next section we provide an algorithm that, given an arbitrary spider diagram of order, produces a spider diagram of order in our normal form.

## 4 An Algorithm to Produce Diagrams in Normal Form

In this section we define a normal form for spider diagrams of order. Our normal form allows the diagram  $\perp$ . Furthermore, compound diagrams are formed from unitary  $\alpha$ -spider diagrams in Venn-form containing all contours in  $\mathcal{C}$  and no ranked spiders. Compound expressions in normal form allow  $\wedge, \vee$  and  $\triangleleft$  as connectives and  $\neg$  as the unary operator. From the 7 reasoning rules, presented in the previous section, we show that any spider diagram of order is semantically equivalent to a diagram in our normal form. Specifically, we define a sequence of applications of reasoning rules for producing a diagram in normal form given an arbitrary spider diagram of order.

**Definition 12** Let  $D$  be a spider diagram of order. It is the case that  $D$  is in **normal form** if the following conditions hold:

- No unitary component of  $D$  contains ranked spider feet.
- Each unitary component of  $D$  is an  $\alpha$ -diagram and contains all contours in  $\mathcal{C}$ , or is  $\perp$ .
- There are no zones missing from any unitary component ( $\neq \perp$ ) of  $D$ .
- The binary connectives  $\wedge, \vee$  and  $\triangleleft$  and the unitary connective  $\neg$  are the allowed connectives in  $D$ .

Given a spider diagram of order as input, the algorithm produces a spider diagram of order in normal form as output. The algorithm is outlined as follows, where applications of rule 7, the rule of replacement, are implicitly assumed:

- Let  $D$  be the input diagram.
- Apply rule 1 to each unitary diagram in  $D$ , producing  $D_{\mathcal{C}}$ , until all contours in  $\mathcal{C}$  are present in the result  $D_{\mathcal{C}}$ .
- Apply rule 2 to each unitary diagram in  $D_{\mathcal{C}}$ , producing  $D_{\mathcal{Z}}$ , until there are no missing zones in the result  $D_{\mathcal{Z}}$ .

- Apply rule 3 to each unitary diagram in  $D_Z$  until there are no spiders with multiple feet in the result  $D_\alpha$ .
- Apply rule 4 to each unitary diagram in  $D_\alpha$ , producing  $D_{\alpha_0}$ .
- Apply rule 5 to each unitary diagram in  $D_{\alpha_0}$  until every unitary diagram either contains no spider feet or spider feet of all the same rank.
- Apply rule 6 to each unitary diagram in  $D_{\alpha'_0}$ , producing  $D_\bullet$ , the final result.

We now present an example of a unitary spider diagram, its corresponding normal form diagram and an illustration of the algorithm to generate the required normal form.

**Example 8** Let  $d_1$  be the unitary spider diagram of order in Fig. 10(a) and let  $\mathcal{C} = \{P, Q, R\}$ . The diagram  $d_1$  contains the contours  $P$  and  $Q$  and the zones  $(\{\}, \{P, Q\})$ ,  $(\{P\}, \{Q\})$  and  $(\{Q\}, \{P\})$ . It is not in Venn form as the zone  $(\{P, Q\}, \{\})$  is missing. Furthermore, the contour  $R$  is not present in  $d_1$  and the diagram contains a spider with more than one foot. Our strategy is to first add all the missing contours to  $d_1$ . The addition of  $R$  to  $d$  gives the diagram in Fig. 10(b). We then add all the zones that are missing from 10(b) forming 10(c). The addition of the zones  $(\{P, Q\}, \{R\})$  and  $(\{P, Q, R\}, \{\})$  to the diagram can be seen in Fig. 10(c). The diagram in Fig. 10(c) is in Venn form and contains no missing zones. It is semantically equivalent to  $d_1$ . From this diagram in Venn form containing all contours in  $\mathcal{C}$  we repeatedly employ a split spiders reasoning rule to produce a disjunction of  $\alpha$ -diagrams. The diagram in Fig. 10(d) is a disjunction of  $\alpha$ -diagrams where each unitary component contains all contours in  $\mathcal{C}$ . It is semantically equivalent to  $d_1$ .

Let  $d_7$  (an arbitrary choice) be the unitary component of the diagram in Fig. 10(d) as annotated in the figure. We now show that  $d_7$  and, by extension, any unitary  $\alpha$ -diagram may be transformed, by application of reasoning rules, into a diagram in our normal form. We first separate order and shading resulting in diagram  $d_{12} \wedge d_{13}$  in Fig. 11(a). Thereafter, we factor lowest spiders from  $d_{12}$  and replace  $d_{12}$  with  $d'_{12} \wedge d''_{12}$  in Fig. 11(b). Finally, the precondition for the drop spider foot order rule is satisfied we drop the ordered foot from  $d'_{12}$  resulting in  $d_{14}$  in Fig. 11(a). Applying the drop spider foot order rule to  $d'_{12}$  and replacing the result into  $(d_{14} \triangleleft d''_{12}) \wedge d_{13}$  yields the diagram  $(d_{14} \triangleleft d_{15}) \wedge d_{13}$  in normal form in Fig. 11(d).

## 5 Conclusion

We have presented a reasoning system and normal form for spider diagrams of order. Spider diagrams of order are

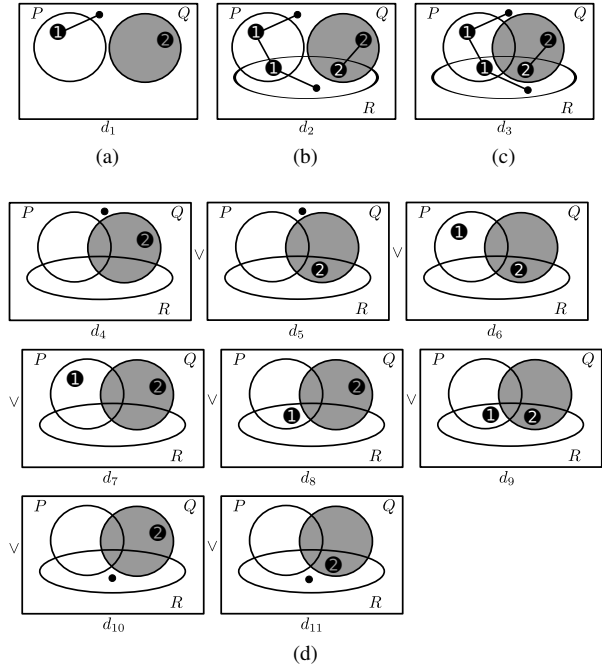


Figure 10: Applying reasoning rules resulting in a disjunction of  $\alpha$ -diagrams.

an interesting recent advance in Euler-diagram based visual logics as they incorporate an order relation into their semantics. Furthermore, it is known that spider diagrams of order are as expressive as star-free regular languages and that star-free regular languages are as expressive as linear temporal logic. In the future, we wish to use the normal form, developed in this paper, to directly compare spider diagrams of order and linear temporal logic.

We view our algorithm for obtaining normal form to be the first step towards providing a completeness result for spider diagrams of order. This is because the completeness proofs for existing spider diagram logics, such as [9], as well as their extension called constraint diagrams [10], rely on obtaining diagrams with the property that all spiders have single feet [2]. This property is delivered by our normal form. Once in this normal form, the completeness proofs use other rules to establish syntactic entailment. As it stands, the spider diagram of order logic does not have sufficient rules to establish completeness. Finding a complete set of rules is an interesting prospect for future work because it will provide insight into how to gain completeness when an order operator is present, contributing to our understanding of diagrammatic logics in general.



## References

- [1] P. Bottoni and A. Fish. Policy enforcement and verification with timed modeling spider diagrams. In *IEEE Symposium on Visual Languages and Human-Centric Computing*, pages 27–34, 2013.
- [2] J. Burton, G. Stapleton, and J. Howse. Completeness proof strategies for euler diagram logics. In *International Conference on the Theory and Application of Diagrams*, 2012.
- [3] P. Chapman, G. Stapleton, and A. Delaney. On the expressiveness of second-order spider diagrams. *Journal of Visual Languages & Computing*, 24(5):327 – 349, 2013.
- [4] L. Choudhury and M. K. Chakraborty. On extending Venn diagrams by augmenting names of individuals. In *International Conference on the Theory and Application of Diagrams*, volume 2980 of *LNAI*, pages 142–146. Springer-Verlag, March 2004.
- [5] J. Cohen, D. Perrin, and J.-E. Pin. On the expressive power of temporal logic. *Computer System Science*, 46:271–294, 1993.
- [6] H. Dunn-Davies and R. Cunningham. Propositional state-charts for agent interaction protocols. In *Proceedings of Euler Diagrams 2004, Brighton, UK*, volume 134 of *ENTCS*, pages 55–75, 2005.
- [7] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Springer, 2 edition, 1991.
- [8] A. Fish and J. Taylor. Equivalences in euler-based diagram systems through normal forms. *London Mathematical Society*, 2014. *to appear*.
- [9] J. Howse, G. Stapleton, and J. Taylor. Spider diagrams. *LMS Journal of Computation and Mathematics*, 8:145–194, 2005.
- [10] S. Kent. Constraint diagrams: Visualizing invariants in object oriented modelling. In *International Conference on Object Oriented Programming, Systems, Languages and Applications*, pages 327–341. ACM Press, October 1997.
- [11] M. Leucker and C. Snchez. Regular linear temporal logic. In C. Jones, Z. Liu, and J. Woodcock, editors, *Theoretical Aspects of Computing ICTAC 2007*, volume 4711 of *Lecture Notes in Computer Science*, pages 291–305. Springer, 2007.
- [12] F. Molina. *Reasoning with extended Venn-Peirce diagrammatic systems*. PhD thesis, University of Brighton, 2001.
- [13] B. Nagy and S. Vályi. Visual reasoning by generalized interval-values and interval temporal logic. In P. T. Cox, A. Fish, and J. Howse, editors, *VLL*, volume 274 of *CEUR Workshop Proceedings*, pages 13–26. CEUR-WS.org, 2007.
- [14] S.-J. Shin. *The Logical Status of Diagrams*. Cambridge University Press, 1994.
- [15] G. Stapleton. *Reasoning with Constraint Diagrams*. PhD thesis, University of Brighton, August 2004.
- [16] G. Stapleton and A. Delaney. Evaluating and generalizing constraint diagrams. *Journal of Visual Languages and Computation*, 19:499–521, August 2008.
- [17] G. Stapleton, J. Howse, P. Chapman, A. Delaney, J. Burton, and I. Oliver. Formalizing concept diagrams. In *19th International Conference on Distributed Multimedia Systems*, pages 182 – 187. Knowledge Systems Institute, 2013.
- [18] G. Stapleton, J. Howse, and J. Taylor. A decidable constraint diagram reasoning system. *Journal of Logic and Computation*, 15(6):975–1008, December 2005.

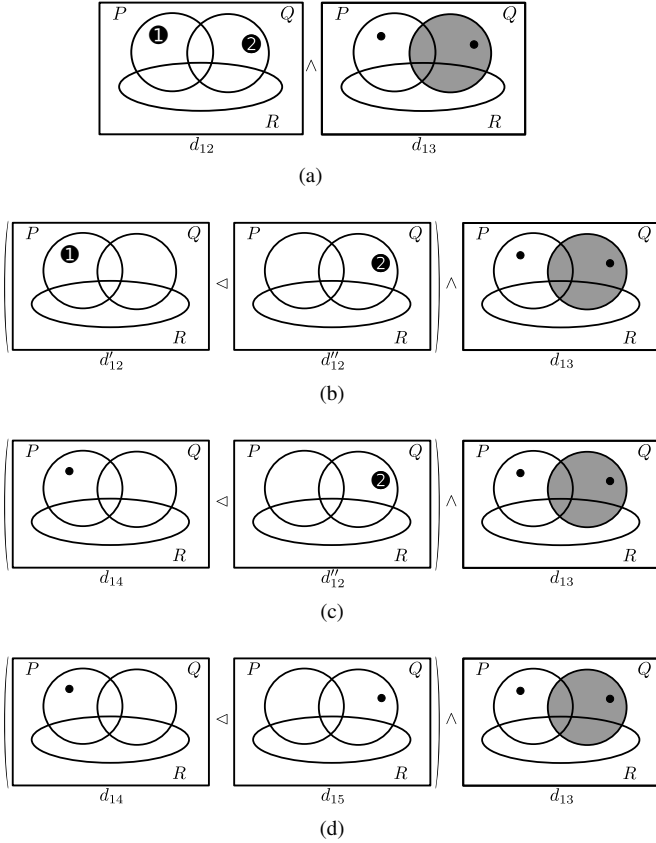


Figure 11: Applying reasoning rules resulting in a diagram in normal form.

# PaL Diagrams: A Linear Diagram-Based Visual Language

Peter Chapman<sup>1</sup>, Gem Stapleton<sup>1</sup>, Peter Rodgers<sup>2</sup>

<sup>1</sup>University of Brighton, <sup>2</sup>University of Kent

E-mail: {p.b.chapman, g.e.stapleton}@brighton.ac.uk, p.j.rodgers@kent.ac.uk

## Abstract

*Linear diagrams have recently been shown to be more effective than Euler diagrams when used for set-based reasoning. However, unlike the growing corpus of knowledge about formal aspects of Euler and Venn diagrams, there has been no formalisation of linear diagrams. To fill this knowledge gap, we present and formalise Point and Line (PaL) diagrams, an extension of simple linear diagrams containing points, thus providing a formal foundation for an effective visual language. We prove that PaL diagrams are exactly as expressive as monadic first-order logic with equality, gaining, as a corollary, an equivalence with the Euler diagram extension called spider diagrams. The method of proof provides translations between PaL diagrams and sentences of monadic first-order logic.*

## 1. Introduction

Linear diagrams have a long history, with the first recorded use of them owing to Leibniz in 1686 [1, 4]. Much like Venn and Euler diagrams, they express information about sets in a visual way. Whilst Venn and Euler diagrams have been put on a formal footing (see [10]), linear diagrams have largely been overlooked, which we begin to address in this paper. In a linear diagram, parallel labelled line segments represent sets. The vertical overlap of lines represents the intersection of the corresponding sets. For example, consider the diagrams in Fig. 1. The three diagrams shown express the same information, namely that  $A \cap B = \emptyset$ , and  $C \subseteq A$ :  $d_1$  is a Venn diagram, using shading to represent the emptiness of certain set intersections;  $d_2$  is an Euler diagram, which uses disjointness of curves to represent emptiness of sets; and  $d_3$  is a linear diagram, where the absence of any vertical overlap between the lines labelled  $A$  and  $B$  represents the emptiness of the corresponding set intersection.

As notations built upon Euler diagrams (hereafter Euler-based diagrams) have been widely used and formalised, the expressiveness of these notations has been well stud-

ied. Venn-II and Euler diagrams exactly as expressive as monadic first-order logic (MFOL) [12, 14]. Although it has not been formally established, the expressiveness of the Euler/Venn system is thought to be somewhere between MFOL and monadic first-order logic with equality (MFOL[=]) [7]. Spider diagrams extend Euler diagrams with points, and are known to be exactly as expressive as MFOL[=] [16]. Of this family of logics, generalised constraint diagrams are at least as expressive as dyadic first-order logic making them the most expressive [13].

Recent research provided empirical evidence that linear diagrams can be more effective for visualisation than Euler-based diagrams. In the restricted setting of representing syllogisms, [11] showed that linear diagrams performed as well as Euler diagrams. In [3], where the context was general set-based reasoning, participants using linear diagrams outperformed those using Euler diagrams in terms of both task completion times and error rates. In order to exploit this interesting result, we propose an extension to linear diagrams, called PaL diagrams, by adding points allowing the representation of both sets and elements. We have two goals: (a) to provide a formal foundation for PaL diagrams in order that we may (b) determine contexts where these new diagrams maintain their advantage over Euler-based notations. It is the first of these goals which is one focus of this paper.

Adequate formal foundations are important for a number of reasons. Firstly, to compare the efficacy of two notations, it is key that the two notations are capable of expressing the same information. Any meaningful comparison between notations can only be performed on information expressible in all. Without formalisation, determining the expressiveness of a notation is not possible. In this paper, we show that PaL diagrams are exactly as expressive as MFOL[=], giving us the corollary that PaL diagrams are equivalent in expressive power to spider diagrams, and more expressive than Euler and Venn-II diagrams. Secondly, while static diagrams are useful, the ability to manipulate and reason with diagrams in a coherent manner is also desirable. The only way in which such reasoning rules can be determined and shown to be sound is through formalisation. The develop-

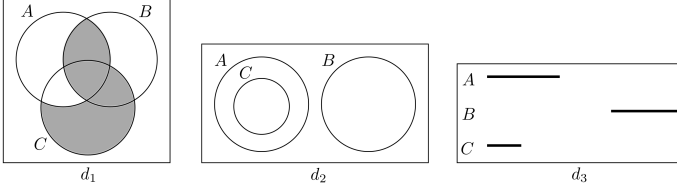


Figure 1. Venn, Euler and linear diagrams

ment of reasoning rules is outside the scope of this paper, but the work contained herein will allow such rules to be defined, and reasoned about, in a rigorous manner.

A number of notations have been derived from linear diagrams. For instance, the parallel bargrams of [18] and the double decker plots of [8] are both closely related to Leibniz's original version of linear diagrams, though neither are formal objects. To our knowledge, only one attempt has been made to formalise diagrams similar in flavour to linear diagrams, called *line diagrams* [6]. These diagrams contain non-parallel lines, and the intersection of these lines asserts the existence of an element in the corresponding sets. However, [9] showed the construction rules for these line diagrams were unsound. There is thus a gap for a formalization of linear diagrams, which is a key contribution of this paper.

The rest of the paper is organised as follows. In section 2 we give formal definitions of the syntax and semantics of PaL diagrams. Section 3 gives an overview of some MFOL[=] concepts necessary for establishing expressive equivalence of PaL diagrams and MFOL[=]. The sections 4 and 5 contain demonstrations that every PaL diagram is equivalent to some sentence in MFOL[=], and that every sentence in MFOL[=] can be equivalently expressed as a PaL diagram, respectively. We conclude and point to future directions in section 6.

## 2. PaL diagrams: Syntax and Semantics

A PaL diagram consists of a set of parallel horizontal line segments (the actual orientation is somewhat irrelevant, all that is important is that the lines are parallel) with a collection of points arranged underneath the lines, as in Fig. 2. How the points and lines are arranged determines the meaning of the diagram. We proceed to present an abstract syntax for PaL diagrams.

In what follows, we take  $\mathcal{L}$  to be a countably infinite set of letters, whose elements are called **line labels**, and  $\mathcal{P}$  to be a countably infinite set of letters, whose elements are called **point labels**, disjoint from  $\mathcal{L}$ . In examples, we use capital roman letters  $A, B, C, \dots$  as elements of  $\mathcal{L}$  and lower case roman letters  $a, b, c, \dots$  as elements of  $\mathcal{P}$ . When making general statements, we use  $L_i$  and  $p_i$  to denote line and

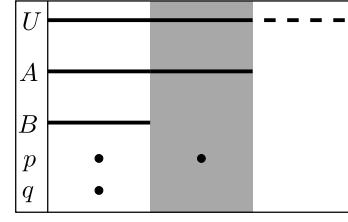


Figure 2. A PaL diagram

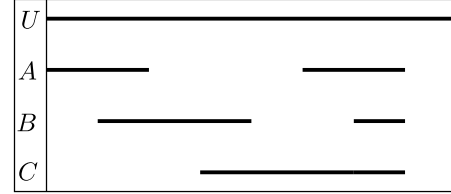


Figure 3. Venn-3

point labels respectively. We reserve the letter  $U$  for a particular purpose, thus it is the case that  $U \notin \mathcal{L} \cup \mathcal{P}$ .

The main interaction between lines is the *overlap* which is where horizontal lines share the same vertical space. Formally:

**Definition 1** An *overlap* is a word  $L_1 L_2 \dots L_n$  where each  $L_i$  is a line label from  $\mathcal{L}$ , possibly adorned with a bar,  $\bar{L}_i$ . Given an overlap,  $O$ , we say that  $L_i$  is **barred** if  $\bar{L}_i$  appears in  $O$ , otherwise  $L_i$  is **unbarred**. The set of all overlaps is denoted  $\mathcal{O}$ .

For example, suppose that  $A, B$  and  $C$  are line labels in  $\mathcal{L}$ . Then the following are overlaps:  $ABC$ ,  $\bar{A}BC$  and  $A\bar{A}B\bar{C}C$ . Note that last overlap is peculiar in that letters appear duplicated, barred and unbarred. We allow such overlaps to make reasoning about contradictions straightforward and intuitive, although that will be future work. In Fig. 2, there are three overlaps, reading left to right:  $AB$  (where the lines labelled  $A$  and  $B$  overlap),  $\bar{A}\bar{B}$  (where the line  $A$  does not overlap with  $B$ ), and  $\bar{A}\bar{B}$  (where neither  $A$  nor  $B$  appear). Notice that the top line, labelled  $U$ , does not appear in the overlaps. The line label  $U$  is special: it represents the universal set and its presence indicates the extreme left and right coordinates of the line segments in the diagram. This limiting behaviour is important when there are overlaps whose letters are all barred; in Fig. 2, without the line for  $U$ , the PaL diagram could be taken to assert that all elements had to be in  $A$ , since the overlap  $\bar{A}\bar{B}$  would be not be visible in the diagram.

Lines may consist of several segments. For example, in Fig. 3 we have a representation of Venn-3. This diagram would be impossible to draw without splitting at least one of the lines into segments. Here  $A$  and  $B$  consist of multiple

segments, whereas  $C$  consists of a single segment. The  $U$  line can only consist of a single segment.

Overlaps can either be **solid** or **dashed**, and this is represented by  $\text{—}$  or  $\text{---}$  respectively. The interpretation of solid overlaps is that the set intersection represented is non-empty, whereas a dashed overlap *could* represent an empty set.

Points, which are visually drawn underneath overlaps as in Fig. 2, are formally defined in a similar way to overlaps, and to each overlap we associate the set of points occupying the same vertical space, called a *clan* of points. Formally:

**Definition 2** A *clan* is a word  $p_1p_2 \dots p_n$  where each  $p_i$  is a letter drawn from  $\mathcal{P}$ , possibly adorned with a bar,  $\bar{p}_i$ . Given a clan  $cl$ , we say that  $p_i$  is **barred** if  $\bar{p}_i$  appears in  $cl$ , otherwise  $p_i$  is **unbarred**.

For example, suppose that  $a, b, c$  are letters from  $\mathcal{P}$ . Then the following are clans:  $a\bar{b}$ ,  $c\bar{c}$ , and  $a\bar{b}c$ . Informally, points which are unbarred are said to lie under overlaps. In Fig. 2, there are three clans, reading left to right:  $pq$  (where both  $p$  and  $q$  are under  $AB$ ),  $p\bar{q}$  (where the point  $p$  is under  $AB$  but  $q$  is not), and  $\bar{p}\bar{q}$  (where neither  $p$  nor  $q$  are under  $AB$ ).

PaL diagrams will comprise a list of overlaps, which may be solid or dashed, together with a set of clans. For (abstract) PaL diagrams to properly correspond to their concrete (drawn) realisations, it must be the case that each overlap and clan in a diagram is in some sense similar. We require that each overlap has the same *underlying* word. In order to formalise this, we introduce the *rem* function that removes bars from letters; in the definition below,  $\lambda$  denotes the empty word and  $\alpha_i$  denotes a single letter.

**Definition 3** The *remove function*,

$$rem : \text{Overlap} \cup \text{Clan} \rightarrow \text{Overlap} \cup \text{Clan},$$

is defined recursively by:

- $rem(\lambda) = \lambda$ ,
- $rem(\alpha_i) = rem(\bar{\alpha}_i) = \alpha_i$ ,
- $rem(\alpha_i \cdot w) = rem(\alpha_i) \cdot rem(w)$ ,

where  $\cdot$  is the standard concatenation operator.

For example  $rem(\bar{A}BC) = ABC$  and  $rem(a\bar{b}c) = abc$ .

The last piece of syntax needed is *shading*. An overlap and clan sharing the same vertical space is either shaded or not shaded, indicated by  $\square$  or  $\blacksquare$ , respectively. Shading is used to place an upper bound on the size of the set represented by the overlap.

**Definition 4** A *unitary PaL diagram*,  $d$ , is a non-empty ordered list of 4-tuples  $(overlap_i, type_i, clan_i, shading_i)$  such that:

- $\forall i, j. rem(overlap_i) = rem(overlap_j)$ ,
- $type_i \in \{\text{---}, \text{—}\}$
- $\forall i, j. rem(clan_i) = rem(clan_j)$ ,
- $shading_i \in \{\square, \blacksquare\}$ .

The word  $rem(overlap_i)$  is called the **line-order** of  $d$ , denoted  $lo(d)$ ; the set of letters in  $lo(d)$  is called the **lines** of  $d$ , denoted  $l(d)$ ; the word  $rem(clan_i)$  is called the **point-order** of  $d$ , denoted  $po(d)$ ; and the set of letters in  $po(d)$  is called the **points** of  $d$ , denoted  $p(d)$ . The point  $p_i$  lies under  $overlap_j$ , denoted  $p_i \downarrow overlap_j$ , whenever  $p_i$  is unbarred in  $clan_j$ . The set of overlaps in  $d$  is denoted  $\mathcal{O}_d$ ; and the set of all  $w$  such that  $rem(w) = lo(d)$  is called the **allowable overlaps** of  $d$ , denoted  $\mathcal{AO}_d$ .

Much like repeated letters in overlaps, overlaps themselves can be repeated in a diagram. The drawn PaL diagram in Fig. 2 is, formally,

$$d = [(AB, \text{—}, pq, \square), (A\bar{B}, \text{—}, p\bar{q}, \blacksquare), (\bar{A}\bar{B}, \text{---}, \bar{p}\bar{q}, \square)].$$

In this diagram,  $l(d) = \{A, B\}$  and  $p(d) = \{p, q\}$ . The following, however, is *not* a PaL diagram:

$$[(ABCD, \text{—}, ab, \square), (AB\bar{C}, \text{---}, \bar{a}\bar{b}, \blacksquare)]$$

since  $rem(ABCD) = ABCD \neq ABC = rem(AB\bar{C})$ .

**Definition 5** Given a unitary PaL diagram  $d$ , we call the set  $\mathcal{O}_d(\text{—}) = \{O_i \in \mathcal{O}_d : type_i = \text{—}\}$  the **solid overlaps** of  $d$ ; we call the set  $\mathcal{O}_d(\blacksquare) = \{O_j \in \mathcal{O}_d : shading_j = \blacksquare\}$  the **shaded overlaps** of  $d$ ; we call the set  $\mathcal{P}_d(O) = \{p \in p(d) : p \downarrow O\}$  the **points lying under** overlap  $O$  in  $d$ ; and we call the set  $\mathcal{O}_d(p) = \{O \in \mathcal{O}_d : p \downarrow O\}$  the **overlaps over** point  $p$ .

We can then build up PaL diagrams using normal logical connectives and unitary PaL diagrams:

**Definition 6** A *PaL diagram* is defined inductively as follows:

- if  $d$  is a unitary PaL diagram then  $d$  is a PaL diagram;
- if  $d_1$  is a PaL diagram then  $\neg d_1$  is a PaL diagram where  $l(\neg d_1) = l(d_1)$  and  $p(\neg d_1) = p(d_1)$ ;
- if  $d_1$  and  $d_2$  are PaL diagrams and  $\diamond \in \{\wedge, \vee, \Rightarrow\}$  then  $(d_1 \diamond d_2)$  is a PaL diagram where  $l(d_1 \diamond d_2) = l(d_1) \cup l(d_2)$  and  $p(d_1 \diamond d_2) = p(d_1) \cup p(d_2)$ ;

Given a unitary PaL diagram, we now show how to draw that diagram. The process will produce a drawn diagram where all overlaps have equal length, although it is a simple matter to drop this restriction.

**Definition 7** Given a unitary PaL diagram

$$d = [(O_1, type_1, cl_1, sh_1), \dots, (O_n, type_n, cl_n, sh_n)]$$

we draw the **concrete diagram for  $d$**  as follows:

1. Write the word  $U \cdot lo(d) \cdot po(d)$  vertically downwards, followed by a vertical line of equivalent length.
2. Draw

$$(O_1, type_1, cl_1, sh_1) = (L_1 \dots L_m, type_1, p_1 \dots p_k, sh_1)$$

as follows:

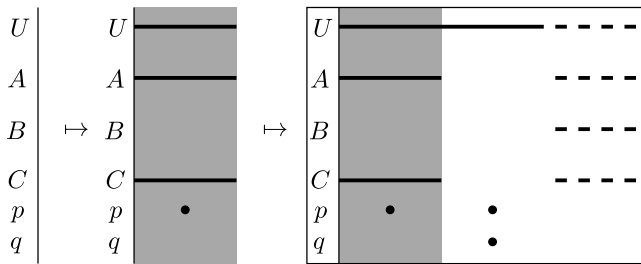
- If  $type_1 = \text{—}$  then draw a solid horizontal line of length 1 unit against  $U$  and, for each  $j = 1, \dots, m$ , if  $L_j$  is unbarred in  $O_1$  draw a solid horizontal line against the letter  $L_j$  of length 1 unit.
- If  $type_1 = \text{---}$  then draw a dashed horizontal line of length 1 unit against  $U$  and, for each  $j = 1, \dots, m$ , if  $L_j$  is unbarred in  $O_1$  draw a dashed horizontal line against the letter  $L_j$  of length 1 unit.
- For  $j = 1, \dots, k$ , if  $p_j$  is unbarred then draw • in the middle of the horizontal space of width 1 unit against the letter  $p_j$ .
- If  $sh_1 = \blacksquare$  then shade the entire vertical column one unit wide.

For each  $i = 2, \dots, n$ , repeat the process as for

$$(O_i, type_i, cl_i, sh_i),$$

moving along  $i - 1$  units before starting to draw lines (resp.  $i - \frac{1}{2}$  units for points).

3. Draw a box around the constructed elements.



**Figure 4. Drawing a PaL diagram**

As an example, consider the unitary PaL diagram

$$[(A\bar{B}C, \text{—}, p\bar{q}, \blacksquare), (\bar{A}\bar{B}\bar{C}, \text{—}, pq, \square), (ABC, \text{---}, \bar{p}\bar{q}, \square)].$$

The drawing stages for this diagram are shown in Fig. 4.

Given drawings of unitary PaL diagrams, we can combine them to form drawings of more general diagrams. Diagrammatically, negation is represented by crossing the diagram out, conjunction by juxtaposition, disjunction by drawing a horizontal line segment between diagrams (as in Shin's Venn-II system [12]) and implication by drawing a horizontal, single-headed arrow ( $\Rightarrow$ ) between diagrams. Concrete representations of these constructions are illustrated in Fig. 5, showing

- (a)  $\neg[(A\bar{B}C, \text{—}, p\bar{q}, \square)]$
- (b)  $[(A\bar{B}C, \text{—}, p\bar{q}, \square)] \wedge [(\bar{A}\bar{B}\bar{C}, \text{—}, pq, \square)]$
- (c)  $[(\bar{A}\bar{B}\bar{C}, \text{—}, p\bar{q}, \square)] \vee [(ABC, \text{—}, pq, \square)]$ , and
- (d)  $[(\bar{A}\bar{B}\bar{C}, \text{—}, p\bar{q}, \square)] \Rightarrow [(A\bar{B}C, \text{—}, p\bar{q}, \square)]$ .

We now have the syntax for PaL diagrams. We give them meaning in a natural way, given we are using them as representations of sets and elements. The lines and points are interpreted as subsets and elements of some universe, respectively. Formally:

**Definition 8** An **interpretation** is a pair  $(\mathcal{U}, I)$  where  $\mathcal{U}$  is called the **universal set** and  $I$  the function  $I : \mathcal{L} \rightarrow \mathbb{P}(\mathcal{U})$  ensures  $I(A) \subseteq \mathcal{U}$ . The function  $I$  can be extended to interpret barred letters and overlaps as follows:

- for each letter,  $A$ ,  $I(\bar{A}) = \mathcal{U} - I(A)$ , and
- for each overlap  $O$ ,

$$I(O) = \bigcap_{A \in \mathcal{L}_u} I(A) \cap \bigcap_{A \in \mathcal{L}_b} I(\bar{A})$$

where  $\mathcal{L}_u$  is the set of letters which appear unbarred in  $O$  and  $\mathcal{L}_b$  is the set of letters which appear as barred letters in  $O$ .

If  $\mathcal{U}$  is finite the **size** of the interpretation is  $|\mathcal{U}|$ .

Interpretations that agree with the intended meaning of a diagrams are called the diagram's **models**:

**Definition 9** Let

$$d = [(O_1, type_1, cl_1, sh_1), \dots, (O_n, type_n, cl_n, sh_n)]$$

be a unitary PaL diagram. An interpretation  $\mathcal{I} = (\mathcal{U}, I)$  is a **model** for  $d$ , denoted  $\mathcal{I} \models d$ , whenever there exists a function, namely  $\Phi : \mathcal{P} \rightarrow \mathcal{U}$ , mapping points to elements of  $\mathcal{U}$  satisfying:

1. **Point-location condition:** each point maps to an element in the set represented by an overlap under which the point lies:

$$\bigwedge_{p \in \mathcal{P}(d)} \left( \Phi(p) \in \bigcup_{O_i \in \mathcal{O}_d(p)} I(O_i) \right).$$

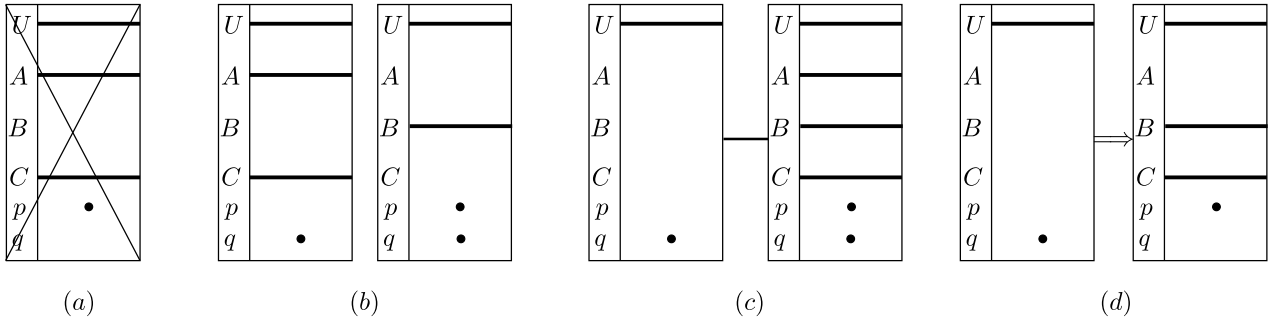


Figure 5. Components of PaL diagrams

2. **Point-distinctness condition:** no two distinct points map to the same element:

$$\bigwedge_{p_i, p_j \in p(d)} (p_i \neq p_j \Rightarrow \Phi(p_i) \neq \Phi(p_j)).$$

3. **Solid-overlap condition:** Solid overlaps represent non-empty sets:

$$\bigwedge_{O \in \mathcal{O}_d(\text{—})} (I(O) \neq \emptyset).$$

4. **Overlap-absence condition:** Overlaps that could be present given the line labels used represent empty sets:

$$\bigwedge_{O \in \mathcal{AO}_d - \mathcal{O}_d} (I(O) = \emptyset).$$

5. **Shaded-overlap condition:** in a shaded overlap, all elements are represented by points:

$$\bigwedge_{O_i \in \mathcal{O}_d(\blacksquare)} (I(O_i) \subseteq \{\Phi(p_i) : p_i \in cl_i\}).$$

If  $\mathcal{I}$  is a model for  $d$  then  $\mathcal{I}$  satisfies  $d$ .

As an example, take the final diagram in figure 4, and consider the interpretation  $(\mathcal{U}, I)$  where  $\mathcal{U} = \{1, 2, 3\}$ ,  $I(A) = \{1, 3\}$ ,  $I(B) = \{3\}$ ,  $I(C) = \{1, 3\}$ ,  $\Phi(p_1) = \{1\}$ ,  $\Phi(p_2) = \{2\}$ . We show that  $(\mathcal{U}, I) \models d$ . For every solid overlap present in  $d$ , we have  $I(O) \neq \emptyset$ . To illustrate, the second overlap is  $\bar{A}\bar{B}\bar{C}$ , giving interpretation  $(\mathcal{U} - I(A)) \cap (\mathcal{U} - I(B)) \cap (\mathcal{U} - I(C)) = \{2\} \cap \{1, 2\} \cap \{2\} = \{2\} \neq \emptyset$ . Further, the overlap  $\bar{A}\bar{B}\bar{C}$  (amongst others) is absent. The interpretation of this overlap is  $(\mathcal{U} - I(A)) \cap I(B) \cap (\mathcal{U} - I(C)) = \{2\} \cap \{3\} \cap \{2\} = \emptyset$ , as required. Consider the point  $p_2$ , lying under the overlap  $\bar{A}\bar{B}\bar{C}$ . We have already seen the interpretation of this overlap is  $\{2\}$ , and since  $\Phi(p_2) \in \{2\}$ , we have that the point-location condition is satisfied for  $p_2$ . Now, to satisfy

the point-location condition for  $p_1$ , we require that  $\Phi(p_1) \in \{1, 3\} \cup \{2\}$ , which holds. The shaded-overlap condition for the first overlap requires that  $I(A) \cap (\mathcal{U} - I(B)) \cap I(C) \subseteq \{\Phi(p_1)\}$ . Now, since  $\Phi(p_1) = 1$ , the condition is satisfied. All conditions are thus true, and so  $(\mathcal{U}, I) \models d$ .

Note that  $(\mathcal{U}, I)$  is not the only model for  $d$ . There is no maximum cardinality restriction on the number of elements in  $I(\bar{A}\bar{B}\bar{C})$ . Thus, keeping the function  $I$  the same, but changing  $\mathcal{U}$  to  $\{1, 2, 3, 4\}$  will still be a model for  $d$ , except now  $I(\bar{A}\bar{B}\bar{C}) = \{2, 4\}$ . By contrast, we cannot add extra elements to  $I(\bar{A}\bar{B}\bar{C})$ , since otherwise we would violate the shaded overlap condition. This observation that the model sets for unshaded overlaps can be extended, but the sets for shaded overlaps cannot necessarily be extended, will be crucial in section 5.

The interpretation  $(\{1, 2\}, I)$  where  $I(A) = I(B) = I(C) = \emptyset$ ,  $\Phi(p) = \{1\}$ ,  $\Phi(q) = \{2\}$  is likewise a model for  $d$ . This model illustrates the shaded-overlap condition requiring a subset relation, rather than equality. For, the interpretation of the first overlap is  $\emptyset$ , since  $p$  is interpreted as lying under the second overlap, so clearly  $I(\bar{A}\bar{B}\bar{C}) = \emptyset \neq \{1\}$ .

Consider the unitary diagram  $d = [(A, \text{—}, \lambda, \blacksquare)]$ , in other words the fully shaded diagram with one solid overlap,  $A$ , with no points lying under it. Consider further the an interpretation  $\mathcal{I}$ . The solid overlap presence condition tells us that  $I(A) \neq \emptyset$ . By contrast, the shaded overlap condition tells us  $I(A)$  is subset of the interpretations of the points lying under the overlap. Since no points lie under the overlap  $A$ , we have that  $I(A) \subseteq \emptyset$ . Thus, the conditions cannot all be true, and so  $\mathcal{I}$  is not a model for  $d$ , that is we say  $d$  is *unsatisfiable*. There are many unsatisfiable diagrams, but the **canonical unsatisfiable diagram**, denoted  $d_\perp$ , is defined to be:

$$d_\perp = [(\lambda, \text{—}, \lambda, \blacksquare)].$$

We need to define models for arbitrary PaL diagrams. This is straightforward:

**Definition 10** Given an interpretation  $\mathcal{I} = (\mathcal{U}, I)$  and a non-unitary PaL diagram  $d$ , we say that  $\mathcal{I}$  is a model for  $d$ ,



denoted  $\mathcal{I} \models d$ , based on the structure of  $d$ :

1. if  $d \equiv \neg d_1$ , then  $\mathcal{I} \models d$  whenever  $\mathcal{I} \not\models d_1$ ,
2. if  $d \equiv d_1 \vee d_2$ , then  $\mathcal{I} \models d$  whenever  $\mathcal{I} \models d_1$  or  $\mathcal{I} \models d_2$ ,
3. if  $d \equiv d_1 \wedge d_2$ , then  $\mathcal{I} \models d$  whenever  $\mathcal{I} \models d_1$  and  $\mathcal{I} \models d_2$ , and
4. if  $d \equiv d_1 \Rightarrow d_2$ , then  $\mathcal{I} \models d$  whenever  $\mathcal{I} \models d_1$  implies  $\mathcal{I} \models d_2$ .

Each dot against a point label represents disjunctive information. For example, in the right-most diagram of Fig. 5, there are two dots for the point  $p$ . This arrangement means that the point  $p$  will be interpreted either as the first dot, or the second, but not both. Conjunctive information about points, meanwhile, will be represented by duplicate point labels. Wherever a point-order for a diagram contains two instances of the same letter, say  $p$ , then unless there is some clan where both instances of  $p$  are unbarred, then the diagram will necessarily represent a contradiction.

The use of multiple dots against a single point label to represent disjunction provides a compact notation. If points were singular (could only consist of a single dot), then disjunction would have to be represented as a disjunction of unitary PaL diagrams. Consider the simple case of a diagram for Venn-2, together with a pair of points  $p$  and  $q$  where the only information we know is that  $p$  and  $q$  are distinct points. In other words, we do not know in which sets their interpretations are. This situation is illustrated in Fig. 6. If points were instead singular, then we would need 10 unitary diagrams connected by disjunctions to represent the same information. Suppose that the interpretation of  $p$  is an element of the intersections of the interpretations of  $A$  and not  $B$ . Then, there are 4 possible locations for the interpretation of  $q$ , requiring 4 separate representative diagrams (the first four components of Fig. 7). Suppose, instead, that the interpretation of  $p$  is an element of the intersection of the interpretations of  $A$  and  $B$ . Recalling that the points represent variables, not constants, there are now only 3 possible locations for the interpretation of  $q$ . Continuing in this way, we see that there are  $4 + 3 + 2 + 1 = 10$  different unitary diagrams needed when we restrict points to be singular. The diagram in Fig. 7 represents the same information as the in Fig. 6 yet the latter is more compact than the former. In general, where  $m$  points each lie on  $n$  overlaps, then the number of disjuncts needed if each point is to be singular is  $O(n^m)$ , although proving this simple result is outside the scope of this paper.

### 3. Monadic First-Order Logic with Equality

To show that PaL diagrams are exactly as expressive as monadic first-order logic with equality (MFOL[=]), we aim

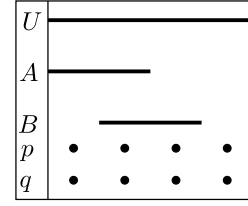


Figure 6. Points as disjunctive information

to provide translations between sentences in MFOL[=] and diagrams. In order to do this, we first give a brief survey of some necessary results about MFOL[=].

A *monadic predicate symbol* is one which takes a single argument. For example,  $L(x)$  uses the monadic predicate symbol  $L$ , whereas  $D(x, y)$  uses the dyadic predicate symbol  $D$ . MFOL[=] has only one special dyadic predicate, that of equality. We use  $\mathcal{L}$  as the set of monadic predicate symbols, thus treating the line labels as monadic predicates. Further, we take the set of variables to include all points in  $\mathcal{P} = \{p_1, p_2, \dots\}$ , and sometimes use the more usual  $x, y, \dots$  as well. Using points as variables makes definitions later in the paper more straightforward (variables will arise from points in our translations). Using  $x$  and  $y$  as variables distinguishes them from those arising from points.

Briefly, formulae in MFOL[=] are defined using monadic predicate symbols, variables,  $=$ , logical connectives and quantifiers in the standard way. Given a formula,  $\psi$ , if  $\psi$  has no free variables (i.e. variables that are not bound by a quantifier) then  $\psi$  is a **sentence**. Every formula can be turned into a semantically equivalent sentence by binding the free variables with universal quantifiers. As we use the standard syntax and semantics of MFOL[=] full details of the syntax and semantics are omitted; unfamiliar readers are referred to [2]. Firstly, we define the notion of expressive equivalence:

**Definition 11** A diagram and a sentence are **expressively equivalent** whenever they have the same set of models.

In order to prove the expressive equivalence between PaL diagrams and MFOL[=], we compare the model sets of each. As an example, consider the MFOL[=] sentence  $S = \exists x. A(x)$ . Take the interpretation  $\mathcal{I} = (\{1, 2\}, \{(A, \{1\}), \dots\})$ . This interpretation is a model for  $S$ , whereas any interpretation with  $I(A) = \emptyset$  is not:  $S$  stipulates that any interpretation of  $A$  must be non-empty.

We briefly summarise some results about models for MFOL[=] contained in [16]. In particular, we include the definitions that are needed to state a key theorem about MFOL[=] sentences, encapsulating the fact that each sentence,  $S$ , has a finite set of ‘small’ models (formally defined later) that can be used to generate all models of  $S$ . These small models are crucial for constructing a diagram with

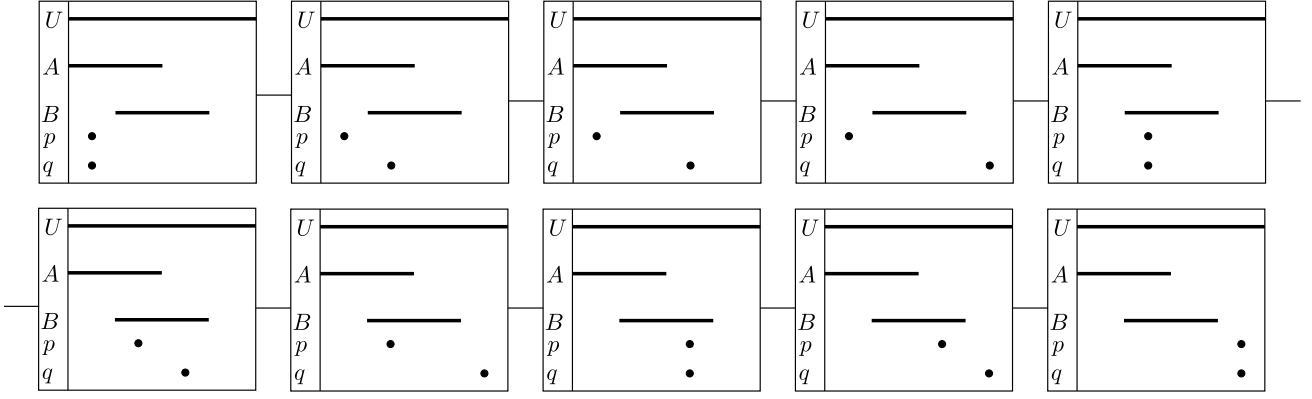


Figure 7. Disjunctive blow-up

the same meaning as  $S$ . Like diagrams, each sentence  $S$  in  $\text{MFOL}[=]$  may have many models. The first step is to identify the interpretation of sets of predicate symbols, akin to the interpretation of overlaps:

**Definition 12** Let  $\mathcal{I}$  be a interpretation with universal set  $\mathcal{U}$  and let  $X$  and  $Y$  be finite subsets of  $\mathcal{L}$ . The **predicate intersection set** in  $\mathcal{I}$  with respect to  $X$  and  $Y$ , denoted  $PI(\mathcal{I}, X, Y)$ , is given by:

$$PI(\mathcal{I}, X, Y) = \bigcap_{L_i \in X} I(L_i) \cap \bigcap_{L_i \in Y} (\mathcal{U} - I(L_i))$$

where  $\bigcap_{L_i \in \emptyset} I(L_i) = \bigcap_{L_i \in \emptyset} (\mathcal{U} - I(L_i)) = \mathcal{U}$  [16].

Given a sentence  $S$ , denote by  $q(S)$  and  $P(S)$  the quantifier rank of  $S$  [5] and set of predicates in  $S$ , respectively; recall, the quantifier rank of  $S$  is the maximum number of nested quantifiers in  $S$ . A sentence with quantifier rank of  $n$  can contain at most  $n$  distinct variables within the body of the sentence. Now, to limit the cardinality of a predicate intersection set to, say,  $m$  we need  $m + 1$  distinct variable names. To see this, consider the sentence  $\exists p_1. \forall p_2. p_1 = p_2$ . Any model for this sentence must have size 1, since the sentence tells us some element exists, and every element is equal to it. This argument is easily extended to arbitrary  $m > 1$ . Given a model for  $S$  and a predicate intersection,  $PI$ , set with cardinality at least  $q(S)$ , elements can be added to  $PI$  and the resulting interpretation is still a model for  $S$ . By contrast, if  $PI$  has cardinality less than  $q(S)$  then elements cannot necessarily be added to it. Given any model for  $S$ , we can identify which predicate intersection sets can safely be extended with extra elements. Formally:

**Definition 13** Let  $S$  be a sentence and let  $\mathcal{I}_1$  be a model for  $S$ . An  **$S$ -extension** of  $\mathcal{I}_1$  is an interpretation,  $\mathcal{I}_2$ , such that for each subset  $X$  of  $P(S)$ :

$$PI(\mathcal{I}_1, X, P(S) - X) \subseteq PI(\mathcal{I}_2, X, P(S) - X)$$

with equality whenever  $|PI(\mathcal{I}_1, X, P(S) - X)| < q(S)$  [16].

**Definition 14** Let  $S$  be a sentence and  $\mathcal{I}$  be a model for  $S$ . If the cardinality of  $\mathcal{I}$  is at most  $2^{|P(S)|} q(S)$  then we say  $\mathcal{I}$  is a **small model** for  $S$  [16].

Given a sentence  $S$  we have that  $q(S)$  and  $2^{|P(S)|}$  are finite, and so there are finitely many candidate interpretations which can be small models for  $S$ . We say two interpretations  $\mathcal{I}_1 = (\mathcal{U}_1, I_1)$  and  $\mathcal{I}_2 = (\mathcal{U}_2, I_2)$  are **isomorphic restricted to  $P(S)$**  iff there exists an isomorphism between  $\mathcal{I}_1$  and  $\mathcal{I}_2$  when the domains of  $I_1$  and  $I_2$  are both restricted to  $P(S)$ .

**Definition 15** Let  $S$  be a sentence. A set of small models,  $c(S)$ , is called a **classifying set of models** for  $S$  if for each small model  $m_1$  for  $S$ , there exists a unique  $m_2 \in c(S)$  such that  $m_1$  and  $m_2$  are isomorphic restricted to  $P(S)$  [16].

In other words, a classifying set for  $S$  is the smallest possible set of small models for  $S$ . We can create  $S$ -extensions of the small models for  $S$  to create more models for  $S$ . Such extensions form a set called the **cone**:

**Definition 16** The **cone** of  $\mathcal{I}_1$  given  $S$ , denoted  $\text{cone}(\mathcal{I}_1, S)$ , is a class of interpretations such that  $\mathcal{I}_2 \in \text{cone}(\mathcal{I}_1, S)$  iff  $\mathcal{I}_2$  is isomorphic to some  $S$ -extension of  $\mathcal{I}_1$  [16].

Finally, we have the key theorem needed for our expressiveness result:

**Theorem 1** Let  $S$  be a sentence and let  $c(S)$  be a classifying set of models for  $S$ . Then  $\bigcup_{m \in c(S)} \text{cone}(m, S)$  is precisely the set of models for  $S$  [16].

## 4. Sentences for diagrams

To show that every diagram can be turned into a sentence in  $\text{MFOL}[=]$ , we translate the conditions from definition 9 into formulae in  $\text{MFOL}[=]$ . Given a unitary PaL diagram we need to know how to translate the overlaps, the clans and the shading into  $\text{MFOL}[=]$  formulae. We first define a formula for an overlap, regardless of whether it is solid, dashed or shaded.

**Definition 17** Let  $O$  be an overlap with unbarred letters  $\mathcal{L}_u$  and barred letters  $\mathcal{L}_b$ . The **overlap formula** for  $O$ , denoted  $\mathcal{F}(O, x)$ , is given by:

$$\mathcal{F}(O, x) = \bigwedge_{L \in \mathcal{L}_u} L(x) \wedge \bigwedge_{L \in \mathcal{L}_b} \neg L(x).$$

We can now define the translation of a unitary PaL diagram to a  $\text{MFOL}[=]$  sentence.

**Definition 18** Let

$$d = [(O_1, \text{type}_1, \text{cl}_1, \text{sh}_1), \dots, (O_n, \text{type}_n, \text{cl}_n, \text{sh}_n)]$$

be a unitary PaL diagram, where  $p(d) = \{p_1, \dots, p_s\}$ .

- The **point-location formula** for  $d$ , denoted  $\mathcal{F}_{PL}(d)$ , is a conjunction, over all points, of the disjunctive information given by each point of  $d$ :

$$\mathcal{F}_{PL}(d) = \bigwedge_{p_i \in p(d)} \left( \bigvee_{O \in \mathcal{O}_d(p_i)} \mathcal{F}(O, p_i) \right).$$

- The **point-distinctness formula** for  $d$ , denoted  $\mathcal{F}_{PD}(d)$ , is a conjunction, over all distinct points, of inequalities:

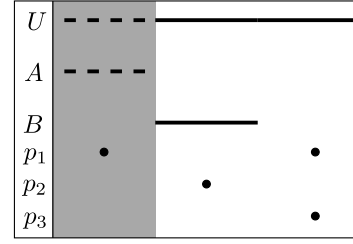
$$\mathcal{F}_{PD}(d) = \bigwedge_{p_i, p_j \in p(d) \wedge i \neq j} p_i \neq p_j.$$

- The **solid-overlap formula** for  $d$ , denoted  $\mathcal{F}_{SoO}(d)$ , is a conjunction, over all solid overlaps, of the existential formulae:

$$\mathcal{F}_{SoO}(d) = \bigwedge_{O \in \mathcal{O}_d(\text{—})} \exists x. \mathcal{F}(O, x).$$

- The **overlap-absence formula** for  $d$ , denoted  $\mathcal{F}_{OA}(d)$ , is a conjunction, over all allowable overlaps that are absent from  $d$ , of negated existential formulae stating that no elements lie under the absent overlaps:

$$\mathcal{F}_{OA}(d) = \bigwedge_{O \in \mathcal{A}\mathcal{O}_d - \mathcal{O}_d} \neg \exists x. \mathcal{F}(O, x).$$



**Figure 8. A PaL diagram to be translated to  $\text{MFOL}[=]$**

- The **shaded-overlap formula** for  $d$ , denoted  $\mathcal{F}_{ShO}(d)$ , is a conjunction, over all shaded overlaps, of universal formulae stating that if an element lies under an overlap, then it must be one of the points lying under that overlap:

$$\mathcal{F}_{ShO}(d) = \bigwedge_{O \in \mathcal{O}_d(\blacksquare)} \left( \forall y. \left( \mathcal{F}(O, y) \Rightarrow \left( \bigvee_{p \in p_d(O)} y = p \right) \right) \right).$$

The  **$\text{MFOL}[=]$  sentence for  $d$** , denoted  $S(d)$ , is then given by:

$$S(d) = \exists p_1, \dots, p_s. \mathcal{F}_{PL}(d) \wedge \mathcal{F}_{PD}(d) \wedge \mathcal{F}_{SoO}(d) \wedge \mathcal{F}_{ShO}(d).$$

Note that, if  $\text{type}_i = \dots$  for some overlap then the overlap itself gives us no information (although the points lying under it might). Hence, there is no formula created from the dashed overlaps, just as no condition arose in definition 9.

Consider the diagram  $d$  in figure 8. We create an  $\text{MFOL}[=]$  sentence for this diagram using the following formulae:

$$\begin{aligned} \mathcal{F}_{PL}(d) &= (A(p_1) \wedge \neg B(p_1)) \vee (\neg A(p_1) \wedge \neg B(p_1)) \wedge \\ &\quad (\neg A(p_2) \wedge B(p_2)) \wedge \\ &\quad (\neg A(p_3) \wedge \neg B(p_3)), \\ \mathcal{F}_{PD}(d) &= (p_1 \neq p_2 \wedge p_1 \neq p_3 \wedge p_2 \neq p_3), \\ \mathcal{F}_{SoO}(d) &= \exists x. (\neg A(x) \wedge B(x)) \wedge \exists x. (\neg A(x) \wedge \neg B(x)), \\ \mathcal{F}_{OA}(d) &= \neg \exists x. (A(x) \wedge B(x)), \\ \mathcal{F}_{ShO}(d) &= \forall y. ((A(y) \wedge \neg B(y)) \Rightarrow (y = p_1)). \end{aligned}$$

Thus the sentence for  $d$  is:

$$S(d) = \exists p_1 p_2 p_3. \mathcal{F}_{PL}(d) \wedge \mathcal{F}_{PD}(d) \wedge \mathcal{F}_{SoO}(d) \wedge \mathcal{F}_{OA}(d) \wedge \mathcal{F}_{ShO}(d).$$

**Theorem 2** Every unitary PaL diagram  $d$  is expressively equivalent to  $S(d)$ .

The proof is straightforward: each part of the sentence corresponds to an encoding in  $\text{MFOL}[=]$  of the conditions in definition 9. It is also immediate how to extend the result to arbitrary PaL diagrams:

**Definition 19** Let  $d$  be a PaL diagram. The **MFOL**[ $=$ ] **sentence for  $d$** , denoted  $\text{Sen}(d)$ , is given by induction on the structure of  $d$ :

- if  $d$  is a unitary diagram then  $\text{Sen}(d)$  is already defined,
- if  $d$  is  $\neg d_1$ , then  $\text{Sen}(d) = \neg \text{Sen}(d_1)$ ,
- if  $d$  is  $d_1 \diamond d_2$ , where  $\diamond \in \{\wedge, \vee, \Rightarrow\}$ , then  $\text{Sen}(d) = (\text{Sen}(d_1) \diamond \text{Sen}(d_2))$ .

**Theorem 3** Every PaL diagram  $d$  is expressively equivalent to  $\text{Sen}(d)$ .

## 5. Diagrams for sentences

In order to construct a diagram for the sentence  $S$ , we need only construct diagrams for the small models in  $c(S)$ , and take the disjunction of these diagrams. In what follows we let  $PI_{X,\mathcal{I},S} = |PI(\mathcal{I}, X, P(S) - X)|$ . The process for drawing a diagram for each small model is straightforward. First, if  $c(S)$  contains the interpretation with  $|\mathcal{I}| = 0$ , then this model gives rise to the **empty diagram**  $d_\emptyset = [(\lambda, \dots, \lambda, \blacksquare)]$ . This diagram contains no points or lines other than the  $U$  line, and the shading asserts that the universe is empty. For non-empty models, we assign solid overlaps to non-empty predicate intersection sets and we create the same number of points lying under this overlap as the cardinality of the predicate intersection set. We shade those overlaps where the associated predicate intersection set has smaller size than the quantifier rank of the sentence. Formally:

**Definition 20** Let  $\mathcal{I}$  be a small model for a MFOL[ $=$ ] sentence  $S$  and suppose  $|\mathcal{I}| = m$  and  $|P(S)| = n$  where  $P(S) = \{L_1, \dots, L_n\}$ . Let the set  $\{X : X \subseteq P(S) \wedge PI_{X,\mathcal{I},S} > 0\} = \{X_1, \dots, X_N\}$  be ordered. The **PaL diagram  $d$  representing  $\mathcal{I}$  given  $S$** , denoted  $\mathcal{D}(\mathcal{I}, S) = d$  is defined as follows:

1. If  $|\mathcal{I}| = 0$ , then  $d = d_\emptyset$ .
2. Otherwise, the line labels are the predicate symbols in  $P(S)$  and set  $\text{lo}(d) = L_1 \dots L_n$ .
3. There is one point label for each element of  $\mathcal{U}$ :

$$p(d) = \{p_1, \dots, p_m\}$$

and set  $\text{po}(d) = p_1 \dots p_m$ .

4. For  $i = 1, \dots, N$  construct the overlap  $(L_1 \dots L_n, \text{---}, p_1 \dots p_m, sh_i)$  where:

- $L_j$  is unbarred iff  $L_j \in X_i$ ,

- $p_k$  is unbarred iff:

$$\sum_{j=1}^{i-1} PI_{X_j, \mathcal{I}, S} < k \leq \sum_{j=1}^i PI_{X_j, \mathcal{I}, S}$$

where we define  $\sum_{j=1}^0 PI_{X_j, \mathcal{I}, S} = 0$ ,

- $sh_i = \blacksquare$  iff  $PI_{X_i, \mathcal{I}, S} < q(S)$ .

We illustrate the process of determining whether points are unbarred with an example. Suppose a model has  $X_1, X_2$  and  $X_3$  as the only sets where  $PI_{X,\mathcal{I},S} > 0$ , and  $PI_{X_1, \mathcal{I}, S} = 3, PI_{X_2, \mathcal{I}, S} = 4$  and  $PI_{X_3, \mathcal{I}, S} = 2$ . The values  $k$  can take such that  $PI_{X_1, \mathcal{I}, S} < k \leq PI_{X_1, \mathcal{I}, S} + PI_{X_2, \mathcal{I}, S}$  are 4, 5, 6 and 7. Then the unbarred points in the second clan would be  $p_4, p_5, p_6$  and  $p_7$ , meaning  $p_1, p_2, p_3, p_8$  and  $p_9$  would be barred. A consequence of definition 20 is that every point lies on exactly one overlap.

Having defined the diagrams representing an interpretation given a sentence, we now define the diagrams for the sentence:

**Definition 21** Given an MFOL[ $=$ ] sentence  $S$  with classifying models  $c(S)$ , the **diagram representing  $S$** , denoted  $\mathcal{D}(S)$ , is given by:

$$\mathcal{D}(S) = d_\perp \vee \bigvee_{\mathcal{I} \in c(S)} \mathcal{D}(\mathcal{I}, S)$$

We demonstrate definition 21 using an example. One of the small models of the sentence  $S = \exists x. A(x) \vee \forall x. A(x)$  is given by  $\mathcal{I} = (\{1, 2\}, I)$  where  $I(A) = \{1\}$ . There is a single predicate symbol in  $S$ , and so the line order for  $d$  is simply as  $A$ . Furthermore, the set  $X_1 = \{A\}$  and  $X_2 = \emptyset$  are the only sets for which  $PI_{X,\mathcal{I},S} > 0$ . So, the number of overlaps in  $\mathcal{D}(\mathcal{I}, S)$  is 2. In this example,  $|\mathcal{I}| = 2$  so we require 2 points in each clan. The point order of  $d$  is given by  $p_1 p_2$ .

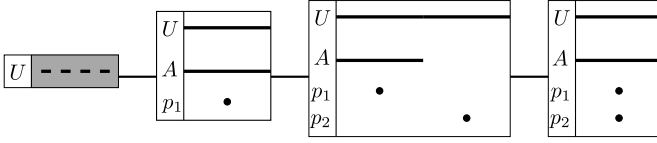
The first overlap,  $(A, \text{---}, p_1 \bar{p}_2, \square)$ , is unshaded since  $PI_{X_1, \mathcal{I}, S} = 1 \geq q(S) = 1$ . Also, since  $0 < 1 \leq 1$ ,  $p_1$  is unbarred but  $p_2$  is not. Similarly, the second overlap is  $(\bar{A}, \text{---}, \bar{p}_1 p_2, \square)$ .

The rest of the small models for  $S$  are:

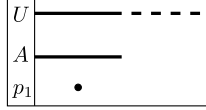
1.  $\mathcal{I}_1 = (\emptyset, \emptyset)$ ,
2.  $\mathcal{I}_2 = (\{1\}, \{(A, \{1\})\})$ ,
3.  $\mathcal{I} = \mathcal{I}_3 = (\{1, 2\}, \{(A, \{1\})\})$ , and
4.  $\mathcal{I}_4 = (\{1, 2\}, \{(A, \{1, 2\})\})$ .

The diagram for this sentence is given in figure 9.

The first model,  $\mathcal{I}_1$ , comes from the vacuous satisfaction of  $\forall x. A(x)$ , giving rise to the empty diagram. We have already shown the diagram for the model  $\mathcal{I}_3$ , and the



**Figure 9.** The PaL diagram for  $\exists x. A(x) \vee \forall x. A(x)$



**Figure 10.** A more natural diagram for  $\exists x. A(x)$

rest are left for the reader to verify. We note that the diagram in figure 9 could be considered a relatively natural PaL diagram for the sentence. We could write  $\forall x. A(x)$  using the law of excluded middle: either no  $x$  exists, or every  $x$  has  $A(x)$ . Thus, the first two disjuncts of figure 9, i.e.  $d_\emptyset \vee [(A, \text{---}, p_1, \square)]$ , represent the universal part of the sentence. The former covers the case where no  $x$  exists (and so  $\forall x. A(x)$  is trivially satisfied), and the latter covers the case where  $\forall x. A(x)$  holds for non-empty universes. The existential part of the sentence can be thought of as follows: we know an  $x$  exists such that  $A(x)$ , but this does not preclude other  $y$  existing such that  $\neg A(y)$ . In order to capture this situation, we would use the diagram in figure 10. The overlap  $(\bar{A}, \dots, \bar{p}_1, \square)$  tells us only that it is possible for such  $y$  to exist, but does not necessitate its existence. We could thus replace the third and fourth components of figure 9 with the PaL diagram from figure 10.

The purpose of definitions 20 and 21 was not to produce the most natural diagrams for sentences (an interesting challenge for future work), however, but rather to demonstrate that every sentence could be translated into some PaL diagram.

**Theorem 4** *Let  $S$  be a sentence. Then  $\mathcal{D}(S)$  is expressively equivalent to  $S$ .*

**Proof. (Sketch)** We show that definition 20 provides a diagram with models for  $S$ , and the corresponding proof for definition 21 follows immediately. Let  $\mathcal{I}$  be a small model for  $S$  in  $c(S)$ , and let  $\mathcal{D}(\mathcal{I}, S)$  be constructed according to definition 20. We show that any model  $\mathcal{I}_1 \in \text{cone}(\mathcal{I}, S)$  is a model for  $\mathcal{D}(\mathcal{I}, S)$ , omitting the details of the converse.

If  $\mathcal{I}_1 \in \text{cone}(\mathcal{I}, S)$  then it has been extended from  $\mathcal{I}$  in such a way that

$$PI(\mathcal{I}, X, P(S) - X) = PI(\mathcal{I}_1, X, P(S) - X)$$

whenever  $PI_{X, \mathcal{I}, S} < q(S)$ , and

$$PI(\mathcal{I}, X, P(S) - X) \subseteq PI(\mathcal{I}_1, X, P(S) - X)$$

otherwise. Let  $O$  be the overlap associated with the set  $X$ , and thus  $|I(O)| = PI_{X, \mathcal{I}_1, S} \geq PI_{X, \mathcal{I}, S}$ . There are  $PI_{X, \mathcal{I}, S}$  distinct points lying under  $O$  and at least  $PI_{X, \mathcal{I}, S}$  distinct elements in  $I(O)$ , so we can thus assign each point lying under  $O$  to a distinct element of  $I(O)$  using a function  $\Phi$ . In this way, we have satisfied the point location and point distinctness conditions.

If  $PI_{X, \mathcal{I}, S} < q(S)$  then  $O$  is shaded and  $|I(O)| = PI_{X, \mathcal{I}, S}$ . Since every point of  $\mathcal{D}(\mathcal{I}, S)$  lies under exactly one overlap, and there are  $PI_{X, \mathcal{I}, S}$  of them by the point distinctness and location conditions being satisfied we have that

$$\bigcup_{p \in \mathcal{P}_d(O)} \Phi(p) = I(O).$$

Trivially, then,  $I(O) \subseteq \bigcup_{p \in \mathcal{P}_d(O)} \Phi(p)$ , and thus the shaded overlap condition is satisfied.

If  $PI_{Y, \mathcal{I}_1, S} = 0$  then we draw no overlap, meaning  $O \in \mathcal{AO}_{\mathcal{D}(\mathcal{I}, S)} - \mathcal{O}_{\mathcal{D}(\mathcal{I}, S)}$ . However, since we associate  $O$  with  $Y$ , we also have  $I(O) = \emptyset$ , and thus the overlap absence condition is satisfied. Put another way, since we only draw an overlap when  $PI_{X, \mathcal{I}_1, S} \neq 0$ , every  $X$  such that  $PI_{X, \mathcal{I}_1, S} \neq 0$  is assigned to an overlap  $O$  which has  $I(O) \neq \emptyset$ , satisfying the solid overlap presence condition. Then, the model  $\mathcal{I}_1$  is also a model for  $\mathcal{D}(\mathcal{I}, S)$ , as required. Hence, by theorem 1, every model for  $S$  is a model for  $\mathcal{D}(\mathcal{I}, S)$ . ■

We have seen how any  $\text{MFOL}[=]$  sentence  $S$  can be translated into a PaL diagram, and any PaL diagram can be translated into an  $\text{MFOL}[=]$  sentence. Therefore:

**Theorem 5** *PaL diagrams and  $\text{MFOL}[=]$  are equally expressive.*

The main result of [16] was that spider diagrams are equally expressive as  $\text{MFOL}[=]$ . We thus have the following:

**Theorem 6** *PaL diagrams,  $\text{MFOL}[=]$  and spider diagrams are all expressively equivalent.*

## 6. Conclusion and Further Work

The results from [3, 11] demonstrated that linear diagrams have the potential to be an effective visual language in the areas where Euler and Venn diagrams are currently used. To explore and exploit this observation, we have extended linear diagrams to PaL diagrams by adding points. Moreover, we have formalised the syntax and semantics of PaL diagrams and shown they are capable of expressing

exactly the same statements as MFOL[=]. As a corollary, we immediately have that PaL diagrams are exactly as expressive as spider diagrams which extend Euler diagrams. PaL diagrams provide a solid foundation for further development of linear-based notations, both in terms of expressiveness and reasoning.

There are a number of clear directions for further work. The creation of a system of inference rules for PaL diagrams is of particular interest. The reasoning rules for Euler-based diagrams with an equivalent level of expressiveness demonstrate that a sound and complete system of rules is possible. Given a sound and complete reasoning system for PaL diagrams, it should be possible to integrate these diagrams into a heterogeneous system such as Diabelli [17].

In the usability direction, we will seek layout guidelines for PaL diagrams that aid understanding. For example, to what extent is the ordering of the lines important? Is interleaving of the points and lines ever beneficial? By answering these questions, and others, empirically we seek to develop conditions akin to the well-formed conditions of Euler diagrams [15]. However, rather than these conditions being purely theoretical, to be tested empirically later, the empirical method will drive their development. In other words, we will develop a set of ease-of-understanding guidelines.

In terms of comparative usability, now that we have a system equivalent to spider diagrams, it is possible to extend the results of [3], which established linear diagrams' superiority over Euler diagrams, to notations equivalent to MFOL[=]. We hypothesise that PaL diagrams, when compared to spider diagrams, will retain the efficacy that linear diagrams have over Euler diagrams.

## References

- [1] F. Bellucci, A. Moktefi, and A.-V. Pietarinen. Diagrammatic autarchy. volume 1132. CEUR, 2014.
- [2] S. Burris. *Logic for Mathematics and Computer Science*. Prentice Hall, 1998.
- [3] P. Chapman, G. Stapleton, P. Rodgers, L. Micallef, and A. Blake. Visualizing Sets: An Empirical Comparison of Diagram Types, 2014.
- [4] L. Couturat. *Opusculs et fragments inédits de Leibniz*. Felix Alcan, 1903.
- [5] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Springer-Verlag, 1991.
- [6] G. Englebretsen et al. Linear diagrams for syllogisms (with relationals). *Notre Dame Journal of Formal Logic*, 33(1):37–69, 1991.
- [7] J. Gil, J. Howse, and S. Kent. Constraint diagrams: A step beyond UML. In *Proceedings of TOOLS USA 1999, Santa Barbara, California, USA*, pages 453–463. IEEE Computer Science Press, August 1999.
- [8] H. Hofmann, A. Siebes, and A. Wilhelm. Visualizing Association Rules with Interactive Mosaic Plots. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 227–235. ACM, 2000.
- [9] O. Lemon and I. Pratt. On the insufficiency of linear diagrams for syllogisms. *Notre Dame Journal of Formal Logic*, 39(4):573–580, 1998.
- [10] P. Rodgers. A survey of Euler diagrams. *J. Vis. Lang. Comput.*, 25(3):134–155, 2014.
- [11] Y. Sato and K. Mineshima. The Efficacy of Diagrams in Syllogistic Reasoning: A Case of Linear Diagrams. In *Diagrammatic Representation and Inference*, Lecture Notes in Computer Science, pages 352–355. Springer, 2012.
- [12] S.-J. Shin. *The Logical Status of Diagrams*. Cambridge University Press, 1994.
- [13] G. Stapleton and A. Delaney. Evaluating and generalizing constraint diagrams. *Journal of Visual Languages and Computing*, 19(4):499–521, 2008.
- [14] G. Stapleton and J. Masthoff. Incorporating negation into visual logics: A case study using Euler diagrams. In *Visual Languages and Computing 2007*, pages 187–194. Knowledge Systems Institute, 2007.
- [15] G. Stapleton, P. Rodgers, J. Howse, and J. Taylor. Properties of Euler diagrams. In *Proceedings of Layout of Software Engineering Diagrams*, pages 2–16. EASST, 2007.
- [16] G. Stapleton, S. Thompson, J. Howse, and J. Taylor. The expressiveness of spider diagrams. *Journal of Logic and Computation*, 14(6):857–880, December 2004.
- [17] M. Urbas and M. Jamnik. Heterogeneous Proofs: Spider Diagrams Meet Higher-Order Provers. In *2nd International Conference on Interactive Theorem Proving*, pages 376–382, 2011.
- [18] K. Wittenburg, T. Lanning, M. Heinrichs, and M. Stanton. Parallel Bargrams for Consumer-based Information Exploration and Choice. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 51–60. ACM, 2001.



# Local Context-based Recognition of Sketched diagrams

Gennaro Costagliola, Mattia De Rosa, Vittorio Fuccella  
Dipartimento di Informatica, University of Salerno  
Via Giovanni Paolo II, 84084 Fisciano (SA), Italy  
{gencos, matderosa, vfuccella}@unisa.it

## Abstract

*We present a new methodology aimed at the design and implementation of a framework for sketch recognition enabling the recognition and interpretation of diagrams. The diagrams may contain different types of sketched graphic elements such as symbols, connectors, text. Once symbols are distinguished from connectors and identified, the recognition proceeds by identifying the local context of each symbol. This is seen as the symbol interface exposed to the rest of the diagram and includes predefined attachment areas on each symbol. We argue that, in many cases, simple constraints on the local context of each symbol are enough to describe diagram languages defined on those symbols. Further refinement and interpretation of the set of acceptable diagrams is then provided through a visual grammar. We also describe the architecture of the framework and provide sample applications for the domains of flowcharts and binary trees.*

**Keywords:** *sketch recognition, multi-domain, methodology, framework, visual languages*

## 1 Introduction

The use of diagrams is common in various disciplines. Typical examples include maps, line graphs, bar charts, engineering blueprints, architects' sketches, hand drawn schematics, etc. In general, diagrams can be created either by using pen and paper, or by using specific computer programs. These programs provide functions to facilitate the creation of the diagram, such as copy-and-paste, but the classic WIMP interfaces they use are unnatural when compared to pen and paper. Indeed, it is not rare that a designer prefers to use pen and paper at the beginning of the design [32], and then transfer the diagram to the computer later [2].

To avoid this double step, a solution is to allow users to sketch directly on the computer. This requires both specific hardware and sketch recognition based software. As

regards hardware, many pen/touch based devices such as tablets, smartphones, interactive boards and tables, etc. are available today, also at reasonable costs. Sketch recognition is needed when the sketch must be processed and not considered as a simple image and it is crucial to the success of this new modality of interaction. It is a difficult problem due to the inherent imprecision and ambiguity of a freehand drawing and to the many domains of application.

A central element of sketching is the stroke. On a touch screen, a stroke starts with the pressure of the pen (or finger) on the screen and ends when the pen is raised. Technically, a stroke is a finite list of triples  $(x, y, t)$  (or samples) where  $(x, y)$  are the pair of coordinates in which the pen was at the time  $t$ . The strokes are often preprocessed in order to extract the basic primitives from them.

In the literature we can find the description of many different frameworks for the recognition of diagrams. There are both solutions developed for specific domains [17, 13, 27] and multi-domain solutions [2, 4, 26, 18, 28, 14]. In the multi-domain frameworks, the low-level recognition is usually performed independently from the context, through the identification of graphical primitives (lines, arcs, ellipses, etc.). In the most advanced products, the domain knowledge is then used at a higher level, to correct possible low-level interpretation errors.

In this paper we present a new methodology aimed at developing a framework for the recognition of sketched diagrams from different domains. The main innovation regards the introduction of a recognition phase based on the analysis of the local context of symbols. This results to be effective since many visual languages need to be simple in order to be used, and as a result their structure happens to be simple enough to be captured with local checks. We prove this statement by showing that even a complex enough flowchart dialect can be fully syntactically modelled through this approach. From the point of view of sketch recognition one of the innovations introduced with this framework is that it learns directly from sample sketches of the specific domain the information used for low-level recognition, taking advantage of the various innovative machine learning-based

techniques produced in recent years (see next section).

The framework is logically composed of four layers: Text/Graphic Separation, Symbol Recognition, Local Context Detection and Diagram Recognition/Parsing. The first three layers are mostly pattern recognition processes to extract intermediate information from the strokes and to perform symbol and local context recognition. They include different modules to perform stroke segmentation, symbol identification in the diagram and the recognition of the attachment areas needed to connect the symbols to each other. The last layer consists of two modules. The first one is the Local Context-based Diagram Recogniser and validates the scanned diagram against simple well formedness rules. If validation succeeds then the diagram is recognised. The second one uses visual parsing techniques and is executed on the well formed diagram only if more checks and/or a syntactic interpretation are needed for further translation or execution of the diagram.

The rest of the paper is organised as follows: the next section contains a brief survey on frameworks for sketch recognition; in section 3 we describe the framework, its architecture and the main recognition techniques; section 4 gives the data to provide in order to instantiate the framework for a particular domain; lastly, some final remarks and a brief discussion on future work conclude the paper.

## 2 Related Work

In the literature of sketch recognition we can find the description of many solutions, both multi-domain or oriented to the interpretation of diagrams from specific domains (e.g., UML diagrams [3], electrical circuits [4], chemical drawings [5], etc.). In this brief survey we will only focus on the former frameworks. We will also briefly outline other proposals which only face specific subproblems (e.g., fragmentation of strokes, identification of primitives, recognition of symbols), since some of these techniques are used in our framework.

### 2.1 Multi-Domain Sketch Recognition

Most approaches exploit the domain knowledge to improve recognition at a lower level. *SketchREAD* [2] is a multi-domain sketch recognition framework which uses a structural description of the domain symbols to perform the recognition. The domain knowledge is also used in the low-level phases, in order to allow the system to recover from low-level recognition errors. *SketchREAD* was evaluated in two different domains: family trees and circuit diagrams.

*AgentSketch* [4] is a multi-domain sketch recognition system in which an agent-based system is used for interpreting sketched symbols. The method exploits the knowledge

about the domain context for disambiguating the symbols recognized at a lower level.

The framework presented in [26] exploits a combination of low level and high-level techniques to be less sensitive to noise and drawing variations. It has been evaluated on two domains: molecular diagrams and electrical circuits.

*InkKit* [5] is a sketch tool framework which works very similarly to ours. It firstly classifies the strokes as either writing or drawing, then identifies basic shapes such as lines, rectangles, and circles, then groups these primitives in text and diagram components and, lastly, identifies the relationships between components.

*LADDER* is a language [18] which enables the definition of sketched elements at different levels (e.g., primitives, symbols, entire diagrams, etc.) by giving a structural description of them, including components, geometric constraints, etc. The framework can automatically generate a domain specific sketch recognition system from each description and has been tested on many domains including UML diagrams, mechanical engineering, flowcharts and others.

Other frameworks working at a lower level than those cited above, but having possible application on a broad range of domains are *Paleosketch* [28] and *CALI* [14]. The former is a recognition and beautification framework that can recognize different classes of primitive shapes and combinations of them. The latter exploits a naive Bayesian classifier to recognize multi-stroke geometric shapes.

### 2.2 Low-Level Techniques for Sketch Recognition and Symbol Recognition

Some frameworks are aimed to the solution of specific subproblems of sketch recognition. In recent years we have seen notable improvements in low-level stroke processing and symbol recognition techniques. The most effective of them are machine learning-based and are aimed to: stroke fragmentation [35, 36, 27, 20, 19, 1] and recognition of unistroke [24, 29] and multi-stroke [31, 25, 21, 23, 16, 26, 33] symbols.

Stroke fragmentation is a very mature subfield of research in sketch recognition. It has produced interesting results in recent times, especially through the use of machine learning techniques. Its objective is the recognition of the graphical primitives composing the strokes and can be used for a variety of objectives, including structural symbol recognition [21, 12]. Most approaches break strokes in corners [35, 36, 27, 20], while some other approaches [19, 1] also use the so called *tangent vertices* (smooth points separating a straight line from a curve or parting two curves). Machine learning-based approaches are based on the extraction of some features from the points of stroke, particularly speed and curvature.

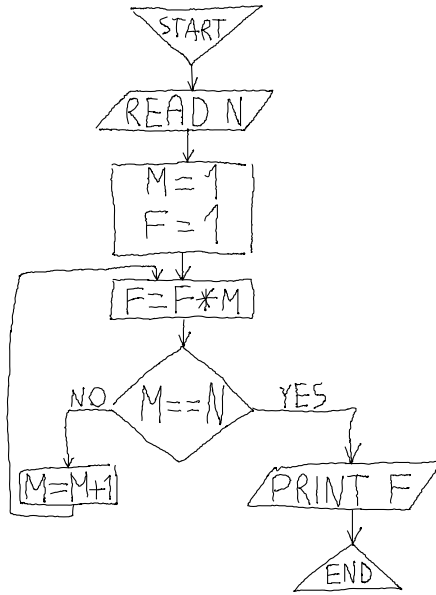


Figure 1: A simple flowchart for computing  $N!$ .

As for symbol recognition, the earliest methods [24, 29] were only able to recognize unistroke symbols. Several specialized methods have been recently proposed for multi-stroke hand-drawn symbol recognition. According to a widely accepted taxonomy [38] the methods are classified in two main categories: *structural* and *statistical*. In *structural* methods, the matching is performed by finding a correspondence between the structures, such as graphs [31, 25, 21] or trees [23], representing the input and the template symbols. In *Statistical* methods [16, 26, 33], instead, a given number of features are extracted from the pixels of the unknown symbol and compared to those of the models.

The recognition of unistroke symbols has had a recent progress and has been especially used in the recognition of gestures [34, 22, 9] for various applications, including interfaces for mobile devices [37, 15].

### 3 The Framework Design

The objective of the framework is to enable the recognition of diagrams from a wide range of domains. A common feature of these domains is the presence of three different types of graphics: symbols, connectors and text.

The framework has a layered architecture composed of the following four Layers, further divided into modules: *Text/Graphic Separation Layer*; *Symbol Recognition Layer*; *Local Context Detection Layer*; *Diagram Recognition/Parsing Layer*.

In the following we will refer to the flowchart in Figure 1 to exemplify the operations performed by the different layers.

#### 3.1 The Text/Graphic Separation Layer

The scope of this layer is to separate freehand drawing (graphics) from handwriting (text) (see Figure 2b). This preliminary operation is necessary because text and graphics need to be managed separately. For the realization of this layer we relied on the technique presented in [3]. For sake of conciseness, in the rest of the paper we focus on the graphic aspects of the diagram.

#### 3.2 The Symbol Recognition Layer

This layer recognizes the user drawn sketched symbols. It is further divided in the following modules:

- **Stroke Preprocessing Module.** This module identifies the graphical primitives present in the graphic domain of the diagram (see Figure 2c). This is done in two distinct phases: a *segmentation* phase, in which a stroke is divided in more primitives and a *clustering* phase in which different segments are put together to form a primitive. Segmentation is performed by detecting corners through *RankFrag* [6], a novel technique derived from previous machine learning-based methods [27, 20].
- **Symbol Identification Module.** This module clusters the primitives identified at the previous step in two different classes: *symbols*, *connectors* (see Figure 2d). For the realization of this module we relied on the technique, based on machine learning, described in [30].
- **Symbol Recognition Module.** Once the primitives composing a symbol have been grouped together, the symbol must be assigned to a class of known symbols (see Figure 2e). The recognition task is performed by this module by using the technique described in [8], which is a point cloud technique invariant with respect to scaling and supports symbol recognition independently from the number and order of strokes.

#### 3.3 The Local Context Detection Layer

In the last years many methods to model a diagram as a member of a visual language have been devised by researchers. Basically, a diagram has been represented either as a set of relations on symbols (the *relation-based approach*) or as a set of attributed symbols with typed attributes representing the “position” of the symbol in the sentence (the *attribute-based approach*) [11]. Even though the two approaches appear to be very different, they both model a diagram (or visual sentence) as a set of symbols and relations among them. Differently from the relation-based

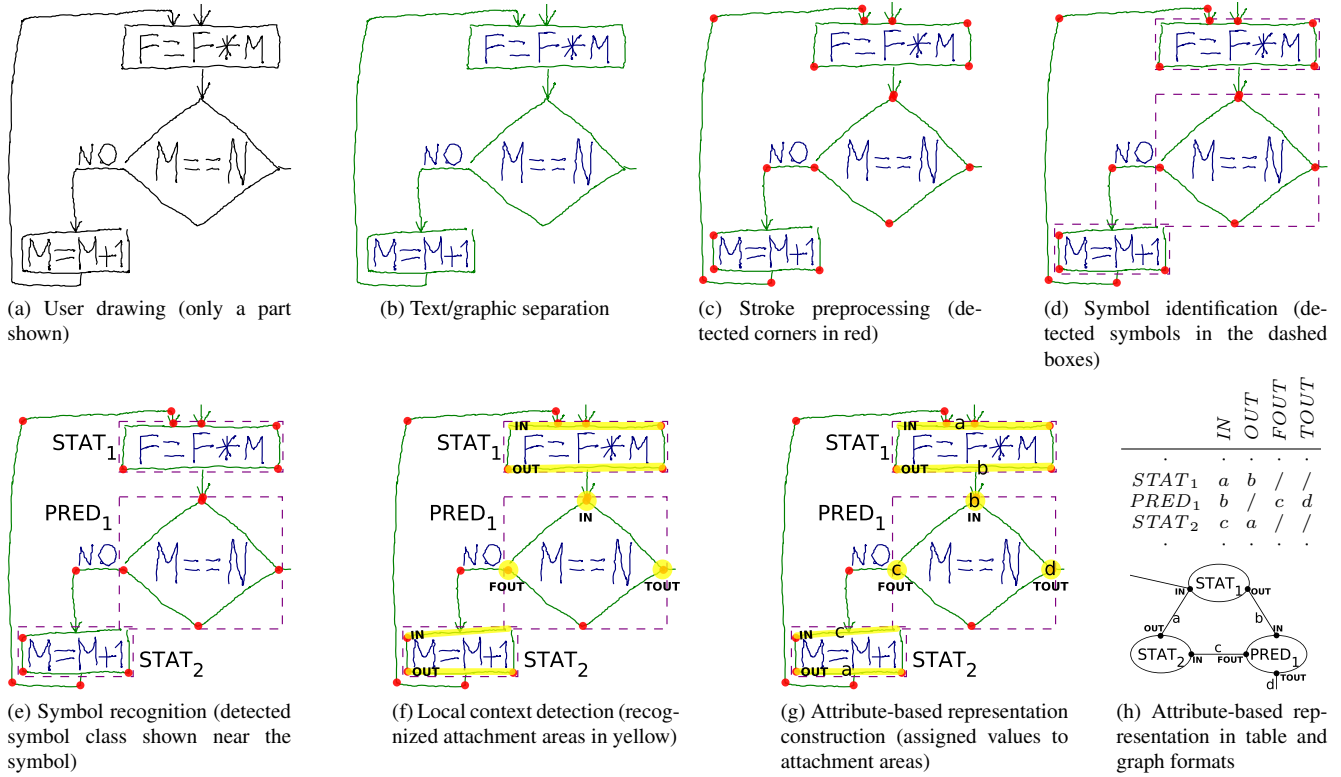


Figure 2: Recognition steps for a part of the flowchart in Figure 1.

approach where the relations are explicit, in the attribute-based approach the relations are implicit and must be derived by associating attribute values. The former approach is therefore at a higher level with respect to the latter. In this paper we adopt the attribute-based approach since it allows us to work at the lowest level possible. Moreover, we define the local context of a symbol as the set of its typed attributes. These are specified together with the visual characteristics of a symbol at definition time and are derived at recognition time from the way the symbol interacts with other symbols in a diagram. As an example, the rhomboid symbol in Figure 1 is defined in Table 1 with three attributes typed as *IN* (input), *FOUT* (output-if-false) and *TOUT* (output-if-true), visually corresponding to the three attaching points of the symbol. The first column of Table 1 shows the definition of flowchart symbol identifiers together with their typed attributes while the second column gives information about the symbol visual aspect and the location of its attributes. (The remaining columns of the Table will be described in the next section). The goal of the Local Context Detection Layer is then to identify the attributes and their types for each symbol (see Figure 2f). In our system, the attributes are identified by using an approach similar to the one proposed in [7]. The approach is independent from the method used to recognize symbols and assumes that the

Token	Graphics	Token occurrences	Constraints
BEGIN: <i>OUT</i>		1	$ OUT  = 1$
END: <i>IN</i>		1	$ IN  \geq 1$
STAT: <i>IN</i> , <i>OUT</i>		$\geq 0$	$ IN  \geq 1$ , $ OUT  = 1$
IO: <i>IN</i> , <i>OUT</i>		$\geq 0$	$ IN  \geq 1$ , $ OUT  = 1$
PRED: <i>IN</i> , <i>FOUT</i> , <i>TOUT</i>		$\geq 0$	$ IN  \geq 1$ , $ FOUT  = 1$ , $ TOUT  = 1$

Table 1: Flowchart symbol specifications.

symbol has already been recognized.

### 3.4 The Diagram Recognition/Parsing Layer

This layer is composed of two modules: the Local Context-based Recogniser and the Diagram Parser. They

execute sequentially and in some uses of the framework, when the language is fully specified by the local constraints or in the case of fast prototyping, only the first module is needed.

### 3.4.1 Local Context-based Recogniser

The output of the recogniser is the attribute-based representation of the diagram if the diagram is well formed. This representation is constructed by giving values to the symbol attributes identified in the previous layer while checking for the well formedness of the diagram against simple constraints. In order to give values to attributes, the module generates value ids and assigns them such that two attributes have the same value only if the corresponding graphical counterparts are connected (see Figure 2g). This will then produce the attribute-based representation of the diagram (shown in Figure 2h both in tabular and graph formats). As for well formedness, constraint rules are given at definition time together with the symbol specifications and are intended to be easy to verify. In our example, the third and fourth columns of Table 1 indicate how many times a symbol may occur in a diagram and the number of values that an attribute may have, respectively. The symbol PRED may then appear zero or more times in a diagram and may have multiple input connections but only one exiting connection for each output attaching point. In our experience, simple constraints such as the ones above are enough to completely describe a visual language. In the case of the flowcharts as the one depicted in Figure 1, it is easy to verify that Table 1 together with the three rules “each connection must be 1-to-1”, “each IN attaching point must only be connected to an (F/T)OUT attaching point” and “the attribute-based (graph) representation must be connected” completely specify a set of flowcharts that is Turing-complete.

The Local Context-based Recogniser takes as input an XML specification file where all the rules about tokens and connections are coded. Figure 3 shows the XML file specification for the case described above coding the information in Table 1 and the three additional rules. In the XML specification, each table row is defined through a `token` element. The name of the token, the file containing its graphical representation, and the number of its occurrences in a language instance are defined through the `name`, `ref` and `occurrences` attributes, respectively. Each `ap` element defines one of the token attaching points by specifying its type (`type` attribute), name (`name` attribute), a reference to its position in the graphical representation (`ref` attribute) and the number of allowed connection (`connectNum` attribute). The constraint “the attribute-based (graph) representation must be connected” is specified by `<constraint>connected</constraint>`.

```
<language name="flowchart">
  <token name="begin" ref="triangleDown.svg"
    occurrences="==1">
    <ap type="exit" name="out" ref="lowPoint"
      connectNum="==1" />
    </token>
  <token name="end" ref="triangleUp.svg"
    occurrences="==1">
    <ap type="enter" name="in" ref="hiPoint"
      connectNum="==1" />
    </token>
  <token name="stat" ref="rectangle.svg"
    occurrences=">=0">
    <ap type="enter" name="in" ref="hiLine"
      connectNum=">=1" />
    <ap type="exit" name="out" ref="lowLine"
      connectNum="==1" />
    </token>
  <token name="io" ref="parallelogram.svg"
    occurrences=">=0">
    <ap type="enter" name="in" ref="hiLine"
      connectNum=">=1" />
    <ap type="exit" name="out" ref="lowLine"
      connectNum="==1" />
    </token>
  <token name="pred" ref="rhombus.svg"
    occurrences=">=0">
    <ap type="enter" name="in" ref="hiPoint"
      connectNum=">=1" />
    <ap type="exit" name="fout" ref="leftPoint"
      connectNum="==1" />
    <ap type="exit" name="tout" ref="rightPoint"
      connectNum="==1" />
    </token>
  <connector ref="arrow">
    <cap type="enter" ref="head" connectNum="==1" />
    <cap type="exit" ref="tail" connectNum="==1" />
  </connector>
  <constraint>connected</constraint>
</language>
```

Figure 3: Flowchart specification.

The `connector` element describes how tokens are connected. It defines its type (from a predefined list of implemented types) and specifies that the head of the arrow must be connected to an *enter* attaching point, while the head must be connected to an *exit* attaching point. The predefined type *arrow* together with the two `connectNum="==1"` conditions guarantee that the property “each connection must be 1-to-1” is satisfied, while the use of the type definitions *enter* and *exit* in the token elements guarantees that the property “each IN attaching point must only be connected to an F/T)OUT attaching point” is satisfied.

As a second example of application, let us now consider the language of the binary trees. In this case, the symbol specification shown in Table 2 and the three constraints “each connection must be 1-to-1”, “each IN attaching point

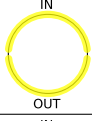
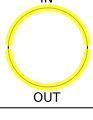
Token	Graphics	Token occurrences	Constraints
ROOT: <i>IN</i> , <i>OUT</i>		1	$ IN  = 0,$ $ OUT  \leq 2$
NODE: <i>IN</i> , <i>OUT</i>		$\geq 0$	$ IN  = 1,$ $ OUT  \leq 2$

Table 2: Binary tree symbol specifications.

```

<language name="binaryTree">
  <token name="root" ref="circle.svg" occurrences
    ==1">
    <ap type="enter" name="in" ref="hiSC"
      connectNum=="=0" />
    <ap type="exit" name="out" ref="lowSC"
      connectNum="<=2" />
  </token>
  <token name="node" ref="circle.svg" occurrences
    ">=0">
    <ap type="enter" name="in" ref="hiSC"
      connectNum=="=1" />
    <ap type="exit" name="out" ref="lowSC"
      connectNum="<=2" />
  </token>
  <connector ref="polyline">
    <cap type="exit" ref="p0" connectNum=="=1" />
    <cap type="enter" ref="p1" connectNum=="=1" /
  >
  </connector>
  <constraint>connected</constraint>
</language>

```

Figure 4: Binary tree specification.

must only be connected to an *OUT* attaching point” and “the attribute-based (graph) representation must be connected”, as coded in the XML specification shown in Figure 4, completely describe the language. Here, ROOT and NODE have the same graphical representation and are only distinguished for the number of occurrences and the constraints on the *IN* attaching point.

### 3.4.2 Diagram Parser

This parser is built only if a syntactic interpretation of the diagram is needed for further processing, such as translation or execution, and/or if the language cannot be completely specified by a set of simple constraints. This is similar to the division of roles between the lexical and syntactic phases for a traditional compiler. The diagram parser takes as input the attribute-based representation produced in the previous module and a visual grammar for the syntax specification. Many visual grammar formalisms and corresponding pars-

ing algorithms may be found in literature and, even though we adopt a parsing technique based on the principles described in [10], the framework is not linked to any specific type of visual parser technology. Moreover, it is important to note that, since the input to the parser is already well formed, the complexity of this module is simplified with respect to other approaches.

In order to show a case when the local context recognition needs to be complemented by a syntax analysis phase let us consider a structured version of the flowchart language described in the previous section (see Figure 5). To structure the language we introduce two more tokens with names B\_BEGIN and B\_END whose roles are the same as the block delimiters “{” and “}” in the C language, respectively. Each of the two tokens is specified similarly to the token STAT with number of occurrences  $\geq 0$  and two attaching points *IN* and *OUT* with types *enter* and *exit*, respectively, and  $|IN| \geq 1$  and  $|OUT| = 1$ . As in any structured language, the block delimiters B\_BEGIN and B\_END are to be used in pairs and then other rules should be added. As already known, these are not constraints that can be solved by locally looking at the properties of a single token. As a consequence, the technique described in the previous section cannot be used to capture the whole structured flowchart language. We now provide a visual grammar describing a limited structured flowchart language including the flowchart in Figure 5. The grammar is composed by a set of terminals given by the tokens as described in Figure 3 in the format: TOKEN(*attaching\_point1*, *attaching\_point2*, ...), a set of non terminals in the same token format: Nterm(*attaching\_point1*, *attaching\_point2*, ...), an initial terminal FlowCh, and the set of productions shown in Figure 6. In each production, the single letters x, y, u, ... represent, when in the right part of the production, connections between token and/or non terminal attaching points. When in the left part of a production, they indicate which attaching points of the subsentence are externally exposed. The notation  $x \oplus z$  indicates that the two attaching points marked by x and z will be connected to the same target attaching point. As an example, the subsentence in Figure 2g matches and instantiates production 7 as follows: Block(a, d)  $\rightarrow$  Block(a,b) PRED(b, c, d) Block(c, a) where Block(a,b) comes from matching STAT<sub>1</sub> with production 5 instantiated as Block(a, b)  $\rightarrow$  STAT(a, b) and Block(c, a) comes from matching STAT<sub>2</sub> with production 5 instantiated as Block(c, a)  $\rightarrow$  STAT(c, a). As already pointed out, in the literature there are many approaches that use visual grammar formalisms, at least as powerful as the one above, to generate visual parsers directly from a grammar.

It can be noted that, without a local context analysis, syntax errors such as adding an extra connection between any token in Figure 5, cannot be easily detected by only using a grammar approach.



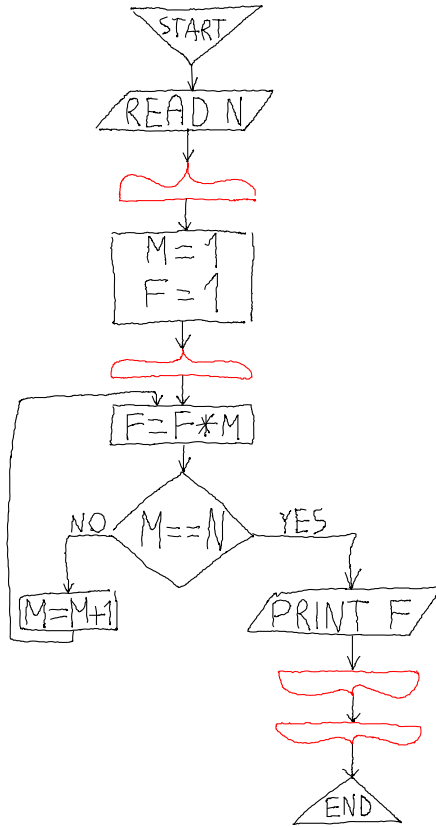


Figure 5: A simple structured flowchart containing the B\_BLOCK and B\_END tokens.

## 4 Instantiating the Framework

In order to instantiate the framework for a specific domain, it is necessary to provide input data. In particular, being mostly based on machine learning, the Symbol Recognition Layer modules need a training phase, while the higher level layers need formal definitions. The following data must be provided:

1. Sample diagrams from the domain, with the following annotations:
  - Types of strokes (text, graphics);
  - Strokes clustering/segmentation;
  - Connector/symbols separation;
  - Class names of the symbols;
  - Attachment areas on the symbols;
2. The XML specification of the language and the referenced files with the graphical definitions of the symbols and attaching points;
3. The specification for the syntax interpretation of a diagram (optional).

1.  $\text{FlowCh} \rightarrow \text{BEGIN}(x) \text{Block}(x, y) \text{END}(y)$
2.  $\text{Block}(x, y) \rightarrow \text{Block}(x, z) \text{Block}(z, y)$
3.  $\text{Block}(x, y) \rightarrow \text{B\_BEGIN}(x, z) \text{Block}(z, u) \text{B\_END}(u, y)$
4.  $\text{Block}(x, y) \rightarrow \text{IO}(x, y)$
5.  $\text{Block}(x, y) \rightarrow \text{STAT}(x, y)$
6.  $\text{Block}(x, y \oplus z) \rightarrow \text{PRED}(x, u, v) \text{Block}(u, y) \text{Block}(v, z)$
7.  $\text{Block}(x \oplus z, y) \rightarrow \text{Block}(x, u) \text{PRED}(u, v, y) \text{Block}(v, z)$
8.  $\text{Block}(x \oplus z, y) \rightarrow \text{PRED}(x, u, y) \text{Block}(u, z)$

Figure 6: Visual grammar describing a structured flowchart language.

## 4.1 Implementation

The framework is being developed in Java.

In addition to the development of the modules for the recognition, we are working at the development of an environment that facilitates the production of input data needed to instantiate the framework for a particular domain. Specifically, we're providing a GUI for making quick annotations on the sample diagrams and to define the constraints and the syntax specification.

## 5 Conclusions

We described a local context-based recognition methodology whose final objective is the development of a framework for multi-domain sketch recognition and interpretation. The diagrams may contain different types of graphic elements (symbols, connectors, text). Future work will include the instantiation of the framework in different domains. At the end of the development phase, we will perform tests to evaluate the effectiveness and efficiency of the individual modules and of the overall framework. We will make comparative evaluations with state-of-art techniques.

## References

- [1] F. Albert, D. Fernandez-Pacheco, and N. Aleixos. New method to find corner and tangent vertices in sketches using parametric cubic curves approximation. *Pattern Recognition*, 46(5):1433 – 1448, 2013.
- [2] C. Alvarado and R. Davis. Sketchread: a multi-domain sketch recognition engine. In *Proc. of UIST '04*, pages 23–32, 2004.
- [3] D. Avola, A. Buono, P. Nostro, and R. Wang. A novel online textual/graphical domain separation approach for sketch-based interfaces. In E. Damiani, J. Jeong, R. Howlett, and L. Jain, editors, *New Directions in Intelligent Interactive Multimedia Systems and Services - 2*, volume 226, pages 167–176. Springer, 2009.

- [4] G. Casella, V. Deufemia, V. Mascardi, G. o Costagliola, and M. Martelli. An agent-based framework for sketched symbol interpretation. *JVLC*, 19(2):225–257, 2008.
- [5] R. Chung, P. Mirica, and B. Plimmer. Inkkit: A generic design tool for the tablet pc. In *Proc. of CHINZ'05*, pages 29–30, 2005.
- [6] G. Costagliola, M. De Rosa, V. Fortino, and V. Fuccella. Rankfrag: A novel machine learning-based technique for hand-drawn sketch segmentation. Submitted for publication, Apr. 2014.
- [7] G. Costagliola, M. De Rosa, and V. Fuccella. Identifying attachment areas on sketched symbols. In *Proc. of VL/HCC '11*, pages 83–86, 2011.
- [8] G. Costagliola, M. De Rosa, and V. Fuccella. Improving shape context matching for the recognition of sketched symbols. In *Proc. of DMS*, pages 289–294, 2011.
- [9] G. Costagliola, M. De Rosa, and V. Fuccella. Investigating human performance in hand-drawn symbol autocompletion. In *Proc. of SMC '13*, pages 279–284, 2013.
- [10] G. Costagliola, V. Deufemia, and M. Risi. Sketch grammars: a formalism for describing and recognizing diagrammatic sketch languages. In *Proc. of ICDAR'05*, pages 1226–1230, 2005.
- [11] G. Costagliola and G. Polese. Extended positional grammars. In *Proc. of VL '00*, pages 103–110, 2000.
- [12] G. Costagliola, M. D. Rosa, and V. Fuccella. Recognition and autocompletion of partially drawn symbols by using polar histograms as spatial relation descriptors. *Computers & Graphics*, 39(0):101 – 116, 2014.
- [13] G. Feng, C. Viard-Gaudin, and Z. Sun. On-line hand-drawn electric circuit diagram recognition using 2d dynamic programming. *Pattern Recognition*, 42(12):3215 – 3223, 2009.
- [14] M. Fonseca and J. Jorge. Using fuzzy logic to recognize geometric shapes interactively. In *Proc. of FUZZ'IEEE*, volume 1, pages 291–296 vol.1, 2000.
- [15] V. Fuccella, M. De Rosa, and G. Costagliola. Novice and expert performance of keyscratch: a gesture-based text entry method for touch-screens (in press). *IEEE Transactions on Human-Machine Systems*, 2014.
- [16] L. Gennari, L. B. Kara, T. F. Stahovich, and K. Shimada. Combining geometry and domain knowledge to interpret hand-drawn diagrams. *Computers & Graphics*, 29(4):547–562, 2005.
- [17] T. Hammond and R. Davis. Tahuti: a geometrical sketch recognition system for uml class diagrams. In *ACM SIGGRAPH 2006 courses*, 2006.
- [18] T. Hammond and R. Davis. Ladder, a sketching language for user interface developers. In *ACM SIGGRAPH 2007 courses*, 2007.
- [19] J. Herold and T. F. Stahovich. Speedseg: A technique for segmenting pen strokes using pen speed. *Computers & Graphics*, 35(2):250–264, 2011.
- [20] J. Herold and T. F. Stahovich. A machine learning approach to automatic stroke segmentation. *Computers & Graphics*, 38(0):357 – 364, 2014.
- [21] W. Lee, L. Burak Kara, and T. F. Stahovich. An efficient graph-based recognizer for hand-drawn symbols. *Computers & Graphics*, 31:554–567, August 2007.
- [22] Y. Li. Protractor: A fast and accurate gesture recognizer. In *Proc. of CHI '10*, pages 2169–2172, 2010.
- [23] Y. Lin, L. Wenyin, and C. Jiang. A structural approach to recognizing incomplete graphic objects. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 1, pages 371–375 Vol.1, aug. 2004.
- [24] J. S. Lipscomb. A trainable gesture recognizer. *Pattern Recognition*, 24:895–907, September 1991.
- [25] J. Lladós, E. Martí, and J. Villanueva. Symbol recognition by error-tolerant subgraph matching between region adjacency graphs. *IEEE Trans. PAMI*, 23(10):1137–1143, Oct. 2001.
- [26] T. Y. Ouyang and R. Davis. A visual approach to sketched symbol recognition. In *Proc. of IJCAI'09*, pages 1463–1468, 2009.
- [27] T. Y. Ouyang and R. Davis. Chemink: a natural real-time recognition system for chemical drawings. In *Proc. of IUI '11*, pages 267–276, 2011.
- [28] B. Paulson and T. Hammond. Paleosketch: accurate primitive sketch recognition and beautification. In *Proc. of IUI '08*, pages 1–10, 2008.
- [29] D. Rubine. Specifying gestures by example. *SIGGRAPH Comput. Graph.*, 25:329–337, July 1991.
- [30] T. F. Stahovich, E. J. Peterson, and H. Lin. An efficient, classification-based approach for grouping pen strokes into objects. *Computers & Graphics*, (0):–, 2014.
- [31] W.-H. Tsai and K.-S. Fu. Error-correcting isomorphisms of attributed relational graphs for pattern analysis. *IEEE Trans. Systems Man Cyber.*, 9(12):757–768, dec. 1979.
- [32] D. G. Ullman, S. Wood, and D. Craig. The importance of drawing in the mechanical design process. *Computers & Graphics*, 14(2):263 – 274, 1990.
- [33] D. Willems, R. Niels, M. van Gerven, and L. Vuurpijl. Iconic and multi-stroke gesture recognition. *Pattern Recognition*, 42(12):3303–3312, 2009. New Frontiers in Handwriting Recognition.
- [34] J. O. Wobbrock, A. D. Wilson, and Y. Li. Gestures without libraries, toolkits or training: a \$1 recognizer for user interface prototypes. In *Proc. of UIST '07*, pages 159–168, 2007.
- [35] A. Wolin, B. Eoff, and T. Hammond. Shortstraw: A simple and effective corner finder for polylines. In *EUROGRAPHICS Workshop on Sketch-Based Interfaces and Modeling*. Eurographics Association, 2008.
- [36] Y. Xiong and J. J. J. LaViola. A shortstraw-based algorithm for corner finding in sketch-based interfaces. *Computers & Graphics*, 34(5):513 – 527, 2010.
- [37] S. Zhai and P. O. Kristensson. The word-gesture keyboard: Reimagining keyboard interaction. *Commun. ACM*, 55(9):91–101, Sept. 2012.
- [38] W. Zhang, L. Wenyin, and K. Zhang. Symbol recognition with kernel density matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):2020–2024, dec. 2006.

# Towards a Lightweight Approach for Modding Serious Educational Games: Assisting Novice Designers

Jacob Dahleen, Alex Hunsberger,  
Ryan Weber, Dennis Brylow  
Department of Mathematics, Statistics  
and Computer Science  
Marquette University  
Milwaukee, U.S.A.

C. Shaun Longstreet  
Center for Teaching and Learning  
Marquette University  
Milwaukee, U.S.A.

Kendra M.L. Cooper  
Department of Computer Science  
The University of Texas at Dallas  
Richardson, U.S.A.

**Abstract**— Serious educational games (SEGs) are a growing segment of the education community’s pedagogical toolbox. Effectively creating such games remains challenging, as teachers and industry trainers are content experts; typically they are not game designers with the theoretical knowledge and practical experience needed to create a quality SEG. Here, a lightweight approach to interactively explore and modify existing SEGs is introduced, a tool that can be broadly adopted by educators for pedagogically sound SEGs. Novice game designers can rapidly explore the educational and traditional elements of a game, with a stress on tracking the SEG learning objectives, as well as allowing for reviewing and altering a variety of graphic and audio game elements.

**Keywords**—*Serious Educational Games, Visualization Environment, Novice Game Designers*

## I. INTRODUCTION

Serious educational games (SEGs) are a growing segment of the education community’s pedagogical toolbox. There is increasing evidence for their efficacy in sustaining engaged learning [11][12][13][22]. SEGs allow for rapid feedback for individually customized educational experiences. Moreover, digital engagement is increasingly ubiquitous, with individuals of all ages growing adept at learning complex tasks online without the use of texts or self-contained tutorials. Nevertheless, the time, financial, and knowledge resources required to develop quality games are out of reach of most pedagogical stakeholders. This is all the more so for those who require a rapid role-out of games with up-to-date, just-in-time information for multiple groups, different grades, and varied subject matter.

The development of a semi-automated game development platform is important because the vast majority of educators do not have the knowledge or time to code their own educational gaming experiences. Semi-automation can assist with two critical functions. First, it can greatly reduce the complexity of game generation, leaving much of the educational content elements to the instructor. Teachers and industry trainers are content experts; typically they are not game designers with the theoretical knowledge and practical experience needed to

create a quality SEG. Without game design experience, instructors are left to storyboard an SEG effectively, determine variants for levels of difficulty, identify opportunities for challenges and then establish how results of the challenges affect subsequent game play.

Second, semi-automated game development situates the learning experience for both instructor and students in a way that draws upon an established body of knowledge and structured modes of learning, thereby insuring an SEG experience that is directly tied to selected learning objectives. This latter part is where many SEG programs can fall short for educators. Instructors would need to outline, embed, and track learning objectives for each specific player choice and challenges for individual games. The quality of the SEG game play experience becomes overly dependent on instructors’ abilities to develop a quality game and too often puts the burden of game immersion on the player.

In this paper, we introduce a lightweight modding approach, SEG<sub>Mod</sub>, that can be broadly adopted by novice designers of SEGs. The approach is part of a larger, on-going project, SimSYS, which explores SEG research issues with a development platform for serious educational games, including Game Generation. To keep this proposed approach straightforward, the educational game designers are provided with a relatively modest set of functionality. They can load the game, browse the structure of the game, browse the learning outcomes of a game (from the top down to challenge levels of game play), modify visual and audio game elements, check for errors that may have been introduced, and save their tailored games. Our work is innovative in that the learning outcomes are consistently and explicitly present throughout the game design process and explicitly embedded in game play.

The SEG<sub>Mod</sub> approach has been investigated using a scenario based approach. First, a set of high level scenarios are defined. For example:

*Scenario 1. Amy, a 4<sup>th</sup> grade teacher, has just finished generating a game to teach her students math. She wants to review the learning objectives and confirm that*

the game has the appropriate assets. Amy opens the game with the Interactive Environment and clicks through each of the screens to insure consistency with the prior week's lessons. As she clicks on the first screen of the opening scene in Act Two, she notices that the protagonist character is too small and out of place. She moves and resizes the character to a suitable location and saves the game.

A set of scenarios such as the one above is used to drive the requirements for the approach and the interface design. Our approach has been validated by developing a prototype tool. The design of the tool has two components and it has been programmed using Java, open source tools, and standard libraries. Substantial testing (feature, performance) has been conducted as preliminary validation, demonstrating that the tool supports the outlined scenarios, thereby providing strong indicators the SEG<sub>Mod</sub> is a promising research direction.

The structure of the remainder of this paper is as follows. Section II presents background on the broader SimSYS project, to provide the context for this research. The proposed SEG<sub>Mod</sub> approach and the validation are presented in Section III and Section IV, respectively. Related Work is in Section V. Conclusions and Future Work are in Section VI.

## II. SIMSYS PROJECT BACKGROUND

SimSYS is an on-going project investigating the effective semi-automated development of SEGs [18]. This is an interdisciplinary project, which integrates three research areas: traditional game design, pedagogical methodologies, and software engineering methodologies. Here, we provide an overview of the SimSYS Game Development Platform (GDP) to provide the context of the new results reported in this paper.

### A. SimSYS Game Development Platform

The SimSYS GDP is being systematically engineered so that it is well-modularized, extensible, and established on a meta-model foundation [4][7][18]. The overall architecture, illustrated in Fig. 1, is based on the repository style, which consists of a repository that stores knowledge with controlled access, and a collection of knowledge sources that directly interact with the repository (i.e., clients of the repository). This organization has been adopted to support the rapid evolution of content; we recognize there may be redundant storage of assets that increases the storage overhead of this approach.

Here, the repository is organized into domain specific and domain independent sub-repositories. The domain specific sub-repository stores game models (scripts), game play data, and re-usable game assets such as collections of domain specific Challenges (e.g., quiz questions/answers/feedback about software engineering, mathematics, physics, and so on) that are used in the games. The challenge content also captures relationships to educational standards (e.g., SWEBOK [15], [26]), Taxonomy of Learning standards (e.g., Bloom's [1]) and the level of difficulty (defined by educators). Re-usable domain independent assets include the formal SimSYS game specification[17], defined in XSD [27], and a collection of game assets (e.g., sound effects, graphic images). A SimSYS game is organized into Acts, with learning objectives coupled

to them. Each Act contains lower-level Scenes, Screens, and Challenges, marking advancing points in the game where

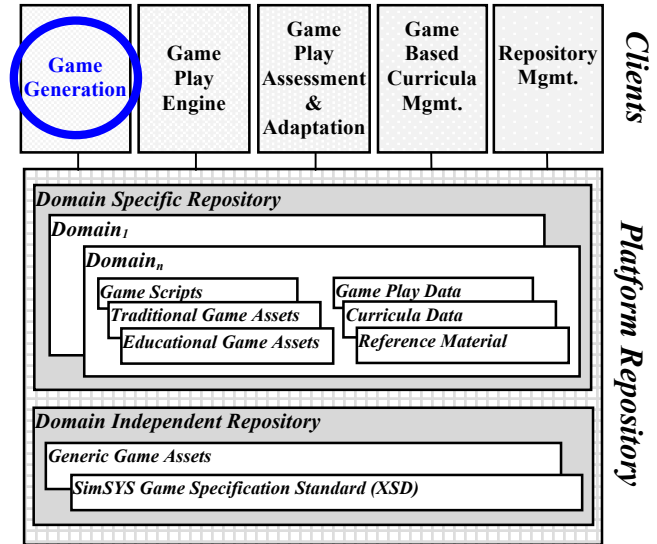


Fig. 1. SimSYS Game Development Platform

learning objectives and game narrative play out. SimSYS games are represented in XML, as Game Scripts.

The knowledge sources for the platform are modules that interact indirectly via changes to the repository: Game Generation; Game Play Engine; Game Play Assessment and Adaptation; Game Based Curricula Management; and Repository Management. These modules embody knowledge of related standards, including body of knowledge, accreditation, and certification. The stakeholder interacting with a module can work with the content in the repository from a particular view.

The Game Generation module is being explored as a collection of three approaches: 1) a highly automated game generator [7]; 2) a semi-automated game generator (interactive wizard to create and modify a game, guided by an experienced game designer), and 3) a preview visualization tool to provide a lightweight, easy to use tool support for exploring an existing game and making relatively minor modifications to it. The preview tool is a particularly appealing prospect because computer games are an inherently visual medium, and because a textual representation of the complex information encoded in an SEG would likely overwhelm all but the most dedicated novices in a design environment. Introducing the interactive, preview visualization approach, hereafter referred to as the SEG<sub>Mod</sub>, will be the focus for the rest of this paper.

## III. SEG<sub>Mod</sub>: LIGHTWEIGHT INTERACTIVE MODDING

The purpose of the SEG<sub>Mod</sub> approach is to support designers who only need to review, affect small changes to an existing game, and to make simple edits. Games can be created using, for example, an interactive semi-automated game generation wizard; individual instructors can use our preview tool to contextualize and frame specific SEG scenarios by editing text in speech bubbles, adding information boxes, or modifying just-in-time hints and prompts. They can play or switch out sounds and background music, swap characters, and even tailor the organization of challenge components. An instructor can

check and adjust relatively minor details within an SEG environment, or build new SEGs out of existing games stored in a file directory. For example, a fantasy-themed reading comprehension game with a castle backdrop for 6<sup>th</sup> grade language arts might not be appropriate for adult learners of English. In another situation, a community college instructor teaching a department's common U.S. History course might switch out game props that are contextually relevant to his specific class discussions. Likewise, there will be times when an instructor will want to add, delete, highlight, move, or resize game elements without having to build an entirely new game.

Experienced game designers can use the SEG<sub>Mod</sub> approach as part of the process for creating a game from scratch. These designers can use the interactive, semi-automated game generation wizard to create a game. The SEG<sub>Mod</sub> approach can be used to quickly visualize the overall structure and content of the game, in addition to making limited modifications to it, as the design progresses. The experienced designers can iteratively design, visualize, and make minor modifications to a game. The lightweight SEG<sub>Mod</sub> is envisioned to provide a complimentary alternative to loading and playing the game in the game play engine.

Error detection and a feature for automatically correcting some specific kinds of errors is part of the preview tool. For example, if one instructor drags an asset off-screen rather than deleting it and saves the game, subsequent users of that game would have difficulty addressing that asset, so the automated error correction would reset off-screen assets back onto the interactive display. Likewise, if the game engine did not find appropriate knowledge data associated with learning objectives or is unable to call items from the game asset repository, these errors would be flagged visually in an error report of the preview welcome display as well as on the relevant lower level screens in the interactive display.

The SEG<sub>Mod</sub> approach is not intended to be a full audio/visual SEG development solution, which would need to support major changes to learning objectives or game structure. For that, the SimSYS game development framework provides the game development wizard, wherein an instructor, lab coordinator, supervisor can input data, learning objectives, general assets and assessment parameters for the SEG. This approach differs from other researchers [8][20] in that alternative tools provide different levels of design engagement, which creates a more flexible game development platform. Moreover, the development of SEG<sub>Mod</sub> will allow for easier scalability and adaptations across institutions.

#### A. Scenarios

A collection of high level scenarios have been defined and refined requirements provide the scope of functionality in the lightweight approach. Several example scenarios are:

**Scenario 1.** We already have mentioned Amy, the 4<sup>th</sup> grade teacher above who uses the SEG<sub>Mod</sub> approach to move and resize characters to more suitable locations and saves the game (refer to Section I).

**Scenario 2.** George, a high school physics teacher, has generated a game to teach his students about kinematics and has opened the game with the SEG<sub>Mod</sub> approach to see what it

will look like. Previewing the game, he decides, based on previous weeks' experience, that his students would benefit from the addition of more instruction. He then adds a new character with a speech bubble to provide more information to his students. He also adds hints that remind his students what they have been working on in class that will help them in the game. Further on, he sees an opportunity in character conversations to add some humour to the game, and make the game more personal for his students. He finds the rest of the game acceptable and saves the game.

**Scenario 3.** Pat, a corporate human resources educator, will use a pre-developed game in her training program for the company's ethics reporting procedure. She wants to preview the game first so that she can make references to it in her face-to-face session with her staff later in the week. After loading the game in the SEG<sub>Mod</sub> approach she runs through the screens in the game. She observes how the game plays out for participants, noting particular scenarios that she thinks will be good for a longer discussion with her class. Furthermore, she sees an opportunity to update some content that will align the game with more recent developments in corporate relations. She adjusts some text, but then also changes out visual assets like logos from the asset repository as well as adjusting elements of game characters' ethnography to be more representative of the company. She edits some of the text and saves the adjusted game.

#### B. Requirements

The scenarios have been refined into requirements. The core functional requirements are capabilities to:

- Browse the organization of the game.
  - Act – the largest organizational unit of the game, in which the parameters of the game context, characters and learning objectives are all contained.
  - Scene – a collection of screens that establish the narrative arc and pedagogical goals for the player.
  - Screen – an individual page that contains information for the player and may or may not require player input.
  - Challenge – the exercise by which a player demonstrates proficiency towards a learning objective.
- Browse game learning objectives at any point of game organization, (i.e. Act, Scene, Screen, and Challenge).
- Preview the content of the game (Screen and Challenge).
- Add and edit the text of speech bubbles, hints, and feedback.
- Add, Move, Delete, and Resize game elements (characters, props) in the game (Screen).
- Change the backdrop (Scene).
- Toggle the display of hints and sound effects associated with visual elements (Screen and Challenge).
- Add, preview, and change sound effects (Screen).
- Add, preview, and change background audio (Scene).
- Save changes to the current or new game.
- Detect, report, and recommend corrections to errors in the game.

#### C. Interface Design

A screenshot of the interface design is provided in Fig. 2. As one part of the Game Generation module (Fig. 1), the preview and visualization capabilities are provided under the

**Preview tab** with an **Interactive Display** area, **Browser**, and **Toolbar**. A main **Menu** provides common features, such as file management (load, save, save as, and error checking).



Fig. 2 Interface Design: Menu, Browser, Toolbar, and Interactive Display areas.

Using the **Toolbar**, in the lower left of the interface, a designer can add a character, a prop (e.g., furniture, coffee cup), or a user interface button. He or she can choose to create an information box, hint, conversation bubble, or replace the backdrop for a Scene. The toolbar allows one to show/hide the icons that indicate a sound effect is present for a game element, the learning objectives associated with a Scene or Challenge, or any hints that are present. For example, clicking the “Character” button in the Toolbar opens the Character Select Window, where the user can browse through a collection of character options and add one to a Screen or Challenge.

To guide the designer, access to the toolbar buttons changes with the level of game currently being viewed. For example, in a Scene, the designer can edit the background. At the Screen level, however, changing the background is not an option, so that button is disabled and greyed out. All buttons offer a similar selection wizard with varying categories for each asset type.

In the **Interactive Display** area, on the right side of the interface, a designer can use the mouse to select a game element (character, props) and move or resize it; these same

selected elements can also be deleted using the delete key. A designer can edit the text of generic interaction elements, including speech bubbles, information boxes, and game hints. A designer can also add, delete, play sound effects. The browser is on the upper left of the interface. The designer can use the **Browser** to conveniently and quickly explore the game. The tree display allows the user to navigate between the Game, organized into Acts, Scenes, Screens, and Challenges. Each Act displays its title along with Learning Objectives that are included in the underlying Scenes/Screens. The challenge screen displays questions as the player would see them, as well as the view revealing hints and the intended learning objectives.

Using the main **Menu**, the designer can manage files (load, save, save as) and access the error checking feature. The game error checker runs following either loading a game, or clicking the “Check XML Errors” menu item. This feature checks the game for 63 kinds of errors and presents the results in the Interactive Display Panel. Each game error consists of a classification, a textual description, and its location in the game structure. Depending on its classification, an “Autocorrect” button may appear enabled to the right of the error text. Clicking this button saves the user from manually editing XML to make corrections.

#### IV. VALIDATION

The preliminary validation of the SEG<sub>Mod</sub> approach includes a prototype tool, which has undergone substantial feature and performance testing. Here, we provide a summary of the validation effort using Scenario 1 (Amy the 4<sup>th</sup> grade teacher, Section I). The results are presented in two parts, in order to simplify the presentation. The first part focuses on interactions to explore the learning objectives (Fig. 3); the second to adjust some of the play aesthetics (Fig. 4). Both parts begin after the designer opens a game and is shown a splash screen with a high level summary of any errors found checking it (Fig. 2); the designer is offered the “autocorrect” option for errors that can be handled by a default operation. These illustrations are

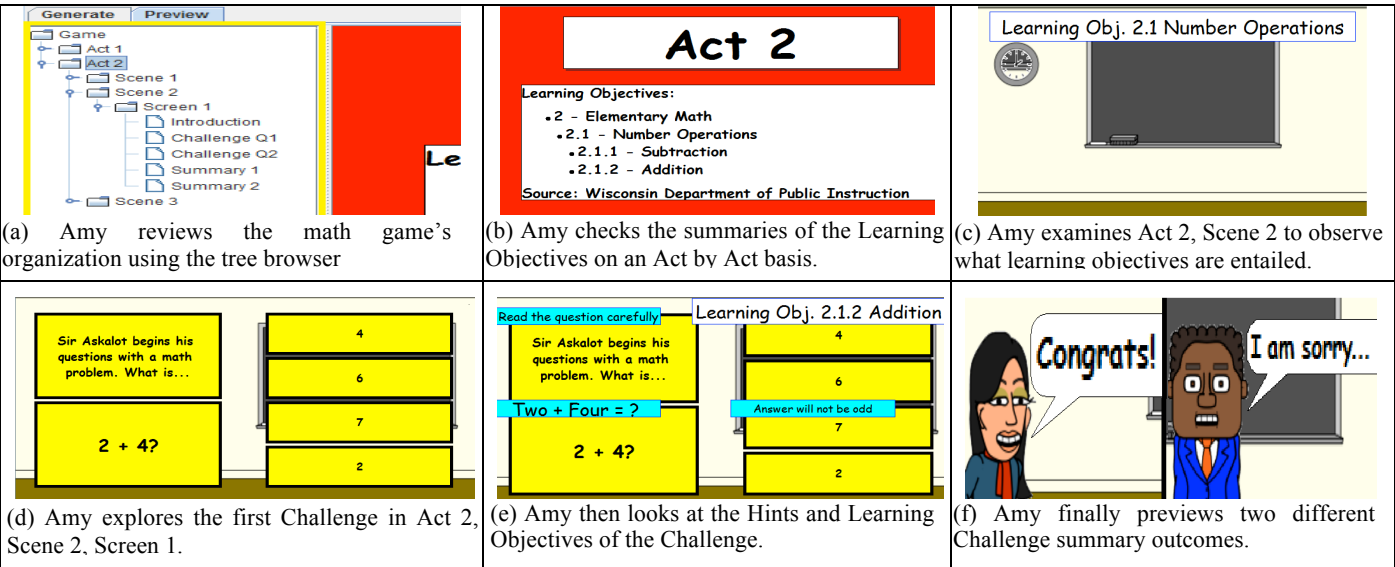


Fig. 3 Scenario 1 Amy Explores the Game’s Learning Objectives



not exhaustive; they are intended to provide a sense of the approach and the prototype tool.

#### A. Learning Objectives Focus

In Fig. 3, Amy explores the overall organization of the

using SEGs to monitor student learning [6]. In many contexts, SEGs have been shown to sustain engaged learning [12], and improve users' feelings of competence [24].

Tools such as ARGILE [10] can help domain experts to

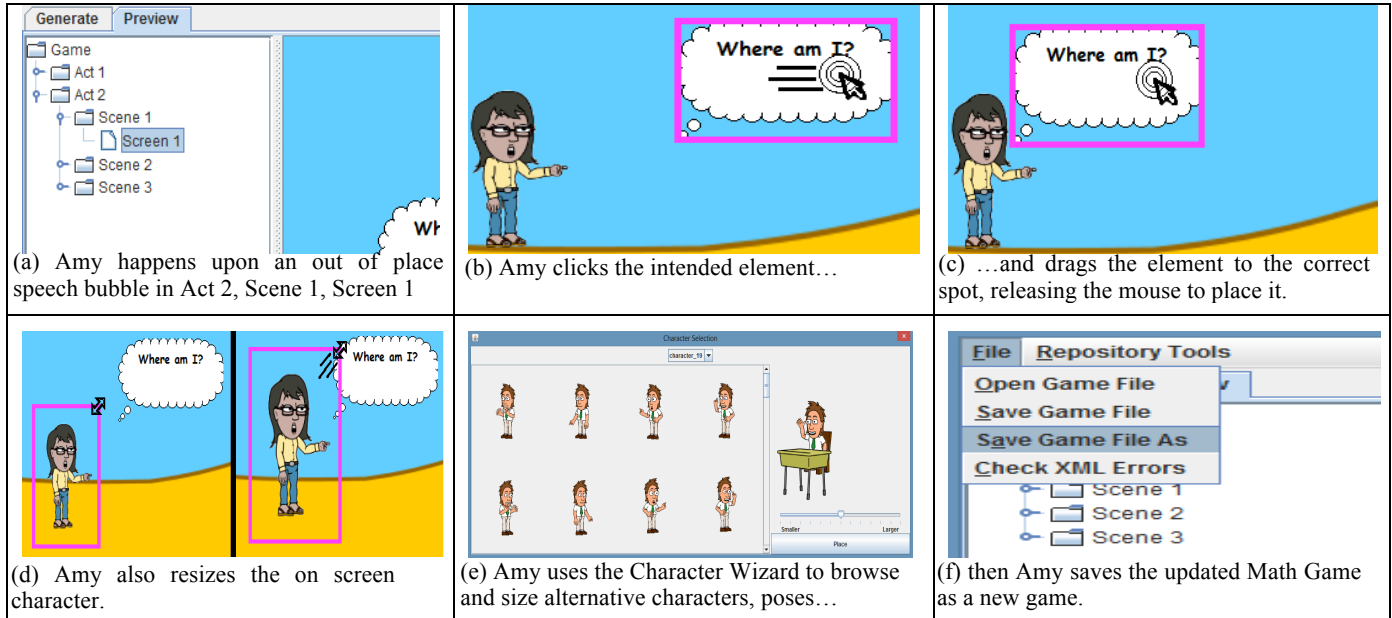


Fig. 4 Scenario 1 Amy Explores the Aesthetics and Traditional Game Content

game using the Browser (Fig. 3a), then progresses to explore the learning objectives covered in Act 2 (Fig. 3b) and scene level (Fig. 3c). She then decides to take a look at the first challenge (Fig. 3d), to see the description, question, and answer options. Amy then takes a more detailed look checking the hints and the learning objectives for the challenge, by hovering the mouse over the challenge in the Interactive Display area. (Fig. 3e), followed by the feedback options presented to the students when they are correct or incorrect (Fig. 3f).

#### B. Traditional Aesthetics and Game Content Focus

In Fig. 4, Amy explores the overall organization of the game using the Browser Fig. 4(a) and notices a speech bubble looks out of place. She mouse selects the speech bubble in the Interactive Display area and drags it closer to the character (Fig. 4b and Fig. 4c). Amy prefers the character to be bigger, so she mouse selects the character and resizes it in the Interactive Display area by dragging the corner of the image (Fig. 4d). She then decides to take a look at alternative characters and poses by selecting the Character button in the toolbar. A wizard is presented, where she has the option to select and add a new element (Fig. 4e). She decides to stay with the current character by cancelling; then continues on to save the work as a new game file (Fig. 4f).

### V. RELATED WORK

Recent research into the efficacy of SEGs has ranged across the spectrum of applications from corporate training [2] to human anatomy courses [19]; in some cases this prior work has also evaluated corresponding tools for SEG development in teaching non-computing students relevant concepts, or in

collaboratively develop and annotate the rules of knowledge-intensive SEGs. Recent Domain Specific Visual Languages (DSVL) have proven useful in programming abstract tasks ranging from algorithmic behavioral specifications (Vibes [14]) to home automation design (Pantagruel [9]). SimSYS combines advantages from both of these bodies of work to allow experts in other domains (besides SEG development) to work with an SEG in a visual programming language.

A wide variety of development environments for SEGs have been proposed [3][8][16][21][25], which focus on serious computer science education through the process of developing games. The eAdventures Project [20] has produced a DSVL for developing serious educational games in other fields of study. This is closest in spirit to our own work, targeting educators and domain-experts as primary users of the tool. Aspects of eAdventures that we consider to be predecessors of our own tool include the chapter/section structure for previewing story flow, the use of hints as game elements, and the ability to transform the visual description automatically into a playable game.

SimSYS differs from eAdventures in the incorporation of learning outcomes as an integral part of the game structure, rather than as optional annotations. The SimSYS tool generates and previews game variants based upon an existing, defined knowledge base with explicit learning outcomes and a repository of applicable game elements, rather than focusing on authoring complex narrative control flow choices from scratch. Our goal from the outset is to equip an educational tool with a DSVL, as opposed to making a game design tool become educational. As a result, assessment is more explicit,

and learning standards are more easily tracked on an institutional basis.

## VI. CONCLUSIONS AND FUTURE WORK

In order to take advantage of the many instructional benefits of SEGs, instructors require powerful, flexible solutions that can help them to effectively navigate the complexity of game development. This then frees them to concentrate on the pedagogical goals of the game. The Game Generation module, part of the SimSYS Game Development Platform, aims to be such a solution, providing a collection of approaches to generate and preview SEGs. Unlike other tools, our system begins with an educational foundation, and helps the instructor to build a high-quality game upon it.

In this paper, we present SEG<sub>Mod</sub>, a lightweight approach for interactively modding SEGs that allows a domain expert (but novice game designer) to explore, visualize, listen to, and tune specific elements of a game. For example, the designer can browse the existing organization and learning objectives of a game, in effect its infrastructure, but cannot modify them as this requires substantial expertise to maintain the consistency of the game. They can, however, modify the aesthetics and traditional game play elements, such as changing the backdrop or background music for a scene; add, modify, or delete the display of characters, props and their sound effects. It is not intended to be a broad, comprehensive solution that supports modifying every aspect of the game design. Instead, to keep it simple and easy to use, we have proposed an approach with a relatively modest subset of functionality.

A prototype tool has been developed as part of the validation work; its features and performance have been tested and the results are encouraging. A comprehensive usability study is needed to fully to evaluate the set of functionality proposed, quantifying its strengths and limitations.

## REFERENCES

- [1] P. Airasian, K. Cruikshank, R. Mayer, P. Pintrich, J. Rath, M. Wittrock, *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*, 2000. L. Anderson and D. Krathwohl, eds., Allyn and Bacon.
- [2] A. Azadegan, J. C. K. H. Riedel, and J. Baalsrud Hauge. Serious games adoption in corporate training. In *Proceedings of the Third International Conference on Serious Games Development and Applications*, SGDA'12, pages 74-85, Berlin, Heidelberg, 2012. Springer-Verlag.
- [3] V. E. Bennett, K. H. Koh, and A. Repenning. CS education re-kindles creativity in public schools. In *Proceedings of the 16th Annual Joint Conference on Innovation and Technology in Computer Science Education*, ITiCSE '11, pages 183-187, 2011. ACM.
- [4] K. Cooper and C. Longstreet, *Towards Model-driven Game Engineering for Serious Educational Games: Tailored Use Cases for Game Requirements*, in *Proceedings of the IEEE 17th International Conference on Computer Games*, 2012, pp. 208-212.
- [5] K. Cooper, E. Nasr, E., and C. Longstreet, "Towards Model-Driven Requirements Engineering for Serious Educational Games: Informal, Semi-formal, and Formal Models". In *Proceedings of the 20th International Working Conference on Requirements Engineering: Foundation for Software Quality*, 2014, pp. 17-22.
- [6] S. Cuomo, V. Fuccella, and A. Murano. Full formative assessment based on educational video games. In *Proceedings of DMS 2010: 16th International Conference on Distributed Multimedia Systems*, pages 228-231. Knowledge Systems Institute, 2010.
- [7] R. Daconceicao, C. Locke, S. Longstreet, and K. Cooper, "Semi-automated Serious Educational Game Generation: A Component-based Game Engineering Approach", in *Proc. of IEEE 18th Int. Conference on Computer Games*, 2013, pp. 222-227.
- [8] S. Dekhane and X. Xu. Engaging students in computing using GameSalad: A pilot study. *J. Comput. Sci. Coll.*, 28(2):117-123, Dec. 2012.
- [9] Z. Drey and C. Consel. Taxonomy-driven prototyping of home automation applications: A novice-programmer visual language and its evaluation. *J. Vis. Lang. Comput.*, 23(6):311-326, 2012.
- [10] N. El Mawas and J.-P. Cahier. Towards participative and knowledge-intensive serious games. In *Proceedings of the Third International Conference on Serious Games Development and Applications*, SGDA'12, pages 86-97, Berlin, Heidelberg, 2012. Springer-Verlag.
- [11] R. Garriss, R. Ahlers, and J. Driskell. Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4): 441-467, 2002.
- [12] J. P. Gee. What Video Games Have to Teach Us About Learning and Literacy. Palgrave Macmillan, 2nd edition, 2007.
- [13] V. Guillén-Nieto and M. Aleson-Carbonell. Serious games and learning effectiveness: The case of *It's a Deal!*. *Computers and Education* 58: 435-448, 2012.
- [14] G. Güleşir, L. Bergmans, M. Akşit, and K. van den Berg. Vibes: A visual language for specifying behavioral requirements of algorithms. *J. Vis. Lang. Comput.*, 24(5):350-364, 2013.
- [15] IEEE Computer Society, SWEBOK 3.0 *Guide to the Software Engineering Body of Knowledge V3.0*, 2014, P. Bourque and R. Fairley eds., available at: <http://www.computer.org/portal/web/swebok>.
- [16] Ioannidou, A. Repenning, and D. C. Webb. AgentCubes: Incremental 3D end-user development. *J. Vis. Lang. Comput.*, 20(4):236-251, Aug. 2009.
- [17] H. Jayaraj, R. Raiker, K. Cooper, and C. Longstreet, *SimSYS Game Specification*, Technical Report UTDCS-06-14, The University of Texas at Dallas, 2014.
- [18] C. Longstreet and K. Cooper, *A meta-model for developing simulation games in higher education and professional development training*, in *Proceedings of the IEEE 17th International Conference on Computer Games*, 2012, pp. 39-44.
- [19] M. Ma, K. Bale, and P. Rea. Constructionist learning in anatomy education: What anatomy students can learn through serious games. In *Proceedings of the Third International Conference on Serious Games Development and Applications*, SGDA'12, pages 43-58, Berlin, Heidelberg, 2012. Springer-Verlag.
- [20] E. J. Marchiori, Á. del Blanco, J. Torrente, I. Martínez-Ortiz, and B. Fernández-Manjón. A visual language for the creation of narrative educational games. *J. Vis. Lang. Comput.*, 22(6):443-452, 2011.
- [21] M. Overmars. Teaching computer science through game design. *Computer*, 37(4):81-83, Apr. 2004.
- [22] M. Papastergiou. Digital-Based Learning in High School Computer Science Education: impact on educational effectiveness and student motivation. *Computers and Education* 52: 1-12, 2009.
- [23] SimSYS Game Development Platform for Serious Educational Games, project homepage [www.utdallas.edu/~kcooper](http://www.utdallas.edu/~kcooper).
- [24] E. D. Van Der Spek. Towards designing for competence and engagement in serious games. In *Proceedings of the Third International Conference on Serious Games Development and Applications*, SGDA'12, pages 98-109, Berlin, Heidelberg, 2012. Springer-Verlag.
- [25] T. Wahner, M. Kartheuser, S. Sigl, J. Nolte, and A. Hoppe. Logical thinking by play using the example of the game "space goats". In *Proceedings of the Third International Conference on Serious Games Development and Applications*, SGDA'12, pages 174-182, Berlin, Heidelberg, 2012. Springer-Verlag.
- [26] *Common Core State Standards for Mathematics* Wisconsin Department of Public Instruction, Madison, WI, 2011.
- [27] World Wide Web Consortium (W3C), XML Schema Definition Language (XSD) 1.1 Part 1: Structures, April 2012, available at: <http://www.w3.org/TR/2012/REC-xmlschema11-1-20120405>.

# Visualization Techniques to Empower Communities of Volunteers in Emergency Management

Sergio Herranz, Rosa Romero-Gómez, Paloma Díaz

DEI Lab – Computer Science Department

Universidad Carlos III de Madrid

Leganés, Madrid, Spain

{sherranz, rmromero, pdp}@inf.uc3m.es

**Abstract—** Information and communication technologies might help to emergency communities of volunteers to both empower community participation and improve their capacity to respond to unexpected events. However, designing technology to support these benefits places unique visualization challenges that go beyond the current state of research on public participation tools and related technologies. Empowering these communities requires developing representations that enable collaborative reflection, promote mutual visibility of volunteers' efforts and sustain a shared view of the community. Similarly, it is necessary to create visualization methods that facilitate sense making of large, simultaneous and distributed pieces of heterogeneous information of different priority levels and with different levels of credibility. Accordingly, this paper analyzes these challenges and proposes a multi-view, multi-abstraction-level visualization approach to address them. In particular, it combines time-oriented visualizations, space-filling visualization techniques, interaction mechanisms and coordinated maps to support community participation as well as collaborative and individual sense making. The application of these visualization techniques is discussed through the development of a set of design prototypes.

*Citizen participation; communities of volunteers; emergency management; sense making; visualization techniques*

## I. INTRODUCTION

The field of emergency management has been evolving over time to adapt to the complexity of disasters associated to modern societies. Originally, emergency management was based on rigid and bureaucratic command-and-control approaches in which responsibilities were exclusive of experts and governmental actors. Nowadays, emergency management is gradually moving from this traditional model to a more collaborative and social one that recognizes the importance of the participation of other actors such as non-profit organizations, volunteer groups or even citizens [32]. In this context, communities of volunteers are considered a fundamental asset to face crisis situations within this field [4]. Communities of volunteers are groups of individuals who altruistically collaborate with official emergency organisms and corps and who have accredited skills and knowledge valuable in specific crisis situations.

Current technological advances on mobile and social computing might empower these communities of volunteers to

participate more actively and to receive additional information from external sources of data [19]. Communities are not isolated structures [33]; they need to complement their perspective with other sources and including citizen-generated data in their perspective might contribute to take profit from the potential of citizens as natural information seekers [24]. For this to be possible, some design challenges need to be addressed, which can be classified into: (1) *community issues* – emergency communities are complex social structures that require to deal with difficult concepts such as membership [23], collaborative discussions [30], or awareness [10]; and (2) *sense making issues* – emergency communities need to make sense of the emergency information to make well informed decisions. It brings to the forefront the need of enable collective reflection, promote mutual visibility of volunteers' efforts, and sustain a shared view of the community. It also highlights the necessity of integrating multiple, large-scale data, including citizen-generated data as well as simultaneous pieces of information of different priority levels and at various geographic locations. Furthermore, each piece of information may come from many different roles of citizens according to their level of credibility [12].

To address these challenges, this paper presents a multi-view, multi-abstraction-level visualization approach aimed at assisting both community participation and sense making activities. We utilize visualization because it can provide varied and richer ways to perceive the community activity and, contribute to improve emergency awareness. In particular, this approach combines time-oriented visualizations, space-filling visualization techniques, interaction mechanisms, and coordinated maps as main visualization mechanisms. This approach provides a flexible, interactive way for volunteers to explore both community and emergency information.

The remaining of the paper is structured as follows. Next section first provides an overview of the areas of communities of volunteers, citizen participation, and information visualization. Then, it discusses related work on public participation tools across application domains. Section four analyzes these design challenges and proposes different visualization techniques to address them. It illustrates the application of these techniques through a set of design prototypes. It also presents a usage scenario of these prototypes

and the insights that they can afford. Finally, some conclusions are drawn.

## II. BACKGROUND

As it has been previously identified, the design of technologies that assist community activities involves design challenges related to both community features and citizen participation. Accordingly, this section first reviews literature on communities and citizen participation in emergency management. Secondly, it examines the role of visualization as a potential tool to support community participation and sense making.

### A. Communities: a Structure to Empower Citizen Participation

A community is defined as a group of people who share common concerns, interest or practices, and who interact frequently becoming socially interdependent [6]. Community members evolve into a social unity whose goals rely on the cooperation with other members. A community emerges around a communal identity that includes “*common values, norms, rules, goals, and so on*” [17]. Furthermore, community members need to share a feeling of membership and attachment with the community and its identity. This is what is being conceptualized as “sense of community” [23], which has a positive influence in volunteer’s involvement and participation [7]. The sense of community defines a community using four dimensions:

- *Membership.* This dimension refers to the feeling of identification with the community. It involves boundaries that determine who is part of the community.
- *Influence.* This bidirectional dimension is defined as the sense of having influence or being influence by the community.
- *Integration and fulfilment of needs.* This dimension corresponds to the feelings of being supported by members of the community while also supporting them.
- *Shared emotional connections.* This dimension refers to the interactions between community members that conform a shared history in the community [23]. This history is considered an important source of cohesion for community members.

Similarly, Carroll and Rosson [10] propose another conceptual model of community composed by three facets: *identity, participation and awareness*, and *support networks*. Identity is built through shared experiences and it enables the feeling of membership and belonging to a community. Participation and awareness transform identity into visible activities. Staying aware of community activity is considered also a form of participation. Finally, participation and awareness requires multiples forms of supporting network in which members with different roles can interact between them. Due to all these complicated social and psychological facets, building communities a non-trivial issue.

Communities are considered powerful structures to channel and empower citizen participation [15]. They provide spaces for interaction that contribute to develop a sense of purpose and belonging, a feeling of being part of something [23]. In this way, citizens could feel that their participation counts and has an influence beyond their role of information seekers, moving upwards in the participation ladder [2]. This helps thus to foster community participation [7] and transforms isolated citizens into work partners that exert a real participation.

### B. Citizen Participation in Emergency Management

Citizens from local and surrounding areas are the first responders in case of an emergency warning or disaster [24]. Participation of citizens in these situations can take a variety of forms such as providing first relief, helping in searching and rescuing activities, etc. Citizens, who have been characterized as information seekers [24], are a valuable source of information [25], particularly now that social networks and mobile computing allow them to capture and disseminate information quicker and broader than ever.

However, citizens are a heterogeneous crowd composed by people with diverse skills and capabilities to inform about and evaluate the situation. Keeping with this reality, a model of ecology of participants in emergencies has been defined in [12] as the basis to design technologies adapted to different roles of participation. This model conceptualizes citizen participants in five different roles: citizen, sensor, trusted sensor, node, and agent. As an example of this conceptualization, while a citizen who sends a tweet on a warning is considered a sensor, a volunteer enrolled in an organization, who has accredited experience, is considered a trusted sensor. This latter means that his data can be directly processed because is reliable from the point of view of the authorities in charge of the emergency management protocol.

### C. Visualization for Sense Making and Community Participation

Visualization is formally defined as “*a graphical representation of data or concepts, which is either an internal construct of the mind or an external artifact supporting decision-making*” [31]. Considering its external nature, DiBiase [13] distinguished two purposes of visualization: assisting sense making in the private domain and facilitating visual communication in the public domain. Sense making is in turn defined as “*the cyclical process in which humans collect information, examine, organize and categorize that information, isolate dimensions of interest, and use the results to solve problems, make decisions, take action, or communicate findings*” [21]. Visual communication has been defined as “*the conveyance of ideas and information in forms that can be read or looked upon*” [29].

He also pinpointed the important role of visualization tools as a medium to enhance group communication. Studies in collaborative intelligent analysis have shown that visualization techniques can help people to rapidly comprehend complexly tangled information in emergency situations. More recently, Convertino et al. [11], have studied that these techniques can increase the quality of the group reasoning by reducing



judgment bias in collaborative conditions. The use of visualization can influence the level of participation of the members and the information sharing process during community decision-making [35]. As an example, shared visualizations can help working groups to communicate more effectively by externalizing the communication process [14]. We posit that proper visualization techniques can be used to enhance the capabilities of communities of volunteers in emergency management. On the one hand, they can provide varied and richer ways to perceive the community activity and, on the other, they can contribute to improve emergency awareness particularly when heterogeneous and multiple sources of information are combined.

### III. RELATED WORK

Emergency management is not the only application domain in which visualization mechanisms have been used to empower public participation. This section reviews this application not only within the emergency management domain but also within other application domains such as policy-making, law or education.

Due to their capability to facilitate complex human activities involving the use and organization of geo-spatial information, geo-visualization tools have been the most applied instruments for representing citizen information in emergency management. This application has led to the concept of *Crisis Mapping* [22], which can be defined as "the display and analysis of data during a crisis, usually a natural disaster or social/political conflict". Examples of crisis mapping tools can be found in most phases of emergency management such as *The Wired*<sup>1</sup> for the preparedness phase, *Usahidi Project*<sup>2</sup> as a very well known tool for the response phase, or *OCHA*<sup>3</sup> (Office for the Coordination of Humanitarian Affairs) for the recuperation stage. Fig. 1 shows the Usahidi map used during the Haiti Earthquake in 2010. One benefit of these crisis-mapping tools is that they can increase the situation awareness of a crisis situation, since they display interactive maps of public information reported by citizens either remotely or from the site of the crisis. However, as a downside, they do not make any distinction between roles of citizens according to their level of credibility.

CSAV (Computer-supported Argument Visualization) mechanisms have been used successfully in the domains of policy-making, law or education. They can be defined as "visualization applications for helping people to participate in various kind of goal-directed dialogues in which arguments are exchanged" [1]. Well-known examples of CSAV tools are Carneades [16] or Araucaria [27], which aim to help both citizens and government officials take part more effectively in dialogues for assessing claims. Fig. 2 shows an example of an argument displayed as a directed graph on the Carneades user interface. They provide services for constructing arguments from formal models of legal concepts, and for evaluating and comparing arguments, through node-link graph views. They are mainly focused on presenting complex information in an organized and easily accessible way.

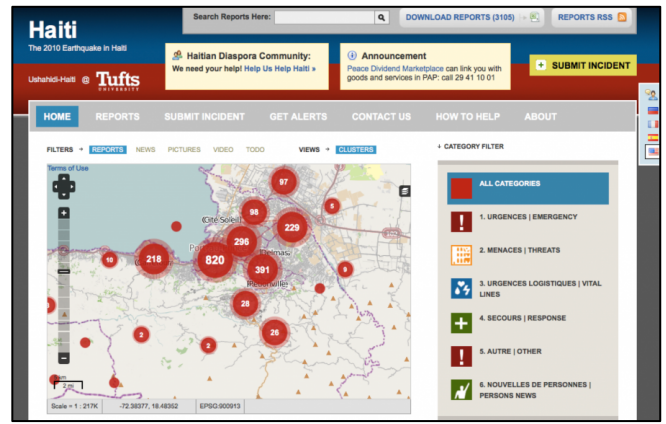


Figure 1. The Usahidi map of Port-au-Prince during the Haiti Earthquake in 2010

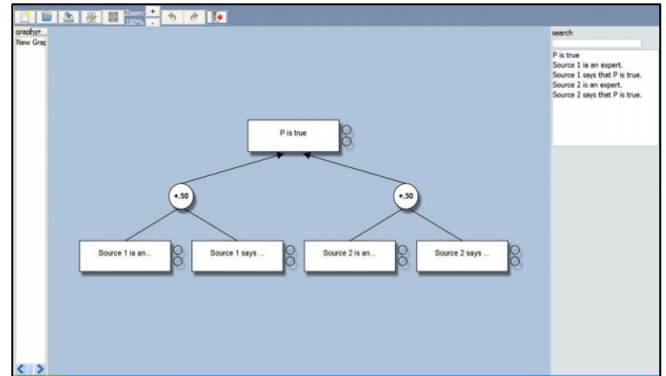


Figure 2. Example of an argument displayed in Carneades

The above summary shows that although visualization is regarded as a powerful tool for empowering citizen participation, most existing tools mainly focus on displaying citizen information in a comprehensible way. There is still very limited support for both encouraging the involvement of citizens through communities and considering the heterogeneous nature of citizen's information. Moreover, these tools still have limitations when high volumes of information are considered. For example, in the case of CSAV tools, they face overlapping issues if large-scale argumentation maps need to be displayed.

### IV. ASSISTING COMMUNITIES OF VOLUNTEERS THROUGH VISUALIZATION

Keeping with the potential of visualization to support sense making and community participation, this section proposes a visualization approach to enhance the capabilities of communities of volunteers through visualization mechanisms. The scenario to illustrate this approach is the participation of communities of volunteers in early warning activities that is a typical contribution in this domain. Communities of volunteers make up a monitoring network that tracks emergency warnings declared by emergency managers in an early stage. Volunteers act then as human sensors, collecting and sharing information about their evolution, the so-called Volunteered Geographic Information (VGI). As communities, volunteers can also collectively reflect upon this information, thus emerging a

<sup>1</sup> The Wire, <http://www.depiction.com/>

<sup>2</sup> Usahidi Project, <http://www.usahidi.com/>

<sup>3</sup> OCHA, <http://www.unocha.org/>

collective picture of the warning that will be analyzed by emergency managers and corps to understand the situation and give a better response. Examples of this type of communities are the “Australian Early Warning Network” (EWN), the “Amateur Radio Emergency Service”, or the Spanish “REMER” (Red Radio de Emergencias).

The following subsections describe both the core design challenges to be addressed; and the visualization and interaction techniques used by our visualization approach through design prototypes implementation. Finally, a usage scenario of how these prototypes might be used to assist volunteers in early warning activities is provided.

#### A. Design Challenges

Based on the previous characterization of both citizen participation and communities in emergency management, a set of design challenges can be identified. These challenges can be classified into two main categories: *community issues* and *sense-making issues*.

Regarding community issues, the following design challenges have been determined analyzing the literature and studying communities of volunteers in Spain [20].

**DC1. Shared view of the community history.** Sharing a history is a fundamental source of cohesion for community members. Indeed, being aware of the activities performed in the community space is also a form of participation. For this reason, it is important to provide a comprehensive and understandable view of the community that allows members to navigate across its history over time.

**DC2. Community reflection.** Communities are social structures that need to collectively reflect upon shared ideas, resources or situations. Reflection involves participating in the community discourse not just by contributing but also by commenting or evaluating existing contributions. Therefore, supporting community work should assist the collaborative exchange of information and facilitate collective reflection.

**DC3. Mutual visibility.** A community achieves real participation when member’s contributions become visible and thus are taken into account. Making clearly visible this investment contributes to increase engagement and generate reliability. Accordingly, it is required to ensure visibility of volunteer’s participation and investment within the community.

The design challenges regarding sense-making issues are the following.

**DC4. Scalability.** The volume of community and citizen participation can vary across crisis situations: the numbers of *tweets* published, the number of messages sent through smartphones, etc. Therefore, visual representations for community information are required to scale from low volumes of information to very high ones.

**DC5. Role distinction.** People with diverse skills and capabilities compose the crowd of citizens who can provide different types of information through different technological platforms and with different level of credibility. Thus, it is necessary to allow the distinction of information across roles of participants.

**DC6. Spatio-temporal interrelationships.** Community volunteers need to track emergency warnings and foresee their evolution across both time and geographical locations in order to support a better response to a situation. Accordingly, enriching community information with citizen participation should allow extracting spatio-temporal interrelationships between these two sources of information.

**DC7. Details-on-demand.** Not all emergencies require the same degree of response or attention, and each incident should be evaluated on a case-by-case basis. Therefore, it is required to provide flexible navigation and interaction to volunteers across information.

#### B. Design Prototypes

The following subsections describe the implications of our visualization approach in terms of visualization techniques according to the design challenges previously defined (DC1 to DC7), grouped in *community issues* and *sense-making issues*. In order to better illustrate our approach, a set of design prototypes was implemented using WPF<sup>4</sup> technology (Windows Presentation Foundation).

##### 1) Supporting Community Issues

In order to support community issues, it is required that volunteers can access and collectively reflect upon the information and items handled at the community scope. As shown the design prototype of Fig. 3, with the purpose of providing a shared view of the community history (**DC1**), an *interactive-annotated timeline* is proposed. Through the use of this technique, community volunteers can navigate among different records that describe the community activity across time, therefore assisting a temporal understanding of the community activity. These records are represented as labels. Trying to facilitate distinction across records, each of these labels is framed with different colors according to the nature of the record including warnings, events, resources, discussions, etc. Hovering the mouse over a label provides an extended summary of the most interesting information about the record. By clicking on it, a volunteer can access to the entire information about the record. In this way, community volunteers can progressively narrowing down the information and focusing on those records of interest.

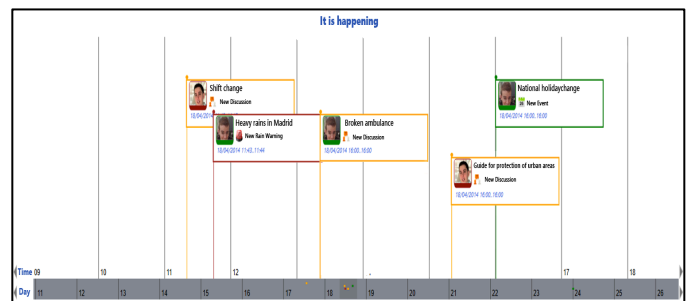


Figure 3. Interactive-annotated timeline for assisting community history

<sup>4</sup> WPF, <http://msdn.microsoft.com/es-es/library/ms754130.aspx>



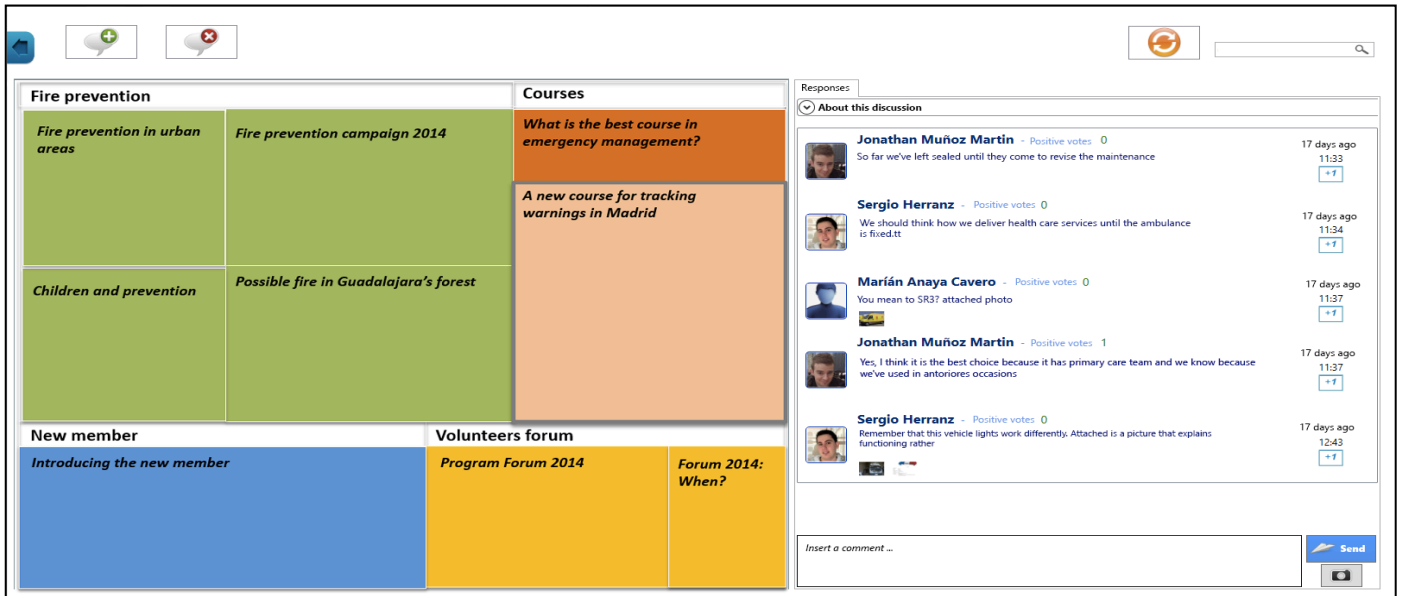


Figure 4. Treemap-based visualization for community reflection

Community reflection (**DC2**) must allow volunteers to initiate topics for collective discussion. This is mainly supported by a combination of diverse visualization and interaction techniques. First of all, volunteers can classify initiated discussions by using *tagging mechanisms*. These discussions are then displayed in a *treemap-based visualization* [34] in order to easily identify which of them are having more impact on the community (see Fig. 4). This treemap can be also adjusted by using direct *dynamic filtering controls* in order to allow determining the discussion topics of interest according to the number of discussions topics classified under such tag. Each tag is then divided in smaller rectangles, which represent the discussion topics, sized according to the number of comments associated to the discussion. Once a discussion is selected, all the comments of this discussion topic are displayed in detail on a list view. In this view, volunteers can explore these comments and contribute not only by sending multimedia or textual comments but also by rating existing contributions. This rating mechanism will help to identify relevant contributions [8] and promote a sense of efficacy [26]. In particular, each tag is then divided in smaller rectangles, which represent the discussion topics, sized according to the number of comments associated to the discussion. Once a discussion is selected, all the comments of this discussion topic are displayed in detail on a list view.

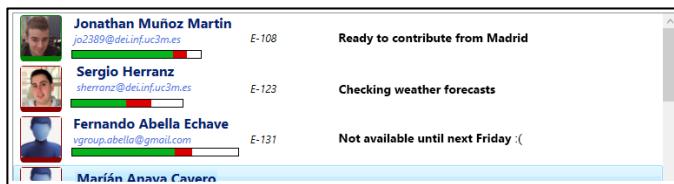


Figure 5. Individual performance bar for mutual visibility

In this view, volunteers can explore these comments and contribute not only by sending multimedia or textual comments but also by rating existing contributions. This rating mechanism will help to identify relevant contributions [8] and promote a sense of efficacy [26].

Finally, supporting mutual visibility (**DC3**) involves making visible member's identity for the community. It is supported by a design prototype divided into two views: a general list of members and a detailed profile area. The list of members goes over the community members by displaying his name, last personal state, and an *individual performance bar* (see Fig. 5). This bar represents a summary of the quantity and quality of member contributions based on the collective rating of his community partners. The length of the bar depicts the quantity of contributions of each member, while the color indicates the quality of these contributions. More specifically, the *green* color represents the proportion of contributions rated mostly positively; the *red* color encodes those contributions rated mostly negatively; and the *white* color those contributions that cannot be classified because either there is not a minimum agreement with their quality or they have been not rated yet by sufficient number of members. By clicking on a particular member on the general list, the detailed profile area displays all the information related to the selected member. Content of these profiles is grouped into personal information such as name, age or interests; professional information including experience, skills, and background; and community information such as community investment and trajectory.

## 2) Supporting Sense-making

In order to achieve a better understanding of the emergencies tracked by the community, volunteers can take advantage of the information provided not only by other volunteers but also by citizens.

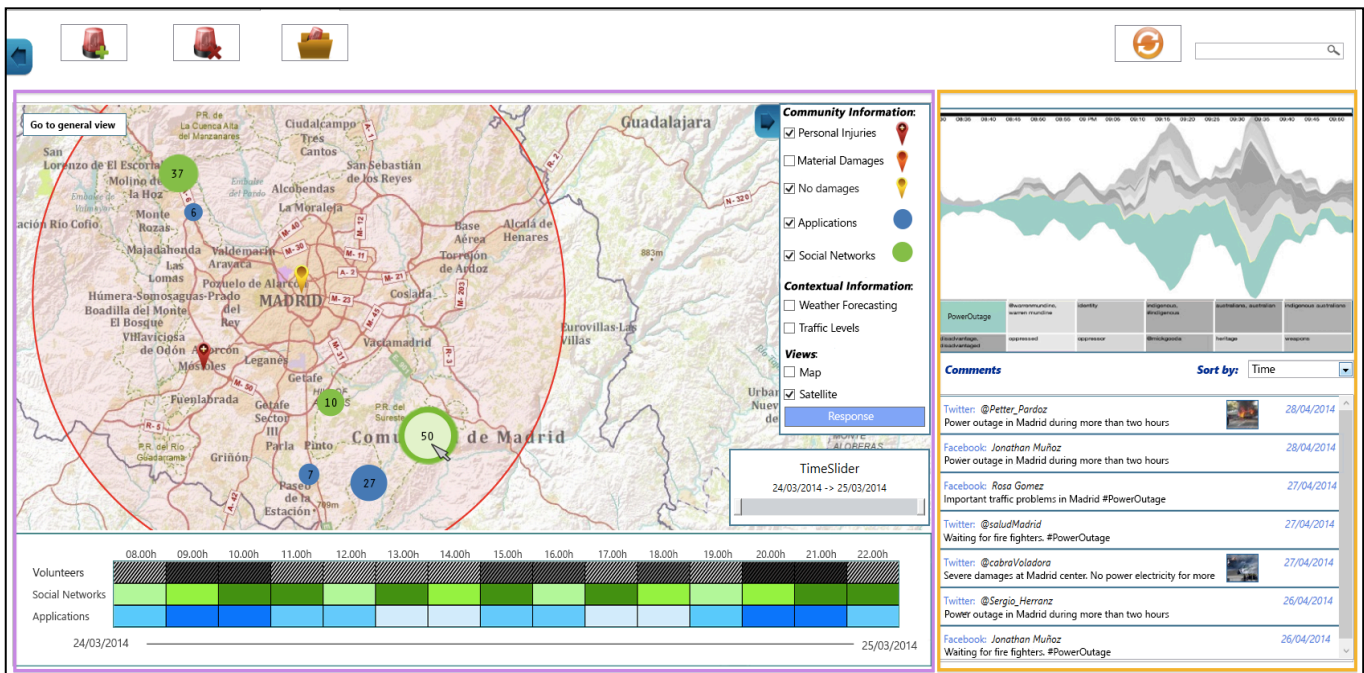


Figure 6. Design prototype for supporting sense-making divided into two main views: Overview (framed in purple) and Detail (framed in orange)

Aiming at assisting this purpose a design prototype have been developed. As shown in Fig. 6, it is divided into two main views: *Overview* (framed in purple), and *Detail* (framed in orange). Overview and detail interfaces have been recognized as effective for coping with scale and complexity (DC4). They allow the user to explore the content methodically, jump around, compare, and contrast [9]. Following paragraphs describe the main constituents of these two views.

The Overview view provides an overall representation of the geographical distribution of warnings and responses provided by volunteers and citizens, as well as their evolution in time (DC6). In particular, this information is provided by means of a combination of a variety of visualization and interaction techniques. A *geographical map* displays warnings and responses in their geographical position. As it was mentioned before (see Section III), geographical maps facilitate complex human activities involving the use and organization of geo-spatial information. In this geographical map, warnings declared by official agencies are encoded as *areas*. Responses to these warnings are represented in the map making use of different visual marks to distinguish between roles of citizens including *icons* for volunteers' responses and *colored circles* for citizen information (DC5). For volunteers' responses, it also uses a *three-color scale* for encoding different levels of priority according to damages. Three or four priority levels are generally recommended [28]. Using too many priority levels, it might lead to situations where higher-priority problems could be delayed too long. This color scale allows assessing response priorities, and therefore developing a better coordination of resources. In particular, it applies *red* for the highest level of priority, *orange* for the middle level, and *yellow* for the lowest one. For citizens' responses, blue circles are used for encoding citizen information reported through mobile applications such

as SafetyGPS<sup>5</sup>. Green circles are used for citizen information reported from social networks such as Twitter or Facebook. This map also provides a set of *layers* that organizes the previous information in categories and supplies additional information (DC7). These layers thus can be turn on and off by interest and are divided into two main groups: *community information*, which classifies all the previous warnings and responses; and *contextual information* such as weather forecasting or traffic levels. Layering seeks to avoid overloading the available space on the map as well as the user's cognitive abilities. Furthermore, aiming at avoiding overlapping issues to high volumes of simultaneous citizen information, it sets the *size* of circles to show the aggregate value of the citizen participation (DC4). Finally, a *double time-slider* is provided in this map in order to allow understanding the temporal evolution of warning and responses. To high volumes of community information, it allows to delimit the information range at will supports the experience of exploration (DC7).

The second graphical element is a *heatmap-based table* that shows an overview of the level of participation of both volunteers and citizens over time. In this table, higher levels of participation are represented by darker colors on each cell and lower levels of participation by lighter ones. It facilitates the visualization and comparison of participation at a certain time interval. It is also highly coordinated with the geographical map through both the *double time-slider* and *brushing and linking* mechanisms [5]. According to previous research in information visualization [3], showing several coordinated views provide useful high information density in context. In this way, it allows to extract spatio-temporal interrelationships between sources of information, without letting lose track of

<sup>5</sup> SafetyGPS, <http://www.safetygps.com/>

the current position of the data (DC6). In particular, the horizontal axis represents a time interval, which can be adjusted by using the *double time-slider*. Consequently, the information displayed on both elements, the geographical map and the heatmap-based table will be the same. The vertical axis represents different roles of citizens including community volunteers and citizens reporting information through different platforms such as mobile applications or social networks. In order to keep consistent across views; the same previous colors are used in this visualization excepting for volunteers' information, which is represented by a gray-scale to sum up the three levels of priority. Accordingly, icons and colored circles on the geographical map are clickable, which form *dynamic queries* whose matching set is immediately shown in the heatmap-based visualization. This heatmap-based visualization will be updated each time new relevant responses come.

In order to support a flexible navigation and interaction across information, the Detail area provides additional views, tightly coupled with the geographical map (DC7). According to the item selected through this map, a different detailed view is displayed (see Fig. 7 and 8). Therefore, this map can be characterized as a visual index to navigate through several levels of detail. Showing both several coordinated views and different levels of detail provide useful high information density in context. Each detailed view resides in a separate panel. In particular, a *bar graph* (shown at Fig. 7) is displayed if citizen information reported through applications, encoded as a blue circle, is selected. It shows comparisons among types of citizen responses. The horizontal axis shows a set of types of responses, which can inform about fires, accidents, need of emergency care and so on. The vertical axis represents the quantity of responses of each type. Selecting on a particular bar shows text lists of responses classified under the type selected. The rest of the bars are adjusted by substituting their color with a gray scale fill. It helps to focus on a subset of interest, while the general context of the data is preserved.

A *streamline graph* [18] is displayed (shown at the Detail view in Fig. 6) if citizen information provided from social networks, encoded as a green circle, is selected. This graph reveals the most relevant topics reported by citizens through social networks over time. The total participation is shown by the varying heights the stripes reach over time. The horizontal axis represents time and each stripe represents a relevant topic. The thickness of a stripe shows the number of messages related to this topic in the given time period. The color encodes a topic. When a stripe is selected, messages related to the topic are shown below the graph. As in the bar graph, the rest of the stripes are adjusted by substituting their color with a gray scale fill.

Finally, if a specific community volunteer response, encoded as a colored icon, is selected, it displays a textual view of the response information (shown at Fig. 8). In particular, it includes a summary of the respondent, textual and multimedia information related to the response, rating information from other volunteers, and the set of comments sent by other members. Both rating information and comments pursue generating reliability on the response.

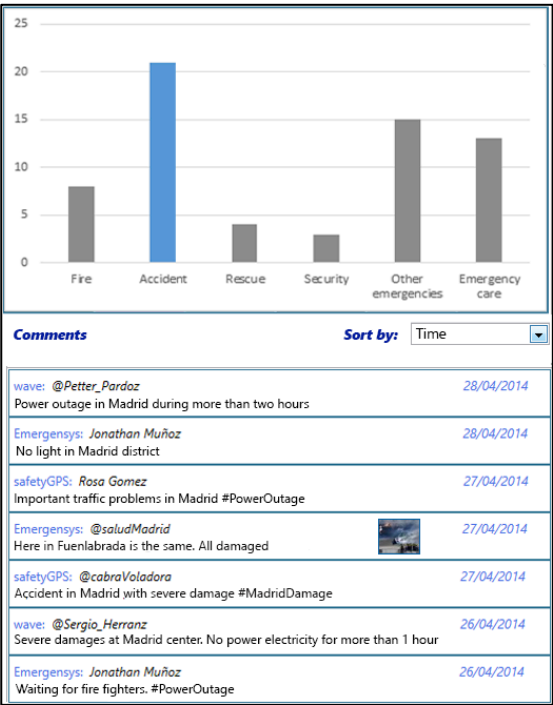


Figure 7. Detailed view for citizen information reported through applications



Figure 8. Detailed view for a specific volunteer response

C. Usage Scenario

Consider the case of a community of volunteers tracking a heavy rain emergency warning declared for the Madrid area.



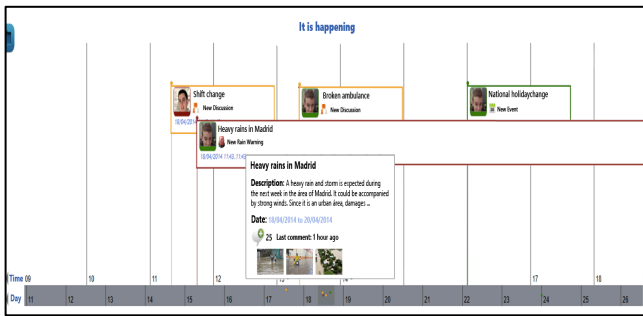


Figure 9. Inspecting the interactive-annotated timeline

Given these circumstances, a volunteer would like to understand relevant information to the situation in order to both make well-informed decisions and inform authorities facilitating a better response.

The volunteer begins by navigating through the interactive-annotated timeline (see Fig. 9). Upon visual inspection, he notices that a warning has been active for the past few days. By hovering the label corresponding to the warning record, he can characterize it as a high-level priority one. Similarly, he can perceive that there have been several recent contributions to this warning by other volunteers of the community. For this reason, he decides to get further information about both its spatial and temporal context and the magnitude of the situation. With this purpose, he clicks on such label in order to display a geographical map that represents warnings and responses in their geographical position. As mentioned before, a warning is represented in the map as a circular area. Within this area, he can explore both community responses, displayed as colored icons, and external citizen-generated information, displayed as colored circles (see Fig. 10). He also can make use of the double time-slider in order to understand their temporal evolution. In particular, the time period of interest is configured to the last twelve hours. Thanks to this mechanism, he can understand that both other volunteers and several citizens at Madrid center have reported several responses related to the warning. Moreover, given the coordination between the time slider and the heatmap-based table, he is able to perceive that participation levels have been higher in the last two hours. In particular, he can characterize social networks as the most active channel of participation (see Fig. 11).

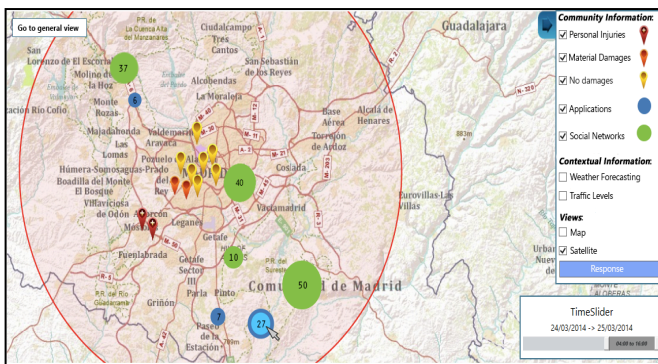


Figure 10. Understanding warning evolution through the Geospatial map. (Time-slider comprises last twelve hours of the warning)

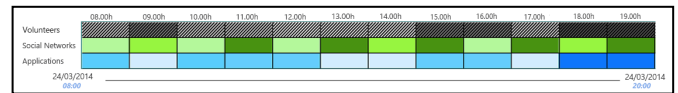


Figure 11. Perceiving levels of participation by using the Heatmap-based table.

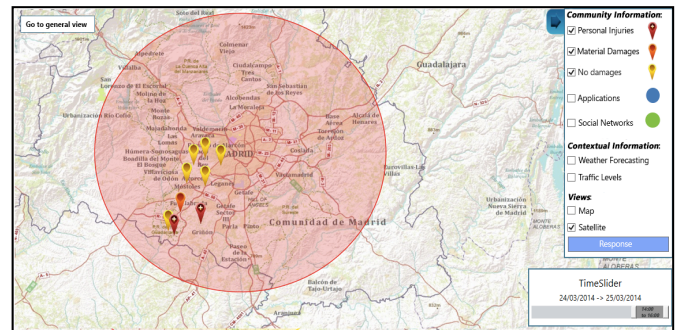


Figure 12. Understanding warning evolution through the Geospatial map. (Time-slider comprises last two hours of the warning)

However, prior to analyze these high volumes of information coming from social networks, he decides to explore the information provided by the community volunteers. In contrast to the information provided by social information coming from social networks, these responses can be characterized as more structured and reliable information. For this end, he updates the time-slider period from the last twelve hours to the last two hours. Similarly, he makes use of the map layers in order to specify the display on the map of just volunteers' responses (see Fig. 12). Based on this customized view of the warning-related information, he can perceive that during this two-hours period most of responses have been reported from Madrid southwest area. In particular, he identifies a set of high-level priority responses referred to personal injuries. He decides consequently to focus on one of them by clicking on the corresponding icon. As a result, a more detailed view is displayed, which shows a summary of the respondent's profile, a textual and multimedia description of the report information, and a set of comments provided by other members. The textual description explains that several pieces of a building cornice fell down, causing injuries to citizens passing by the building. The multimedia description includes different pictures that display the current state of the building. Aiming at knowing the reliability of the respondent, he decides to access to the respondent's profile. According to the individual performance bar (see Fig. 13), it seems that most of the contributions of this respondent have been rated positively. Consequently, the volunteer can characterize this response as a reliable piece of information. Similarly, by exploring the comments associated to this response, he perceives that some members have been discussing about the number of injured citizens provoked by this event. According to these comments, the number of slightly injured citizens varies from ten to fifteen citizens and none of them is severely injured. He therefore gets a better idea of the seriousness of this situation. Nevertheless, given the previously identified high-level of citizen participation through social networks, he wants to explore further such information and identify if there are relationships with the volunteers' responses.

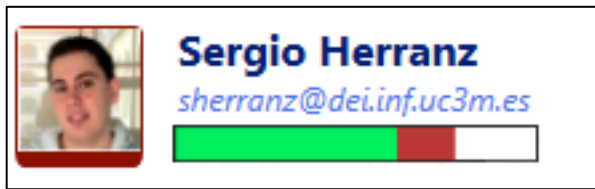


Figure 13. Checking reliability of the respondent through his Individual performance bar

With this goal in mind, the volunteer makes use again of the map layers in order to specify the addition to the map of citizen-generated information. In this new view of the map, he can observe that there exists a big green circle, located at southwest Madrid, which represents a high volume of citizen-generated messages. Given the geographical proximity to the previous high-level priority volunteers' responses, he clicks on this circle in order to get a more detailed view. This action displays a streamline graph that shows the discussion over time for the topics most relevant associated with the warning (see Fig. 14). By clicking on the widest stream of this graph, encoded in green and tagged as "power outages", he can explore the messages and multimedia content associated to this topic. After exploring them, he discovers that the cornice falling has lead to electrical damages in such street. The risks associated to an extended loss of electrical power in that area makes this information particularly relevant to be communicated to authorities.



Figure 14. Understanding most relevant topics in Social Networks channels by using the Streamline graph

## V. CONCLUSIONS

Visualization has proven its worth over time as a relevant instrument to assist sense making and enhance group activities. In keeping with this importance, this paper envisions a visualization approach aimed at empowering communities of volunteers focused on early warning activities. The techniques here proposed seek to enhance community participation and facilitates sense making of emergency situations by integrating external citizen information. The design prototypes show how using elements such as space-filling or interactive time-oriented techniques could assist collective reflection and community visibility. In terms of facilitating sense making, this solution must deal with large pieces of information from a variety of sources with different levels of credibility. The proposed visual approach copes with these issues by utilizing coordinated maps and different levels of information abstraction. It also provides a number of design ideas that could be employed by others working on similar problems or even in different domains containing similar design issues.

Further work will be focused on evaluating the design prototypes proposed in this paper. However, it has to be considered that technologies for EM cannot be tested in complex real situations where agencies and corps are not willing to try new technologies but to be as efficient as possible applying the procedures they are familiar with. For that reason, an evaluation under controlled settings could prove valuable. Volunteers of communities supporting early warning activities could interact with the design prototypes on a set of pre-defined and realistic scenarios. Afterwards, we would like to explore their perception about the usability of the design prototypes as well as the usefulness as an instrument to support community activities.

## ACKNOWLEDGMENT

This work is supported by the project emerCien grant funded by the Spanish Ministry of Economy and Competitiveness (TIN2012-09687).

## REFERENCES

- [1] I. Alexander, "Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making", *European Journal of Information Systems*, vol. 12, pp. 163-164, June 2003.
- [2] S. R. Arnstein, "A Ladder of Citizen Participation", *Journal of the American Institute of Planners*, vol. 35, no. 4, pp. 216-224, 1969.
- [3] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky, "Guidelines for Using Multiple Views in Information Visualization". *Working conference on Advanced Visual Interfaces*, pp. 110-119, 2000.
- [4] T. Baxter-Tomkins, and M. Wallace, "Emergency services volunteers: What do we really know about them?" *Australian Journal on Volunteering*, vol. 11, no. 2, pp. 7-12, 2006.
- [5] R. A. Becker and W. S. Cleveland. "Brushing scatterplots", *Technometrics*, vol. 29, no. 2, pp. 127-142, May 1987.
- [6] R. N. Bellah, "Habits of the Heart: Individualism and commitment in American life". New York: Harper and Row, 1985.
- [7] A. L. Blanchard, and M. L. Markus, "The experienced sense of a virtual community: Characteristics and processes", *Data Base Adv. Inform. Syst.*, vol. 35, no. 1, pp. 65-79, 2004.
- [8] M. Bourimi, "Collaborative design and tailoring of web based learning environments in CURE". (S. B. Heidelberg, Ed.) *Groupware: Design, Implementation, and Use*, pp. 421-436, 2006.

- [9] S. Card, J. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: using vision to think*. San Francisco: Morgan Kaufmann, 1999.
- [10] J. M. Carroll, and M. B. Rosson, "Wild at Home: The Neighborhood as a Living Laboratory for HCI", *ACM Transactions on Computer-Human Interaction*, vol. 20, no. 3, 2013.
- [11] G. Convertino, D. Billman, J. Shrager, P. Pirolli, and J. P. Massar, "The CACHE study: group effects in computer-supported collaborative analysis", *Computer-Supported Cooperative Work*, vol. 17, no. 4, pp. 353–393, 2008.
- [12] P. Díaz, I. Aedo, M. Romano, and T. Onorati, "Ecology of participants. A Design Framework to Support Citizen 2.0 in Disaster Response", *Information Systems Management*, vol. 31, no. 4, 2014. In press.
- [13] D. DiBiase, "Scientific visualization in the earth sciences", *Bulletin of the College of Earth and Mineral Sciences*, vol. 59, no. 2, pp. 13–18, 1990.
- [14] J. M. DiMicco, K. J. Hollenback, D. C. Neale, P. L. Isenhour, J. M. Carroll, and M. B. Rosson, "The impact of increased awareness while face-to-face", *Human-Computer Interaction* vol. 22, pp. 47–96, 2003.
- [15] P. Florin, and A. Wandersman, "An introduction to citizen participation, voluntary organizations, and community development: Insights for empowerment through research", *American Journal of Community Psychology*, vol. 18, no. 1, pp. 41–54, 1990.
- [16] T. F. Gordon, "Visualizing carneades argument graphs", *Law, Probability and Risk*, vol. 6, pp. 109–117, 2007.
- [17] M. Gurstein, "What is community informatics (and why does it matter)?", *Polimetrika sas*, vol. 2, 2007.
- [18] S. Havre, B. Hetzler, and L. Nowell, "ThemeRiverTM: In search of trends, patterns, and relationships", *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 1, pp. 9–20, 2002.
- [19] S. Herranz, P. Díaz, D. Díez, D., I. Aedo. "Studying Social Technologies and Communities of Volunteers in Emergency Management", *International Conference on Communities and Technologies*, 2013.
- [20] S. Herranz, D. Díez, P. Díaz, and S. R. Hiltz. "Exploring the Design of Technological Platforms for Virtual Communities of Practice", *Information Systems for Crisis Response and Management Conference (ISCRAM)*, 2012.
- [21] G. Klein, B. Moon, and R. R. Hoffman, "Making sense of sensemaking 1: Alternative perspectives", *Intelligent Systems, IEEE*, vol. 21, no. 4, pp. 70–73, 2006.
- [22] S. B. Liu, and A. A. Iacucci, "Crisis Map Mashups in a Participatory Age", *ACSM Bulletin*, June 2010.
- [23] D. W. McMillan, and D. M. Chavis, "Sense of community: A definition and theory", *Journal of community psychology*, vol. 14, no. 1, pp. 6–26, 1986.
- [24] L. Palen, K. M. Anderson, G. Mark, J. Martin, D. Sicker, M. Palmer, and D. Grunwald, "A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters", *ACM-BCS Visions of Computer Science Conference* (pp. 8). British Computer Society, 2010.
- [26] J. Porter, *Designing for the social web*. New Riders, 2010.
- [27] C. Reed, and G. Rowe, "Araucaria: Software for argument analysis, diagramming and representation", *International Journal on Artificial Intelligence Tools*, vol. 13, no. 4, pp. 961–979, 2004.
- [28] D. H. Rothenberg, "Alarm management for process control: A best-practice guide for design, implementation, and use of industrial alarm systems", McGraw-Hill Education, 2009.
- [29] D. Sless, "Learning and visual communication", *Educational Communication and Technology*, Spring, vol. 31, no. 1, pp. 58–60, 1983.
- [30] K. Stanoevska-Slabeva, and B. F. Schmid, "A typology of online communities and community supporting platforms", *Hawaii International Conference IEEE*, 2001.
- [31] C. Ware, *Information visualization: Perception for design*. Elsevier, 2012.
- [32] W. L. Waugh, and G. Streib, "Collaboration and leadership for effective emergency management", *Public Administration Review* 66, pp. 131–140, 2006.
- [33] E. Wenger, *Communities of practice: learning, meaning and identity*. Cambridge: Cambridge University Press, 1998.
- [34] B. Shneiderman, "Tree visualization with tree-maps: 2-d space-filling approach", *ACM Transactions on graphics (TOG)*, vol. 11, no. 1, pp. 92–99, 1992.
- [35] A. Wu, G. Convertino, C. Ganoë, J. M. Carroll, and X. Zhang, "Supporting collaborative sense-making in emergency management through geo-visualization", *International Journal of Human-Computer Studies*, vol. 71, no. 1, pp. 4–23, 2013.



# PetroAdvisor: A Volunteer-based Information System for Collecting and Rating Petroglyph Data

Vincenzo Deufemia, Luca Paolino, Giuseppe Polese  
DISTRA MIT  
University of Salerno  
84084 Fisciano (SA), Italy  
{deufemia, lpaolino, gpolese}@unisa.it

Viviana Mascardi  
DIBRIS  
University of Genoa  
16146 Genova, Italy  
mascardi@unige.it

Henry de Lumley  
Laboratoire Départemental de Préhistoire du Lazaret  
Parc de la villa La Côte, 33 Bis Boulevard Franck-Pilatte  
06300 Nice, France  
lazaret@lazaret.unice.fr

## Abstract

*In this paper we exploit a volunteer-based information paradigm for archaeological aims. In particular, we present PetroAdvisor, a system supporting several fundamental activities to digitally preserve petroglyph sites. The system also uses a rewarding strategy in order to stimulate people participation to the project, so that those entering useful information gain free archaeological data, tips on excursions and tours, opinions and rating from previous tourists, and so forth. User provided information typically consists of petroglyph pictures, descriptions, and several useful meta-data, such as geo-referenced information, petroglyph contours, and so forth, empowering the work of the archaeologists, and enabling them to tackle technology shortfalls.*

## 1. Introduction

In the last years, the World Wide Web has been used as a means to stimulate knowledge sharing and collaborative work, also letting people launch new models for developing services. This paradigm is generally referred to as “crowd” followed by a term identifying the aim of the service, e.g., crowdsourcing and crowdfunding. Moreover, a crowd service permits to gain information, money, or other services by leveraging the strength of the participating web community.

The World Wide Web has become one of the most powerful medium to start collaborations and to tackle prob-

lems that would otherwise be simply unmanageable. Nowadays, there are already meaningful examples of services provided by means of people combining their resources on the Web. These include the world’s largest knowledge base Wikipedia<sup>1</sup>, several complementary services for GPS-assisted navigation, such as Waze<sup>2</sup>, and the emergency coordination platform Ushahidi-Haiti, which has been used to coordinate disaster response after the 2010 Haiti earthquake [14]. This new way of providing services is commonly referred to as *crowdsourcing* [9, 16], and it may be adapted to gather information while solving real problems in several application domains. One of the problems that may arise in crowdsourcing is due to possible conflicts among the data provided by the crowd. Moreover, there are contexts in which few people have the knowledge necessary to provide reliable data. As an example, this happens in the context of geographic information, where the process of gathering information from users is known as *volunteered geographic information* (VGI) [12].

The aim of this work is to exploit the volunteered geographic information paradigm for several archaeology activities. In particular, we address issues related to the exploration of archaeological sites, in order to derive a complete map of carvings, and digitally capture their images. Indeed, rock carvings represent an invaluable cultural and natural heritage, since they are often located in wonderful natural sites and represent an irreplaceable resource for understanding our history [6, 11]. Unfortunately, they are constantly exposed to weathering and vandalism, and it is up to hu-

<sup>1</sup><https://www.wikipedia.org>

<sup>2</sup><https://www.waze.com>

manity to preserve them for future generations [2]. Moreover, site exploration and carving cataloguing is a complex and expensive work, which cannot always be completely supported only by domain experts, such as the archaeologists. For this reason, VGI appears to be a promising road to explore [1]. In fact, the growing will of people to participate in this kind of projects, together with the capillarity of smartphones, tablets, and wireless connections promises to increase the possibility of having fully mapped sites [10].

In this paper, we present PetroAdvisor, a system supporting archaeologists in the digital preservation of petroglyph sites. In particular, the user provided information gathered by PetroAdvisor consists of petroglyph pictures, their contours, descriptions, comments, ratings, geo-referenced information, and so on. This way of collecting data not only reduces the necessity of man power, but it also allows archaeologists to tackle several technology shortfalls, such as the automatic segmentation and identification of petroglyphs within pictures [27].

However, in order for a VGI-based strategy to be successful, it is necessary to stimulate the participation of people in the data collection process. To this end, we use a rewarding strategy based on a “do ut des”: users entering useful information gain free tourist services, e.g., archaeological data, tips on excursions and tours, opinions and ratings from previous tourists.

Finally, the main objectives of PetroAdvisor are (i) to map, collect, and analyze carving data for digitally preserving petroglyphs, also yielding new research opportunities, (ii) to collect user provided information useful for automating petroglyph recognition, and (iii) to reward tourists with information guiding them during their visits to petroglyph sites.

The remainder of the paper is organized as follows. Section 2 introduces background information on rock art archaeology and provides motivations of this research. Successively, Section 3 shows some details about PetroAdvisor system design and architecture, with particular emphasis on the actors and their functionalities. Some examples of usage scenarios from both user and archaeologist points of view are described in Section 4. Section 5 provides a discussion of works related to the proposed system. Finally, conclusion and future works end the paper in Section 6.

## 2. Context and Motivations

Our work falls in the domain of rock art archaeology, where we worked with prehistorian archaeologists involved in the study of rock carving sites. In what follows we introduce basic knowledge about rock art and the motivations of this work.

### 2.1. Rock Art Archaeology

Rock art is a term coined in archaeology for indicating any human made markings carved on natural stones [4]. Most of the symbols concerning rock art are represented through petroglyphs, which are created by removing parts of a rock surface by incising, picking, carving, and abrading. They are among the oldest forms of art known to humans. Indeed, although early petroglyphs have traditionally been related to the appearance of modern humans in Europe, recent work has shown that they were created much earlier, that is, about 77,000 years ago in South Africa [15].

Usually the archaeologists working on a rock art site collect petroglyphs, classify them based on shape, and define dictionaries. Although it is not possible to give precise interpretations to petroglyphs, archaeologists have proposed many theories to explain their meaning, e.g., astronomical, cultural, or religious [26]. For example, Figure 1(a) and 1(b) show the pictures of two petroglyphs interpreted as a Christ [2] and the stellar cluster of Pleiades [11], respectively, while Figure depicts a digitalized relief interpreted as a priest making water spout from the water basin [6].

In order to digitally preserve, study, and interpret these artifacts, the archaeologists have created repositories containing heterogeneous information, like pictures, 3D images, textual descriptions, GPS coordinates, black and white reliefs, and so on. The exponential growth of these repositories and the high dimensionality of the stored data challenged rock art archaeologists to make deep analysis on them [8]. However, such analysis would never be complete without the contributions of volunteers, due to the amount of necessary work, and to technological limits, especially in the image processing domain. Indeed, although in the recent years several image recognition approaches have been proposed for automating the segmentation and classification of petroglyphs [7, 23, 27], their accuracy is not satisfactory.

To witness the relevance of the above mentioned problems, the European Community, together with other funding agencies, are financially supporting several research projects in this domain, such as Prehistoric Rock Art Trail<sup>3</sup> and 3D-Pitoti<sup>4</sup>. Among them, the IndianaMAS project [20], supported by Italian Minister of Education, involving our computer science research groups from Universities of Genoa and Salerno, together with archeologists from Laboratoire départemental de Préhistoire du Lazaret in Nice and from the Italian Superintendence of Cultural Heritage. The project aims to promote the awareness and the preservation of rock art, and to support archaeologists in their investigation activities. To this end, we are developing a platform that integrates and complements the techniques usually adopted to preserve cultural heritage sites. The plat-

<sup>3</sup><http://www.prehistour.eu>

<sup>4</sup><http://3d-pitoti.eu/>

form exploits ontologies to provide a shared and human-readable representation of the application domain [13], intelligent software agents to analyze the digital objects analysis and to perform reasoning and comparison activities on them [17], together with standard tools and technologies for Digital Libraries to manage and share digital objects. IndianaMAS provides the means to organize and structure petroglyph data in a standard way, supplying domain experts with facilities for issuing complex queries on the data repositories, making assumptions about the lifestyle of ancient people.

## 2.2. Motivations

Archaeologists and interested tourists frequently visit prehistoric rock art sites. By exploiting the advances in digital photography they can examine and investigate the petroglyphs without traveling, e.g., by highlighting the petroglyphs in the image and analyzing their locations, sizes, and orientations.

Unfortunately, there is no robust, automatic segmentation algorithm capable of determining the exact shapes and spatial locations of petroglyphs from rock art pictures [27]. In particular, the exact boundaries of petroglyphs are hard to identify also due to the direction of the light [23]. Therefore, rather than solely relying on developing new and better algorithms to handle such tasks, we propose to exploit volunteered-based solutions, so as to benefit from the contributions of an external community of people. In particular, with respect to petroglyph segmentation, we ask humans to trace the petroglyph contour by means of a touch-screen. This allows us to determine the exact shapes and spatial positions of petroglyphs in pictures, and to successively classify them based on their shapes, which enables retrieving similar petroglyphs from different archives.

Another limitation in the study of petroglyphs comes from the existence of many different repositories, even for a single rock art site. As an example, the Bicknells legacy collection and the ADEVREPAM database are two repositories of Mt. Bego petroglyph site. Thus, in order to support archaeologists in the study of correlations between petroglyphs, it is necessary to create a centralized repository storing all information about rock art sites. Such a repository should contain information provided by archaeologists, like interpretations and reliefs, and those provided by tourists, like pictures and comments.

Finally, in order to motivate the participation of tourists to the previous tasks, it is necessary to implement rewarding services, such as providing them with means for calculating the most appealing visiting paths.

## 3. PetroAdvisor System Design

In this section we present the PetroAdvisor system design. First of all, we identify the actors, namely the users interacting with the system, including their roles. Then, we present a set of system functionalities defined based on the above mentioned issues, and on archeologist suggestions. Finally, we provide details of the PetroAdvisor architecture. In particular, we show the layers and the modular parts of the system.

### 3.1. Actors and Functional Requirements

During the requirement analysis phase performed with the archaeologist assistance, we identified three user categories: *Generic User*, *Archaeologist*, and *Moderator*. In the following, we describe each category together with its functional requirements.

The *Generic User* represents the majority of the people who will use the system. A user adds new pictures upon finding petroglyphs during walks, looks for descriptions about already managed petroglyphs, adds comments, and generates new itineraries when planning the site visit. Moreover, s/he should be able to communicate with other users, exchanging information. As a result, we identified the following *User Functional Requirements* (UFRs):

- UFR1: inserting a new picture
- UFR2: showing petroglyph information
- UFR3: adding comments to petroglyphs
- UFR4: rating petroglyphs
- UFR5: searching petroglyphs through custom queries
- UFR6: managing itineraries.

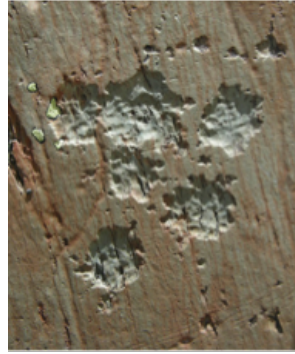
The user's comments contain information about the experience they had during the visit of the archaeological site, their opinions about the beauty of the photographed petroglyphs, and further details that can foster or discourage other users to visit the site. Such comments are also summarized by the user with a rate, in the range [1-5], for the quality and the beauty of the petroglyph.

The second system actor is the *Archaeologist*. S/he is in charge of verifying the quality of the user provided data, and of making them available to *Generic Users*. To this end, other than the previous UFRs, we identified the following additional requirements (AFRs):

- AFR1: creating new petroglyphs
- AFR2: updating petroglyph information
- AFR3: writing comments



(a)



(b)



(c)

**Figure 1.** The picture of a petroglyph supposed to represent a Christ (a) [2], a picture interpreted as the stellar cluster of the Pleiades (b) [11], and the relief depicting christ priests making water spout from the rock (c) [6]

- AFR4: identifying territories to explore.

The involvement of users in the data acquisition process requires an actor in charge for monitoring user behavior and the quality of their contributions. For this reason, we introduced the role *Moderator*, whose task is to control the provided information, and to manage user accesses. More specifically, for the moderator we identified the following requirements (MFRs):

- MFR1: managing users
- MFR2: removing comments
- MFR3: managing discussion groups
- MFR4: removing pictures.

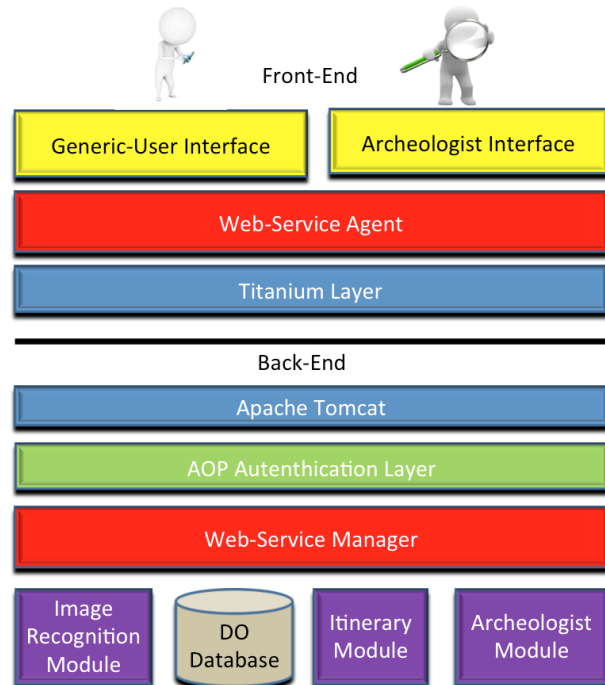
### 3.2. Architecture

Figure 2 shows the web service-based PetroAdvisor architecture. It consists of some third-party software managing communication and development, a web service agent and manager handling requests and responses, an authentication module, two specific interfaces: one for generic users and another for archaeologists, several modules for managing requests, and a database storing information regarding the digital objects.

**Third-party software.** Apache Tomcat is a popular application server. Its native support for the SOA architecture makes it appealing for our purposes.

Less popular is the Titanium SDK used for building the Titanium Layer<sup>5</sup>. In particular, Appcelerator Titanium is a platform for developing mobile applications based on the

<sup>5</sup>[www.appcelerator.com](http://www.appcelerator.com)



**Figure 2.** The PetroAdvisor architecture.

Eclipse SDK. Titanium supports the development of applications for iPhone, Android, iPad, and BlackBerry. One of the most interesting features is the ability to develop mobile applications using well-known web technologies, such as HTML and Javascript.

**Web Service modules.** Web service is a communication paradigm between electronic devices over a network. In particular, it consists of a software function provided at a web address, which is always on like in utility computing.

Web services have been particularly useful in the development of the PetroAdvisor system, since they allowed us to focus our attention on the definition and implementation of the system functionalities, relieving us from presentation issue concerns.

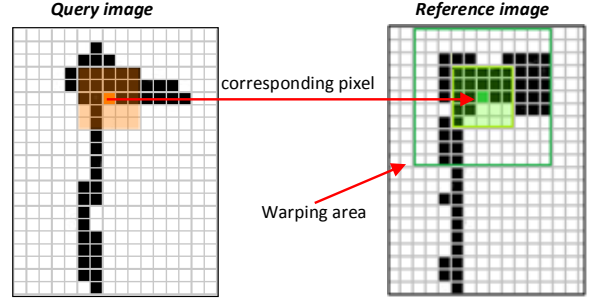
The combination of Web services and the Spring technology allowed us to exploit the AOP (Aspect Object Programming) for the management of security. In fact, we only needed to declare the permissions for each functionality in order to restrict their access.

**Image Recognition Module.** this module aims to recognize petroglyph symbols from petroglyph contours. The recognition process uses a petroglyph symbol classifier based on a flexible image matching algorithm. The idea is to measure the similarity between petroglyphs by using a distance function derived from the image deformation model (IDM) [18], which has been successfully applied to handwritten character recognition [18]. Such a distance measures the displacements of single pixels from the two compared images within a warp range, also taking into account the surrounding pixels (local context). This method is well-suited for petroglyph reliefs since it is less sensitive to local changes often occurring in the presence of symbol variability, making our method tolerant to the types of visual variations. In particular, the IDM model yields a distance measure that is tolerant with respect to local distortions. In fact, in case two images have different values for few pixels, possibly due to noise or artifacts irrelevant to classification, the distance between them is compensated by specifying a region in the matching image where it is allowed to detect a best matching pixel.

The IDM performs a pixel-by-pixel value comparison between the query and the reference images, determining the best matching pixel within a region surrounding the corresponding position in the reference image, for each pixel in the query image. The IDM has two parameters: warp range ( $w$ ), and context window size ( $c$ ). Figure 3 illustrates how the IDM works and the contribution of both parameters, where the warp range  $w$  constrains the set of possible mappings, whereas the  $c \times c$  context window computes the difference between the horizontal and vertical gradients for each mapping.

The algorithm requires each pixel in the query image to be mapped to a pixel within the reference image not more distant than  $w$  pixels from the place it would take in a linear matching. Over all these possible mappings, the best matching pixel is determined using the  $c \times c$  local gradient context window, by minimizing the difference with respect to the test image pixel. In particular, the IDM distance  $D$  between two symbols  $S_1$  (the query input) and  $S_2$  (the template) is defined as:

$$D^2 = \sum_{x,y} \min_{d_x,d_y} ||S_1(x + d_x, y + d_y) - S_2(x, y)||^2$$



**Figure 3. Example of areas affected by the comparison of pixels with IDM, where  $w = 3$  and  $c = 2$ . The query pixel context (indicated by the orange area in the query image) is compared with each equal-sized rectangle within the warping area (dark-green rectangle of the reference image).**

where  $d_x$  and  $d_y$  represent pixel shifts and  $S_i(x, y)$  represents the feature values in  $S_i$  from the patch centered at  $x, y$ .

**Itinerary module.** This module generates personalized tourist route on the basis of several requirements, which can be either of the user or the archaeologist. Basically, users parameters may be alphanumeric or geographical. As an example, one parameter is the available time to walk through the path. The calculated path also depends on the users trekking ability, which can be customized through the interface. Instead, the geographical parameters correspond to the places that the user wants to mandatorily visit. The latter might also be set by the archaeologists, which in some cases need to explore new regions, and/or want to invoke volunteers assistance.

The algorithm used to calculate personalized tourist route is an extension of the single optimal route computation algorithm for transit networks [22]. In particular, the extension concerns the generation of multiple routes for a given departure and arrival point [24]. The generation process consists of three steps. First, the system creates a list of recommended Points of Interest (POIs) by combining the tourist information with the user profile. Then, the routing algorithm combines this information with users constraints (available time, starting POI, trekking ability, and so on), POI data, and itinerary properties (slope degree, kind of track, and so on) to generate personalized routes. Finally, users have the opportunity to customize the proposed route to better fit their needs.

**Archaeologist Module.** The aim of this module is twofold: (i) manage petroglyphs, their information, descriptions, keywords, dating, and so on; (ii) analyze the user provided information. The latter is very important for the



research phase. Basically, PetroAdvisor aims to support archaeologists in the exploration of wide portions of territory by leveraging the volunteers community. In this sense, the module is the final phase of this exploration, when experts analyze the collected data and, in some way, make them available to tourists.

#### 4. Usage Scenario

In this section, we propose three user scenarios, aiming to show the main functionalities of the PetroAdvisor system. Basically, they are described by following a hypothetical mental action road, which starts from the itinerary selection, expects the insertion of a petroglyph, and ends with its check and validation.

This meta scenario is particularly significant for two reasons: it has been proposed by some archaeologists, meaning that the scenario satisfies the specific requirements in terms of data collection quantity and quality. In terms of quantity, because the system takes advantage of the tourists for exploring large parts of a territory in a short time, and in terms of quality, because every information must be checked and validated by an archaeologist.

The first scenario, also described in Figure 4, concerns with the itinerary selection. As previously stated in the functional requirements, the system should be able to propose new itineraries on the basis of information provided by the user and/or the archaeologists. The user can specify the trekking ability or the time available for walking through the itinerary. Instead, the archaeologist can specify the regions of interest like for example those not explored yet, or those that do not require further effort to be explored. S/he can also assign a score for indicating which petroglyphs should be visited. Such a score is mediated through the scores assigned by the users who visited them.

In what follows, we describe the scenario in Figure 4, which represents the user point of view. The first frame (Figure 4(a)) shows the map of the site indicating the petroglyph locations and the region of interest. Notice that petroglyphs are differently colored based on the score they achieved. Here, the user may see comments, descriptions, and any other information related to the petroglyphs. Based on such information, users may choose those to mandatorily see, those to optionally see, and those to absolutely avoid. The next step allows users to set own parameters (Figure 4(b)). Finally, the itinerary will be generated and submitted to the user, who can either accept or request the generation of a new one.

Figure 5 shows the second scenario, which corresponds to the main task a generic user should perform. After selecting an itinerary, like the one highlighted in the previous scenario, the user walk can start. During the excursion the user can see his/her position on the map, as shown in Figure



**Figure 4. The itinerary selection scenario. (a) petroglyph locations and the region of interest, (b) itinerary parameter settings, and (c) the generated itinerary (white colored line).**

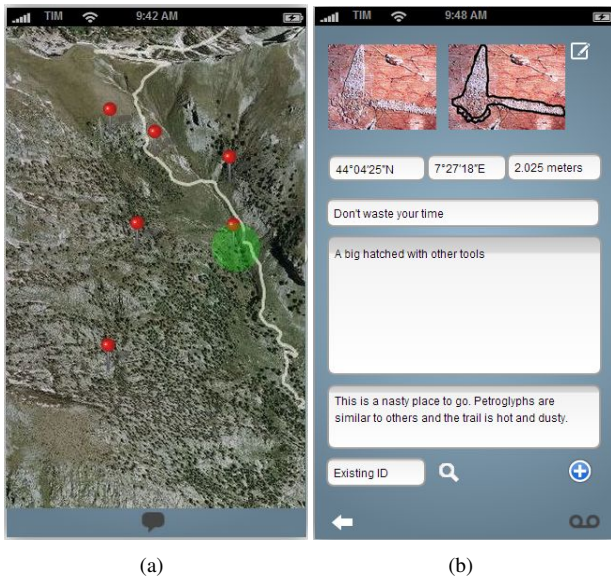
5(a), and tap the screen to see information about the reached petroglyphs. In case the user identifies a new petroglyph, or if s/he wants to add a description or comment to an existing one, s/he can access its properties. A workflow guides the user during the accomplishment of this task, where the first step is to add a new picture of the identified petroglyph (see Figure 5(b)), which can be accepted or rejected. Once completed this step, the user is enabled to insert the information: descriptions, comments, scores, and so on (Figure 5(c)), and the petroglyph contour (Figure 5(d)). After depicted the latter, the user can also search for similar petroglyphs, and look for ones elsewhere (Figure 5(e)). Finally, s/he can submit the form and the data to complete the task.

The final scenario is shown in Figure 6. It does not involve *Generic Users*, but *Archaeologists*. The latter can check and validate the user provided information. Initially, the archaeologist sees a map where the validated petroglyph and those to be validated are highlighted by using different colors. By clicking on each of them, the archaeologist accesses the mask where s/he can modify the submitted information, associate it to an existing petroglyph, or create a new one.

#### 5. Related Work

In the recent years, several systems involving human participation have been developed in the context of cultural heritage. CAPTCHA-ROCK is a system for helping archaeologists to extract data from petroglyph images [27]. In particular, it is a CAPTCHA system where users have to trace petroglyph images with the mouse pointer for access control purposes. The collected contours are exploited by a data mining algorithm for classifying petroglyphs. PetroAdvisor





**Figure 6. Petroglyph check and validation scenario. (a) map of petroglyph location highlighting those to be validated by archaeologists, (b) form through which the archaeologists can view and modify tourist provided information.**

provides a similar contour collection feature, but with an increased usability, since it is conceived for mobile devices, hence it provides touchscreen based interaction modality.

The Wicklow Rock Art Project<sup>6</sup> aims to explore the potential of photogrammetry in rock art recording, and to examine ways of protecting and promoting prehistoric open-air rock art in a sustainable fashion. Moreover, the project has the goal of creating a public engagement and interaction environment based on the crowdsourcing model. The latter should encourage people to identify with the rock art in their area, and cultivate a sense of guardianship and protection of such a fragile resource.

Heritage Crowd<sup>7</sup> is a project providing participants with tools for defining the cultural heritage of their place by using sms messages, voicemail, and other channels. In particular, through the use of a number of technologies, contributors have the possibility of creating a database of local history knowledge, which are accessible through a public website.

MicroPasts<sup>8</sup> is a community platform for conducting, designing, and funding research on human past [3]. In particular, it aims to provide an online space, where mixed groups of archaeologists and other volunteers collaborate

to produce innovative open datasets, develop new research projects on archaeology, history and heritage, and micro-fund those new collaborative projects via crowdfunding.

Ancient Lives<sup>9</sup> is a system that helps researchers transcribe Greco-Roman texts recovered from fragments of papyrus found in Egypt to better understand the periods culture. This is accomplished by involving people in transcribing items from the Oxyrhynchus Papyri and determining if they are parts of already known texts or if they are new texts. After transcriptions have been digitally collected, the system combines human and computer intelligence to identify known texts and documents. Similarly, the Transcribe Bushman<sup>10</sup> project aims at preserving the extinct *!xam* and endangered *!kun* languages of the Bushman people. People can contribute to the project by transcribing some pages of materials, which include rock art painting, drawings, and notebooks.

The Portable Antiquity Scheme<sup>11</sup> is a project to encourage volunteers in recording archaeological objects found in England and Wales. Every year many thousands of objects are discovered, many of which through metal-detectors, but also by people whilst out walking, gardening, or going about their daily work. As a consequence, the project has also the aim of stimulating public involvement and promoting best practice.

The Valley of the Khans<sup>12</sup> project aims to identify the site of Genghis Khans Tomb using noninvasive technologies ranging from aerial and satellite imaging, human computation, and non-invasive geophysical surveying. Users can join the research team by examining the satellite images and searching for clues guiding in the discovery of the lost tomb of Genghis Khan. Within a year, the participants created more than two million notes in about 6.000 km<sup>2</sup>, which include lakes, rivers, and potential archaeological sites.

In 2008 British Library started the BL Georeferencer<sup>13</sup> research project for geo-referencing historical maps [21]. In particular, volunteers scanned and georeferenced maps of the 17th, 18th, and 19th century from England and Wales. The results of this work led to the digitization and distribution of more maps via Internet [19]. Similarly, the New York Public Library's MapWarper project<sup>14</sup> aimed at correcting historical maps through an environment enabling volunteers to browse and correct old historic maps from the collection of the New York Public Library. Another relevant and successful collaborative geo-referencing project is eHarta<sup>15</sup>, which focuses on historical series maps of Roma-

<sup>6</sup><http://www.ahiddenview.com/>

<sup>7</sup><http://www.heritagecrowd.org/>

<sup>8</sup><http://micropasts.org/>

<sup>9</sup><http://ancientlives.org/>

<sup>10</sup>[http://boinc.cs.uct.ac.za/transcribe\\_bushman/](http://boinc.cs.uct.ac.za/transcribe_bushman/)

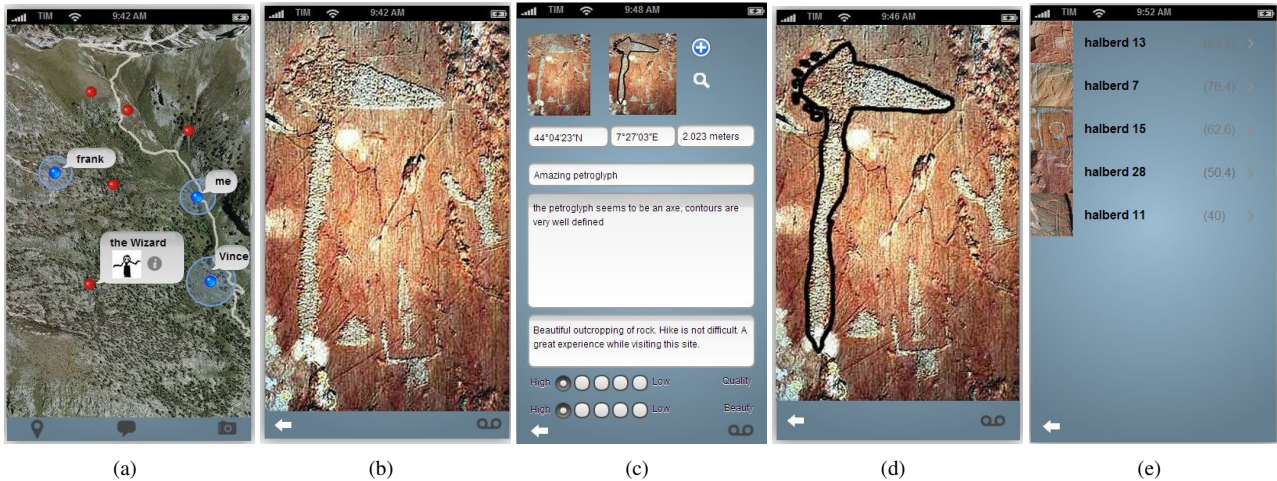
<sup>11</sup><http://finds.org.uk/>

<sup>12</sup><http://www.nationalgeographic.com/explorers/projects/valley-khans-project/>

<sup>13</sup><http://www.bl.uk/maps/index.html>

<sup>14</sup><http://maps.nypl.org/warper/>

<sup>15</sup><http://earth.unibuc.ro/articole/eHarta>



**Figure 5. Submission of petroglyph info scenario. (a) visualization of tourist info, (b) tourist picture of a candidate petroglyph, (c) form through which tourists enter the petroglyph info, (d) petroglyph contour traced by a tourist, (e) petroglyphs with a contour similar to the one highlighted in (d).**

nia [5].

MayaArch3D has the goal of creating virtual models of the Maya archaeological site of Copan in Honduras [25]. Among its research program objectives, there is the creation and visualization of three-dimensional models of cities and landscapes by volunteers provided information.

The HEIR<sup>16</sup> (Historic Environment Image Resource) project aims at digitalizing and keyword indexing old photos of monuments, landscapes, and environments taken across the world, and to re-photograph their modern settings. It is a crowdsourcing initiative for creating a world-wide accessible, interdisciplinary research resource, which can potentially provide a greater understanding of all aspects of the society and of the environment.

The above-mentioned projects have similar goals and objectives with respect to PetroAdvisor. In particular, all of them share the spirit of involving volunteers to raise awareness of cultural heritage, to increase identification of new archaeological information, and to bridge the gap in current technology. The originality of PetroAdvisor resides in the adoption of a rewarding strategy, which stimulates tourists to submit information to the system. In particular, the tourists are rewarded with information guiding them during their visits to petroglyph sites, such as visiting route based on user abilities.

## 6. Conclusion and Future Works

We have presented PetroAdvisor, a volunteer-based system supporting archaeologists in the digital preservation of

petroglyph sites. In particular, PetroAdvisor provides a centralized repository containing both archeologist and tourists provided information. The system also collects user provided petroglyph contours, which are particularly useful due to the limits of the current image recognition algorithms.

To stimulate tourist involvement, PetroAdvisor rewards tourists with information useful to visit archaeological sites, such as comments from previous petroglyph visitors, the recommendation of personalized tourist route, and archaeological information.

In the future we plan to perform massive experiments to evaluate the system usability, and the quality of the services it provides. Moreover, we would like to experimentally evaluate the usefulness of the whole approach from the archeologist point of view. Another important issue that deserves to be investigated is the detection of diverging comments on the same petroglyph, as those shown in Figure 5(c) and 6(b). Currently, we rely on the archeologist to accomplish this task. However, this task might result overwhelming for the number of available experts. Thus, other than devising volunteer-based strategies, it would be interesting investigating the natural-language processing approaches to partially automate this task.

In terms of implementation, we also plan to integrate the system within the IndianaMAS software platform so as to benefit from the availability of a repository storing a broad collection of petroglyph site and, access more sophisticated functionalities for searching and classifying petroglyphs.

Finally, we would like to add social network services in order to further stimulate system usage and active participation to information provisioning.

<sup>16</sup><http://www.arch.ox.ac.uk/HEIR.html>

## Acknowledgments

This research is supported by the “Indiana MAS and the Digital Preservation of Rock Carvings” FIRB project funded by the Italian Ministry of University and Research (MIUR): under grant RBFR10PEIT.

## References

- [1] N. Beale and G. Beale. The potential of open models for public archaeology. In *Proceedings of the Third Annual Digital Economy All Hands Conference*, 2012.
- [2] N. Bianchi. Mount bego: prehistoric rock carvings. *Adoranten*, pages 70–80, 2010.
- [3] C. Bonacchi, D. Pett, and A. Bevan. Developing ‘crowd and community-fuelled archaeological research’: methodological, technical and ethical challenges. In *Proceedings of Conference Computer Applications and Quantitative Methods in Archaeology*, 2014.
- [4] C. Chippindale and P. Taon, editors. *The archaeology of rock-art*. Cambridge University Press, 1998.
- [5] V. Crăciunescu, Ș. Constantinescu, I. Ovejanu, and I. Rus. Project eHarta: A collaborative initiative to digitally preserve and freely share old cartographic documents in Romania. *e-Perimetron*, 6(4):261–269, 2011.
- [6] H. de Lumley and A. Echassoux. The rock carvings of the chalcolithic and ancient bronze age from the Mont Bego area. The cosmogonic myths of the early metallurgic settlers in the Southern Alps. *L’Anthropologie*, 113(5P2):969–1004, 2009.
- [7] V. Deufemia and L. Paolino. Petroglyph recognition using self-organizing maps and fuzzy visual language parsing. In *Proceedings of IEEE International Conference on Tools with Artificial Intelligence (ICTAI’12)*, pages 852–859. IEEE, 2012.
- [8] V. Deufemia, L. Paolino, G. Tortora, A. Traverso, V. Mascardi, M. Ancona, M. Martelli, N. Bianchi, and H. de Lumley. Investigative analysis across documents and drawings: visual analytics for archaeologists. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 539–546. ACM, 2012.
- [9] A. Doan, R. Ramakrishnan, and A. Halevy. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4):86–96, 2011.
- [10] M. Dodge and R. Kitchin. Crowdsourced cartography: mapping experience and knowledge. *Environment and Planning A*, 45(1):19–36, 2013.
- [11] A. Echassoux and H. de Lumley. Rock carvings of the Pleiads in the sacred mont Bego mountain, Tende, Alpes-Maritimes, France. *Comptes Rendus Palévol*, 8(5):461–469, 2009.
- [12] M. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.
- [13] T. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.
- [14] J. Heinzelman and C. Waters. Special report: Crowdsourcing crisis information in disaster affected Haiti. Technical report, United States Institute of Peace, 2010.
- [15] C. S. Henshilwood, F. d’Errico, R. Yates, Z. Jacobs, C. Tribolo, G. Duller, N. Mercier, J. Sealy, H. Valladas, I. Watts, and A. Wintle. Emergence of modern human behavior: middle stone age engravings from South Africa. *Science*, 295:1278–1280, 2002.
- [16] J. Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.
- [17] N. Jennings, K. Sycara, and M. Wooldridge. A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, 1(1):7–38, 2006.
- [18] D. Keysers, T. Deselaers, C. Gollan, and H. Ney. Deformation models for image recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1422–1435, 2007.
- [19] K. C. Kowal and P. Pridal. Online georeferencing for libraries: the British Library implementation of georeferencer for spatial metadata enhancement and public engagement. *Journal of Map & Geography Libraries: Advances in Geospatial Information*, 8(3):276–289, 2012.
- [20] V. Mascardi, D. Briola, A. Locoro, D. Grignani, V. Deufemia, L. Paolino, N. Bianchi, H. de Lumley, D. Malafronte, and A. Ricciarelli. A holonic multi-agent system for sketch, image and text interpretation in the rock art domain. *International Journal of Innovative Computing, Information and Control*, 10(1):81–100, 2014.
- [21] P. Pridal and P. Zabicka. Tiles as an approach to on-line publishing of scanned old maps, vedute and other historical documents. *e-Perimetron*, 3(1):10–21, 2008.
- [22] H. Ruihong. A schedule-based pathfinding algorithm for transit networks using pattern first search. *Geoinformatica*, 11(2):269–285, 2007.
- [23] M. Seidl and C. Breiteneder. Automated petroglyph image segmentation with interactive classifier fusion. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*, pages 66:1–66:8. ACM, 2012.
- [24] G. Tumas and F. Ricci. Personalized mobile city transport advisory system. In W. Höpken, U. Gretzel, and R. Law, editors, *Proceedings of the International Conference Information and Communication Technologies in Tourism (EN-TER’09)*, pages 173–183. Springer, 2009.
- [25] J. von Schwerin, H. Richards-Rissetto, G. Agugiaro, F. Remondino, and G. Girardi. The MayaArch3D project: A 3D GIS web system for querying ancient architecture and landscapes. In *Proceedings of Digital Humanities Conference*, 2012.
- [26] D. Whitley. *Handbook of rock art research*. AltaMira Press, 2001.
- [27] Q. Zhu, X. Wang, E. Keogh, and S. Lee. An efficient and effective similarity measure to enable data mining of petroglyphs. *Data Mining and Knowledge Discovery*, 23(1):91–127, 2011.

# Combining personal diaries with territorial intelligence to empower diabetic patients

Monica Sebillio, Maurizio Tucci, Genny Tortora,  
Giuliana Vitiello

Department of Management and Information Technology  
Università di Salerno - Italy

Athula Ginige

School of Computing, Engineering & Math  
University of Western Sydney - Australia

Pasquale De Giovanni

School of Computer Science and Informatics,  
University College of Dublin, Ireland

**Abstract**— Information is today recognized as a major source of benefit/profit, for those who are able to properly create and manage it. With the advent of new computing, storing and networking technologies, transforming data into useful, ‘marketable’ information has become a major business goal. The myriads of data available today may be profitably aggregated into information, and effective information management is critical to gain competitive advantage. The healthcare domain makes no exception. Governments and healthcare companies are paying increasing attention to patient-centred care and to its positive effects on business metrics such as finances, quality, safety, satisfaction and market share. The term patient-centred care is referred to health care that respects and satisfies the preferences, needs and values of patients. Appropriate information sharing and communication is recognized to be one of the key factors for patient-centred care. In this paper, we propose a software framework for the development of special-purpose applications meant to improve care experience of diabetic patients while creating public value for services. This is achieved by a profitable combination of territorial knowledge with personal data and events.

**Keywords**— *spatio-temporal metadata collection; mobile application development; patient-centred services for patient empowerment*

## I. INTRODUCTION

### A. Premise

Information is today recognized as a major source of profit, for those who are able to properly create and manage it. In the last decades we have progressively assisted to a shift from the industrial age, where technology was the key to increase efficiency in production, to the capitalistic age, where investments towards the most profitable markets were considered as important as production rates, and lately to the modern information age, where information is the core of any production or business activity. With the advent of new computing, storing and networking technologies, transforming data into useful, ‘marketable’ information has indeed become a major business goal for many companies. The chain

Data → Information → Profit

expresses the shared idea that the myriads of data available today may be profitably aggregated into information, and that effective information management is critical to gaining competitive advantage.

In recent years, the explosive growth of mobile devices such as smartphones or tablets has radically changed the way people communicate, access, share and modify information. One of the key factors behind the success of these devices is represented by the wide availability of third-party applications that allow users to perform several different tasks, from checking emails on-the-go to remote management of the whole back-end of a company.

A great support for the development of innovative solutions derives from the combination of the unique features of mobile devices and the mobile operating systems that let developers fully exploit those features, including the growing number of sensors, that are usually not available on traditional personal computers. Some statistics about the increasing number of mobile applications that take advantage of the data coming from such sensors to accomplish their main tasks can be found in [1]. A concrete example of the use of sensors-related information is represented by the context aware applications that provide users with new types of services or user interfaces able to adapt to the ongoing situation.

### B. Motivation

According to Dey, "context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." [2]. The increasing number of sensors available on a mobile device facilitates the development of context-aware, user-adaptive solutions. However, probing or making sense of all the information that can be retrieved from the available sensors can be a challenging task [3]. Usually, only a small subset of such information is exploited for the development of new types of applications, also due to technical difficulties. In fact, in order to adapt contents to user's needs, some form of external middleware is often required and, more important, such information is

collected and used only when it is actually needed, hence affecting the overall performance.

Yet, there are a number of sources of user-generated data/information, whose potentials are today underestimated. These need to be analyzed and combined to be useful for a wider set of application fields.

### C. Contribution

In this paper, we propose a software framework that allows to leverage personal data and events captured by a smartphone for the development of (third-party) mobile applications supporting users' daily activities. This is achieved thanks to the continuous background collection of metadata produced by the hosting smartphone device and of those directly generated by user's interaction with the smartphone. Within the framework, developers are allowed to access, manage and combine spatial and temporal metadata in order to provide end-users with special purpose geographical and temporal information. A spatio-temporal database has been designed to collect, aggregate and manage both the metadata generated by user-performed activities and those captured through the sensors available on the smartphone device. The framework also provides a library, which includes high level methods by which third party developers can effectively exploit the data collected through the background module. Thus, for example, the metadata related to the camera settings could be exploited by a third-party advanced photo editing application. Similarly, the combination of the last visited shopping center with the list of IT products saved in the browser's favorites could be exploited by a recommender system, and so on.

## II. RELATED WORK

The importance that the contextual information and related metadata assume for the proper management and use of growing types of data is well recognized both in the scientific and in industrial communities and is well discussed in literature.

In this section we will focus only on that works which show how metadata content can be usefully exploited in software systems supporting everyday activities. The description of papers that analyze in detail actual metadata standards or describe how to efficiently structure a context-aware middleware is outside the purposes of our discussion.

All the papers discussed here share, some underlying ideas on the need to classify user-generated information reducing as much as possible tedious, error-prone and time-consuming operations like the manual insertion of labels. In particular, for what concerns the user generated contents (particularly multimedia files) there is a general agreement that, besides the acquisition of the actual multimedia object, also the greatest amount of contextual related metadata should be acquired. Such type of additional information can then be analyzed and used, as an example, to automatically add cataloging labels to the multimedia objects.

In [4] Graham et al., the authors recognize the importance of exploiting time information to automatically generate collection and summaries from a set of photos. They propose

two photo browsers for collections with thousands of time-stamped digital images, which exploit the timing information to structure the collections and to automatically generate meaningful summaries. Users are provided with multiple ways to navigate and view the structured collections. Having structured the set of images into clusters, various summarization schemas can be created. In addition such schemas can be specialized whenever additional metadata information become available (such as the location or in presence of a face recognition algorithm).

Davis and Sarvas [5] insist on the need to manage the growing number of media files produced by final users and address the use of metadata as a feasible solution. Moreover they recognize the need to exploit the spatio-temporal context and social community of media capture to deduce media content. They propose a client-server system that combines the features of a traditional camera phone with a remote web server. It gathers all the available information at the point of capture, and uses such metadata to find similar media previously annotated.

In [6] Lahti et al. again the importance of using metadata to describe the content of mobile data is recognized. The authors' proposal consists of a search engine that is able to analyze image and audio content, which supports two types of search methods. Both methods rely on an automatic metadata extraction done for new files.

A detailed analysis of the importance of annotating personal multimedia files with context-related information was also performed by Viana et al. [7]. They categorize research about multimedia annotations into context-based and content-based approaches. They propose a two-step method which leverages the collection of the largest amount of available information about user's context when a multimedia document is created. It then enriches that information through ontologies and semantic reasoning.

Also Kim et al. [8] deal with the problem of effectively managing the photo libraries stored on mobile devices. Even in this case it is recognized that the manual annotation of all the information needed for an efficient retrieval and management, is unfeasible mainly due to the huge amount of time required. They proposed solution is to automatically gather such information from the metadata directly stored on the mobile device. Therefore, they designed a mobile Android application, Photo Cube, which extracts several metadata from photos, combines them with mobile device metadata and uses them to improve their management and searching. The application provides also hierarchical search and browsing facilities using parameters such as address or date/time.

Lee et al. propose a mobile prototype which recommends 3 applications to best match the user's context [9]. An adaptive mobile interface is created, based on five parameters characterizing the user's context of use, namely time, location, weather, emotion and activities. The state of such variables and the history of past context information are input to a probabilistic learning and inference algorithm, which derives the 3 recommendations.



In [10], the metadata management issues are analyzed from different point of view. The authors observe that usually different mobile applications operate on the same aspects of the whole user context and manage the same types of data but they store information in private databases. As a direct result, this lack of interoperability represents a common source of information redundancy. Therefore they argue that on mobile platform there is the need of a greater interoperability at the data management level. Moreover, they observe that on modern mobile device a large amount of data usually owns also a spatial component. A greater interoperability at the data management level could result in a join criterion for several resources (CIT). Exploiting interoperability of data management is useful not only on the single device but also across devices. For example, suppose that a group of users shares information about a meeting; when a certain user add information about the meeting location, such data could be easily exchanged among the other participants. In order to address the interoperability issue, the authors present and discuss an architecture for interoperability between installed application, co-located devices and web application. Their approach is based on a central data repository on mobile devices that all applications use cooperatively.

#### REFERENCES

- [1] Holzinger, A., Geier, M., and Germanakos, P. 2012. On the Development of Smart Adaptive User Interfaces for Mobile e-Business Applications. In Proceedings of the International Conference of e-Business (Rome, Italy, July 24 - 27, 2012).
- [2] Dey, A.K. 2001. Understanding and Using Context. *Personal Ubiquitous Comput.* 5, 1.
- [3] David, L., Endler, M., Barbosa, S. D. J and Filho, J. V. 2011. Middleware Support for Context-aware Mobile Applications with Adaptive Multimodal User Interfaces. In Proceedings of the Fourth International Conference on Ubi-Media Computing (Sao Paulo, Brazil, July 3 - 4, 2011). pp. 106 - 111.
- [4] Graham, A., Garcia-Molina, H., Paepcke, A., and Winograd, T. 2002. Time as Essence for Photo Browsing Through Personal Digital Libraries. In Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries (Portland, Oregon, USA, July 13 - 17, 2002). pp. 326 - 335.
- [5] Davis, M., and Sarvas, R. 2004. Mobile Media Metadata for Mobile Imaging. In Proceedings of IEEE International Conference on Multimedia and Expo, pp. 1707 - 1710 (Vol. 3).
- [6] Lahti, T., Pietarila, P., Liu Yingfei, Pylvänäinen, T., Vuorinen, O. 2007. Audio and Image Browser for Mobile Devices. In Proceedings of International Conference on Mobile Data Management (Mannheim, Germany, May 1, 2007). 206 - 208.
- [7] Viana, W., Miron, A. D., Moisuc, B., Gensel, J., Villanova-Oliver, M., and Martin, H. 2010. Towards the Semantic and Context-aware Management of Mobile Multimedia. *Multimedia Tools and Applications*, 53, 2 (June 2011), 391-429. DOI=10.1007/s11042-010-0502-6
- [8] Kim, J., Lee, S., Won, JS., and Moon, YS. 2011. Photo Cube: An Automatic Management and Search for Photos Using Mobile Smartphones. In Proceedings of 9th IEEE International Conference on Dependable, Autonomic and Secure Computing (Sydney, NSW, December 12 - 14, 2011). 1228 - 1234. DOI=http://dx.doi.org/10.1109/DASC.2011.199
- [9] Lee, H., Choi, Y., and Kim, Y. 2011. An Adaptive User Interface based on Spatiotemporal Structure Learning. In Proceedings of IEEE Consumer Communication and Networking Conference (Las Vegas, NV, January 9 - 12, 2011). 923 - 927. DOI=http://dx.doi.org/10.1109/CCNC.2011.5766642
- [10] Brodt, A., Schiller, O., Sathish, S., and Mitschang, B. 2011. A Mobile Data Management Architecture for Interoperability of Resource and Context Data. In Proceedings of the 12th IEEE International Conference on Mobile Data Management (Lulea, Sweden, June 6 - 9, 2011). 168 - 173. DOI=http://dx.doi.org/10.1109/MDM.2011.81
- [11] Android Developers. 2014. <http://developer.android.com/index.html>
- [12] Meier, R. 2012. Professional Android 4 Application Development. John Wiley and Sons, Inc. ISBN: 978-1-118-10227-5



# Predicting Traffic Congestion in Presence of Planned Special Events

Simon Kwoczek  
Group Research  
Volkswagen AG

simon.kwoczek@volkswagen.de

Sergio Di Martino  
Group Research  
Volkswagen AG

sergio.di.martino@volkswagen.de

Wolfgang Nejdl  
L3S Research Center  
University of Hanover  
nejdl@l3s.de

## Abstract

*The recent availability of datasets on transportation networks with high spatial and temporal resolution is enabling new research activities in the fields of Territorial Intelligence and Smart Cities. Within these domains, in this paper we focus on the problem of predicting traffic congestion in urban environments caused by attendees leaving a Planned Special Events (PSE), such as a soccer game or a concert. The proposed approach consists of two steps. In the first one, we use the K-Nearest Neighbor algorithm to predict congestions within the vicinity of the venue (e.g. a Stadion) based on the knowledge from past observed events. In the second step, we identify the road segments that are likely to show congestion due to PSEs and map our prediction to these road segments. To visualize the traffic trends and congestion behavior we learned and to allow Domain Experts to evaluate the situation we also provide a Google Earth-based GUI. The proposed solution has been experimentally proven to outperform current state of the art solutions by about 35% and thus it can successfully serve to reliably predict congestions due to PSEs.*

**Keywords:** traffic prediction, planned special event, event analysis

## 1. Introduction

In times of ongoing urbanization and steady growth of mobility demands, traffic congestions cost billions of dollars to the society every year [12]. Within the last years, thanks to the advances in sensor technologies (like Smartphones, GPS handhelds, etc.) and storage capabilities, datasets about traffic with high spatial and temporal resolution have become available. This has led to advanced investigations on the impact of different influencing factors, as traffic lights, daily rush hour, construction zones, etc., on traffic congestions (i.e. [2, 7, 9]). Most of these factors are either recurring on a regular base (i.e. rush hour), exist only once for limited time (i.e. construction zones) or

their occurrences are unpredictable (i.e. accidents). Current state of the art commercial solutions are able to work well with recurring traffic situations as it's behavior can easily be learned from historical data [18]. On the other hand, also non-periodic events with an expected large attendance (known also as Planned Special Events, or PSE as introduced in [6]), such as concerts, soccer games, etc., play a major role for delays in everyday transportation [9]. As example, the concert of Rihanna in Johannesburg (South Africa) in October 2013 caused people to sit in traffic for as long as five hours, trying to reach the stadium. Similarly, the concert of Robbie Williams in London, in 2003, caused tailbacks up to 10 miles on the highway A1 towards the stadium. An interesting aspect is that the traffic due to PSEs has a quite typical behavior, having two subsequent waves of congestion [10]. The first one is caused by people going to the event, while the second one is due to people leaving the venue, and may be even bigger than the first wave. Very few research attention has been devoted to predict the congestion due to a PSE. At the same time, even the most advanced available commercial systems are incapable of predicting this kind of non-recurring traffic ahead of time.

To address this open issue, in this paper we describe a solution we developed to predict the spatio-temporal impact of the second wave of traffic due to a PSE around its venue. In particular, the proposed approach is meant to be executed while the event is happening, and takes as input the category of the PSE, like Concert, Entertainment, etc., and the information on the first wave of traffic, coming from traditional traffic providers. Then, using an adaptation of the K-Nearest Neighbors (K-NN) algorithm, we look for the most similar past PSEs (cases) among historic observations, in order to derive a prediction of the impact of the second wave of traffic. Such a prediction is done in terms of average delay over the road segments around the venue that have been found to be highly correlated with the congestions due to PSEs. To graphically visualize the results, we provide a Google Earth-based tool <sup>1</sup>, recalling the one de-

<sup>1</sup><http://earth.google.com>

veloped in [4] or in [13]. Thanks to this tool, it is possible for Decision Makers to understand how the traffic situation is influenced by a PSE.

To assess the proposed approach, we used data about traffic and events from June to December 2013 in the inner city of Cologne, Germany. The traffic data has been provided by one of the most prominent traffic providers in Europe. It comes mainly from Floating Car Data, i.e. data collected from GPS sensors in vehicles, and it covers most of the streets within the inner city. As for the PSEs, we considered all the 29 events hosted in the Cologne LANXESS arena, in the same temporal span. By performing a leave-one-out cross validation on the event dataset, we compared our proposal with current state of the art, intended as real time traffic informations, and with a baseline consisting simply of replicating the impact of the first wave as predicted second wave. Results show that our proposals outperforms both the alternative solutions, providing prediction that are better up to 41%.

The remainder of the paper is structured as follows: section 2 presents related work within the field of traffic predictions and PSEs and some background terminology. In section 3 we present the approach to predict the congestions caused by outbound traffic after an event, plus the GUI to visualize the results. In section 4 we describe the Research Questions and the evaluation protocol to assess the proposed approach. In section 5 the results of the empirical assessment are presented and discussed. Finally in section 6 conclusions are outlined, together with some future research directions.

## 2. Background and Definitions

Since years, traffic congestion predictions have been widely studied within the research communities of ITS, Smart Cities and Territorial Intelligence, leading to a rich body of literature on these topics. Indeed, by knowing in advance the traffic patterns it is possible to optimize mobility. For instance, route calculation engines can compute more energy efficient routes, able also to save time for the drivers.

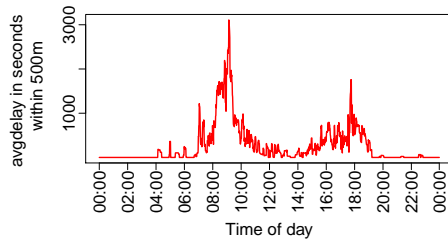
In general, predicting traffic congestion in urban environments is a highly complex task. Early approaches for traffic predictions used simulations and theoretical modeling (e.g. [2, 3]). Nowadays thanks to the availability of new massive datasets on traffic, several different statistic and data driven approaches have been presented to the community. Examples include generalized linear regression ([21]), nonlinear time series ([8]), Kalman filters ([11]), support vector regression ([20]), and various neural network models ([11, 17, 19]). A combination of some of the latter kind of approaches is used also by current commercial navigation solutions, able to predict recurring congestions by identi-

fying characteristic traffic flow patterns for street segments from historical data. On top of that, these commercial systems can also optimize the route planning based on the real-time traffic situation [18]. In general, traffic congestion can be divided into recurring congestions, usually caused due to a mobility demand that exceeds the capacity of the road network (e.g. due to rush hour), and congestions that are non-recurring (e.g. due to incidents or special events) [9]. The effects of nonrecurring traffic congestions and their prediction is a widely investigated topic within the research community (e.g. [14–16]). Although these approaches showed a significant improvement in prediction they use data from stationary loop sensors that are not always capable of reflecting the traffic state in a granularity required for urban scenarios. In addition, their focus lies on one-directional street segments as highways whereby usually in the inner cities the impact is a multidimensional problem, evolving in a 2D, more complex route network.

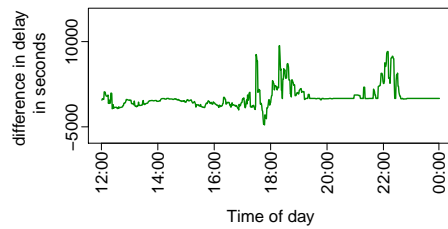
Previous researches have highlighted that also PSEs are a possible influencing factors [7, 9], since they may lead thousands (or even hundreds of thousands) of people to travel towards the same destination in a very limited time interval, and then to leave the venue again in a very short time span. To the best of our knowledge, the only work available that has its focus on the influence of PSEs on traffic is presented in [10]. The authors report a generic overview about the influence of PSEs on the road network, derived from an event classification defined by the Chinese State Council. They also introduce management plans for the different types of events, but there is no quantifiable solution for the prediction of the traffic.

### 2.1. Definitions

A PSE may have an impact on the traffic behavior in a specific region over time. Such an *impact region* can be defined as the list of congested segments of the road network. A *congested segment* is a piece of road network where the difference between expected and actual traffic speed is bigger than a certain threshold. The expected traffic velocity on a given segment can be obtained from a number of historical observations, coming from special sensors in the infrastructure and/or Floating Car Data. In this way it is possible to learn the typical traffic behavior for a given road segment on a certain day of the week, at a given time. If a segment is congested it takes more time to pass by it. This additional time is defined as *delay time* and is measured in seconds. As an example, Figure 1 shows the summed up delay time in a circular impact region, with a radius of 500 meters around the LANXESS Arena in Cologne, Germany for all Tuesdays in our considered dataset, where no events happened. The red line presents the model that is derived from historical observations as the average delay time on all Tues-



**Figure 1. Average historic congestion around the LANXESS arena on Tuesdays**



**Figure 2. Additional delay due to the concert of Mark Knopfler on top of the historic trend line for Tuesdays**

days. From this *trend line* it is possible to detect the rush hour behavior in the morning and in the evening. As stated above, the graph only contains Tuesdays without events and the generated trend line can therefore be seen as the regular pattern around the venue on Tuesdays. As described in the Introduction, a PSE leads to two waves of traffic. An example can be seen in figure 2. It shows the traffic behavior around the LANXESS Arena on Tuesday the 2<sup>nd</sup> of July 2013, when Mark Knopfler gave a concert in the arena. The figure shows the difference between the observed delay time at that day and the trend line for that day of the week. From the figure we can clearly see the two peaks of congestion, between 18:00 and 20:00, and between 22:00 and 23:00. Since the concert started at 20:00 and ended around 22:00, the assumption arises that the two peaks are caused by people going to the event (*inbound traffic*) and leaving after its end (*outbound traffic*). To quantify the impact of PSEs on traffic, we focused on the traffic data happening before and after the events. From the 29 events we analyzed, and for this specific location, we found out that the inbound traffic is always contained in a time frame within two hours before the scheduled begin of the event. As for the outbound traffic, in our dataset the time frame starts two hour after the begin of the event and lasts for two hours. These time frames ensure that the entire event-caused traffic is captured. For both time frames, we define a normalized timescale, using the difference to the planned begin of the event as scale. In

this way, we can compare traffic situations happening in the same relative time corridor, for events that start at different times. For both time frames we observe the average delay time, defined as *avgdelay* (measured in seconds) as measure for the impact of the PSE on traffic.

### 3. The Proposed Approach

In this section we define two different approaches to predict the outbound traffic, namely the Category-Based Modeling Approach (CBMA) and the Category Specific Inbound-Based Prediction Approach (CIPA). Afterwards, we introduce a tool to visualize the results of the predictions.

#### 3.1. Impact of event category

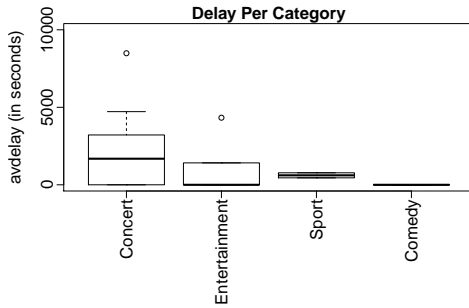
The LANXESS arena hosts different events of different categories. During the time span of our study, we observed 18 *concerts*, 6 *entertainment* events, 3 *comedy* events and 2 *sporting* events. The naming of the categories is taken from the official arena schedule.

Although most of the considered events started roughly at the same time (between 19:00 and 21:00) the different categories show significant differences in their impact on traffic. In total, 12 out of the 18 concerts showed significant inbound congestions and 11 of these events showed also strong outbound congestions. While comedy events did not account for any significant raise in congestions, entertainment events caused congestions in 2 out of 6 times for the inbound and 1 time in the outbound traffic period. The only events that differed significantly in their start time were the sporting events (in the time span of our research the sporting events were two handball games). Both of them showed a comparable congestion behaviour for in- and outbound traffic. To further illustrate the correlation between event category and congestion behaviour, we report in figure 3 the behavior for each category using the *avgdelay* observed as congestion measure. The boxplot highlights the different behavior of the different event categories.

These differences motivated us to develop a first approach that focuses on the typical congestion behavior for different categories that we named *Category-Based Modeling Approach (CBMA)*.

#### 3.2. Category-Based Modeling Approach (CBMA)

The CBMA is aimed at detecting the specific congestion patterns of each observed category and uses these patterns for predictions. We learn the behavior of each category by accumulating the congestion behavior during the outbound traffic for all events of that category in our training set. From that, we consider the average of all observations and



**Figure 3. Outbound congestion behavior for the different observed categories**

use the generated trend line for predicting the outbound traffic congestion behavior for all events in the test set. Since the only information CMBA uses for predictions is the category of the events, it can be used without any knowledge about the traffic situation on the specific day of the targeted event. Thus, it can be applied early ahead of time.

### 3.3. Impact of Inbound Traffic Congestion

While each category of events shows a different pattern, some events of the same category also vary a lot in their congestion behavior. For example, the Rihanna concert on the 26<sup>th</sup> of June 2013 and the concert of Barbra Streisand on the 12<sup>th</sup> of June 2013 are both of the same category (concert) and start almost at the same time. Thus, one would expect that they have a similar impact on traffic. However while the outbound congestion after the Rihanna concert shows an *avgdelay* of around 4600 seconds, the outbound congestion after the Barbra Streisand concert only shows an *avgdelay* of 2100 seconds. A likely explanation for the phenomenon lies in the different amount of people going to these two concerts. Although the actual number of attendees is not available, we were willing to explore if the inbound congestion of an event is a good indicator of how severe the outbound congestion is going to be. In other words, we expect the congestion during the inbound traffic to describe the "hype" about the event by serving as an indicator about how many people are going to the event. This motivated us to develop a second approach, named *Category Specific Inbound-Based Prediction Approach (CIPA)* that includes this information into the prediction model.

### 3.4. Category Specific Inbound-Based Prediction Approach (CIPA)

The CIPA is based on the idea that both category and inbound traffic are explanatory variables for the expected outbound traffic. One possible way to incorporate these in-

formation into the prediction process lies in applying Machine Learning techniques on an adequate dataset of historic data to automatically learn the correlations between these variables. For CIPA, the dataset contains information about the *category* of an event and the observed *avgdelay* during the inbound traffic. We use these features as predictors of the *avgdelay* during the outbound traffic. In our case, we applied the K-Nearest Neighbor regression [1] Machine Learning technique. In order to do a prediction, this technique aggregates the values the *k* "closest" examples in the training set, where *k* is an input parameter to the algorithm. To compute the distance among observations we used the Euclidean distance function, while as aggregation formula, we used the mean of the observed values. Further details on K-NN can be found in [1]. A drawback of this approach compared to CBMA lies in the fact, that it can only be used for predictions on the event day, after the inbound traffic has already been observed.

### 3.5. Mapping Delay on the Road Network

In order to use the generated predictions in Advanced Driver Information Systems, the resulting delay time needs to be mapped onto the street network. We identify road segments that are likely of being congested due to outbound traffic from historic data by selecting all segments that were affected in at least 1/3rd of all PSE caused congestions. Then we spread the predicted outbound delay time for an PSE over the segments, assuming a normal distribution. From the results, we calculate the delay per meter *DM* index (in seconds) for that specific event, which gives an overview about the severity of congestion in the area. The index is also used in the GUI we describe in the next section, to allow Domain Experts to get an overview about the expected impact of the PSE.

### 3.6. The proposed interactive GUI

Once the results are produced by the one of the previous prediction engines, they can feed a visualization layer. To this aim, we have developed a Graphical User Interface, shown in figure 4, that embeds Google Earth to render in 3D the predicted spatio-temporal impact area, over a geo-referenced satellite image, optionally also enhanced by additional informative layers. The goal is to facilitate Domain Experts and Decision Makers to visually understand spatial relationships among the datasets and their spatial context. The proposed GUI contains a main frame that holds the current representation of the predicted traffic severity level in Google Earth. On the left side, the user can pick a location from a list of available data. After selecting a location, the set of scheduled PSEs for the venue is shown in the left panel. If the information is available, the inbound traffic of

the event will be represented in a graphic way at the bottom of the left panel. By selecting one of the developed algorithms (CBMA or CIPA) the prediction process starts and the user gets a visual representation of the predicted congestion area as a new layer on the map. The severity level can be manually adjusted by the settings tab, allowing Domain Experts to easily customize the representation of the results.

## 4. Experiments

This section describes the setup of the experiments conducted to assess the validity of our proposals.

### 4.1. Dataset Description

The traffic information used in this research is collected by one of the most prominent traffic providers all over Europe. It covers the main road network within the inner city of Cologne. The information is generated from a combination of various different sources. These include Floating Car Data (originated from millions of vehicles equipped with GPS sensors), GSM probe data and data from stationary sensor (e.g. loop detectors, camera sensors) obtained from local traffic management centers. More information about the different sources and their aggregation can be found in [18]. From these combined sources, an accurate description of detected traffic congestions can be derived. Specifically, for each detected congestion, the data contains the resulting *delay time* on a segment of the road network (in seconds) and the *congestion level*, on a scale from 1 to 5, where 1 means no relevant traffic congestion, 4 is the highest level and 5 means that the congestion level is unknown. We received data from June the 1<sup>st</sup> 2013 until December the 31<sup>st</sup>, 2013. The dataset we used in this research accounts for about 50 Gigabytes which leads to a favorable coverage of the area around the arena in Cologne. All measurements presented in this section show the accumulated values of all road segments within an area of 500 meters around the stadium. As for the events, we collected all the PSEs happening in the LANXESS arena in the same time span covered by the above described traffic information. In these seven months, 29 events were scheduled. Regarding these observed events, 16 showed congestion due to inbound traffic. For these events, the observed an accumulated *avgdelay* on the considered road segments varied from very minor with 420 seconds (Andreas Gabalier, 17th of October 2013) to severe traffic congestion with accumulated 9422 seconds (Rihanna, 27th of June 2013), that is a delay of almost 3 hours in the considered area. The outbound traffic caused congestion in 17 of the 29 events whereas the *avgdelay* varied from 124 seconds (Cirque du Soleil, 25th of October 2013) up to 4632 seconds (Rihanna, 26th of June 2013).

### 4.2. Baseline Approaches

As stated before, to the best of our knowledge, there is no research that has studied the impact of PSEs on traffic in a similar manner before. Therefore, we introduce two approaches that serve as baseline for our proposed CBMA and CIPA, named the *Zero approach* and the *Start approach*.

**Zero approach:** In the Zero approach, we simply assume that there is no traffic in the area during outbound traffic. While this approach seems fairly simplistic, it actually reflects the current state of the art in navigation solutions, and is therefore suitable for comparing newly developed algorithms against it. As matter of fact, by relying just on historic data, current navigation solutions are unable to predict congestions due to nonrecurring events.

**Start approach:** While the Zero approach simply assumes no additional traffic at all above the historic model, the Start approach goes one step further. It is based on the idea that the amount of people going to the event is the same as the people leaving the event. Consequently, the outbound traffic could be estimated as the same as the inbound traffic. Thus, we could use the exact information about the inbound *avgdelay* as a prediction for the outbound *avgdelay*.

### 4.3. Research Questions (RQ)

Given the above described approaches, the research questions underlying this paper can be formulated as follows:

1. Are the predictions of expected delay due to outbound traffic using CBMA better than those obtained with the Zero and the Start approaches?
2. Are the predictions of expected delay due to outbound traffic using CIPA better than those obtained with the Zero and the Start approaches?
3. Are the predictions of expected delay due to outbound traffic using CIPA better than those obtained with CBMA?

For these research questions, the performance of the different approaches is measured using the root mean square error (RMSE) defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (1)$$

where  $y_i$  and  $\hat{y}_i$  represent the actual and predicted delay time respectively and  $N$  is the number of predictions.

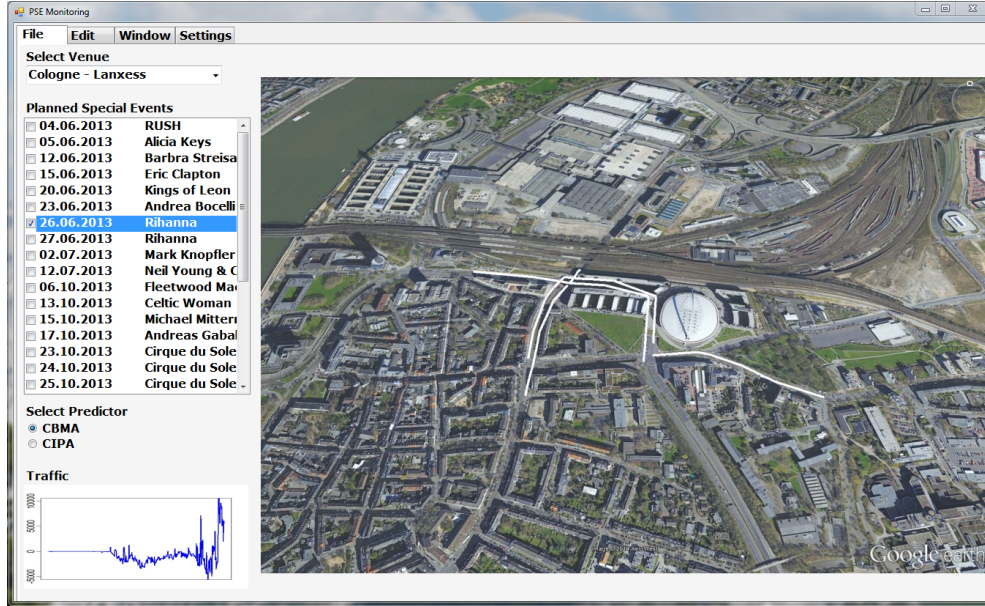


Figure 4. The Google Earth-based GUI for the visualization of the results

## 5. Results

To evaluate the validity of the proposed approaches, and thus to answer the Research Questions, we applied the four solutions described in the previous section on the dataset of 29 events by means of a cross-validation, that is splitting the data set into training and validation sets. Training sets are used to build models, while validation sets are used to validate the obtained prediction models. In particular, we exploited a leave-one-out cross validation, which means that the original data set is divided into  $n=29$  different subsets of training and validation sets, where each validation set has just one event.

As for the use of the K-Nearest Neighbors algorithm, one of the issue is the optimal choice of the parameter  $k$ , that is the number of closest training examples to be considered in the feature space. This choice usually depends upon the data. It is outside the scope of this research to investigate hyperparameter optimization techniques (e.g. as in [5]). In our case we investigated the use of  $k=3$  and  $k=5$ , obtaining very similar results. Consequently, in the following we report only on the use of  $k=5$ .

Figure 5 shows the RSMEs of the predicted outbound traffic *avgdelay* for each approach separately. The figure shows, for the entire dataset of PSEs, the prediction performance for each method in terms of the RMSE values obtained from the cross validation. The data shows, that the baseline which uses the *avgdelay* from the inbound traffic as predictor has the highest RMSE of 2095 seconds. The Zero approach, which has no a-priori information about the

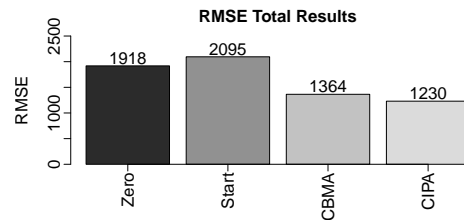


Figure 5. RMSE of the four approaches

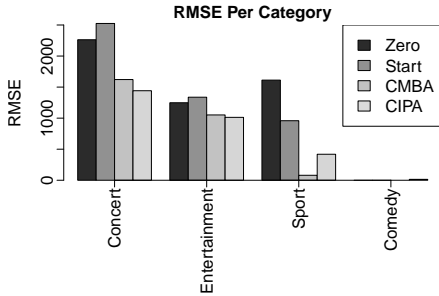
PSEs or the traffic information at all, shows an RMSE of 1918 and therefore slightly outperforms the Start approach.

To answer the RQ1, we can see that the RMSE for the CBMA is 1364, outperforming both the Zero- and the Start approach by far. Indeed, the proposed approach has a predicted delay error that is 28% lower than current state of the art, and 35% lower than the Start approach. Consequently, we can positively answer the RQ1, since CMBA outperforms the two baselines.

As for RQ2, the RMSE for the CIPA, which incorporates the most information about the situation, is 1230. This reflects a reduction of the prediction error of around 41% percent compared to the Start approach and still 35% percent compared to the Zero approach. Consequently, we can positively answer the RQ2, since CIPA outperforms the two baselines.

As for the RQ3, we can see from 5 that CBMA and CIPA provide quite similar results. This motivated us to further investigate the performances, splitting the results for each





**Figure 6. RMSE per category of events**

category of events. The results are shown in figure 6. We found that for very homogeneous categories (as for *Sport* and *Comedy* in our examples), the CIPA was outperformed by the CBMA. For instance, CBMA showed an RMSE of 80 seconds for the *Sport* category while the RMSA for CIPA is 418 seconds. At the same time, for high fluctuating categories (as for the categories *Concert* and *Entertainment* in our examples), the CIPA shows much better results of for instance an RMSE of 1441 seconds for the concert category, compared to 1621 using CBMA. Consequently we cannot provide a definitive answer to RQ3, since CIPA and CBMA have comparable performances.

### 5.1. Mapping results on the Road Network

We identified a set of 52 street segments that showed high correlation of the congestion with PSEs, which account for a total distance of 2105 meters. For these links, calculating the delay per meter index *DM* shows the different severity levels of the events. As an example, the CIPA predicted a delay time of 2129 seconds during the Mark Knopfler concert, that, given the total length of the considered road segments, leads to a severity index of around 1s/m. In a future integration with a route calculation engine, this means that the travel time of each route passing in the vicinity of the arena, in the considered timeframe, takes about 100 additional seconds per 100 meters of the route. Consequently, the route calculation engine has now new travel times for each segment, and can reroute the driver on the most efficient path.

## 6. Conclusion and Future Work

Improving mobility is a paramount issue due to the huge environmental and social impact of vehicular traffic. Thanks to the recent availability of high quality spatio-temporal datasets coming from Floating Car Data and other data sources, new and unprecedented research opportunities are arising in the direction of smarter solutions for mobility. Within this topic, to the best of our knowledge, a

very limited attention has been dedicated to the prediction of the impact of Planned Special Events, such as concerts or sport games, on urban traffic, even if some researches highlight that they are a significant fraction of the total traffic. PSEs have a very specific impact on mobility, inducing two waves of traffic. The first one is due to the incoming spectators while the second one is due to the people leaving the event. In this paper we have proposed an approach to predict the impact of the second wave of traffic due a PSE, given information on its category and on the first wave of traffic. To this aim, we have defined a two-step solution, where at first we created models able to predict the out-bound congestion based on historic data about past events and, in the second step, mapped these congestion prediction on to the route network by identifying the road segments that are most likely of being congested due to PSEs. The approach is complemented by a GUI that allows Decision Makers to visualize on an interactive 3D map the dynamic extension of the impact area and the severity of the expected congestions. An empirical assessment has been done using floating car data covering 7 months of mobility information in the city of Cologne, Germany, and using all the PSEs hosted in the LANXESS arena in the same period.

The results show that our proposal can positively identify the influence of the PSE on the traffic, being up to 35% better than actual state of the art solutions. Consequently it can be a viable module to be integrate in any navigation device, in order to reroute drivers away from the location of the event and optimize urban mobility.

As future work, the next obvious step is to predict also the first wave of traffic, but our experiments showed that relying only on the time schedule is not sufficient, since there is no way to predict the amount of people that will attend the event. The idea we are starting to investigate is to use social networks, such as Facebook or Twitter, to understand what is the popularity of a specific event or artist, considering also the geographical and age distribution of the supporters.

## 7. Acknowledgment

The research leading to these results has been partly funded by the European Communitys Seventh Framework Programme (FP7/2007-2013) under grant agreement n 610990 - Project COMPANION. The authors wish also to thank TomTom NV for providing the traffic data used in this research.

## References

- [1] N. S. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

- [2] R. Chrobok, O. Kaumann, J. Wahle, and M. Schreckenberg. Different methods of traffic forecast based on real data. *European Journal of Operational Research*, 155(3):558–568, 2004.
- [3] S. Clark. Traffic prediction using multivariate non-parametric regression. *Journal of transportation engineering*, 129(2):161–168, 2003.
- [4] P. Compieta, S. Di Martino, M. Bertolotto, F. Ferrucci, and T. Kechadi. Exploratory spatio-temporal data mining and visualization. *Journal of Visual Languages and Computing*, 18(3):255 – 279, 2007. Visual Languages and Techniques for Human-GIS Interaction.
- [5] A. Corazza, S. Di Martino, F. Ferrucci, C. Gravino, F. Sarro, and E. Mendes. Using tabu search to configure support vector regression for effort estimation. *Empirical Software Engineering*, 18(3):506–546, 2013.
- [6] W. M. Dunn and S. P. Latoski. Managing travel for planned special events. In *Institute of Transportation Engineers (ITE) 2003 Technical Conference and Exhibit*, number CD-020, 2003.
- [7] E. Horvitz, J. Apacible, R. Sarin, and L. Liao. Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service. In *In Twenty-First Conference on Uncertainty in Artificial Intelligence*, 2005.
- [8] S. Ishak and H. Al-Deek. Performance evaluation of short-term time-series traffic prediction model. *Journal of Transportation Engineering*, 128(6):490–498, 2002.
- [9] J. Kwon, M. Mauch, and P. Varaiya. Components of congestion: Delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. *Transportation Research Record: Journal of the Transportation Research Board*, 1959(1):84–91, 2006.
- [10] D. Lei-lei, G. Jin-gang, S. Zheng-liang, and Q. Hong-tong. Study on traffic organization and management strategies for large special events. In *System Science and Engineering (ICSSE), 2012 International Conference on*, pages 432–436, June 2012.
- [11] J. W. C. Van Lint. Online learning solutions for freeway travel time prediction. *Intelligent Transportation Systems, IEEE Transactions on*, 9(1):38–47, March 2008.
- [12] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. Big data: The next frontier for innovation, competition, and productivity. Technical report, McKinsey Global Institute, 2011.
- [13] S. Di Martino, S. Bimonte, M. Bertolotto, and F. Ferrucci. Integrating google earth within olap tools for multidimensional exploration and analysis of spatial data. In *Enterprise Information Systems*, volume 24 of *Lecture Notes in Business Information Processing*, pages 940–951. Springer Berlin Heidelberg, 2009.
- [14] M. Miller and C. Gupta. Mining traffic incidents to forecast impact. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pages 33–40. ACM, 2012.
- [15] B. Pan, U. Demiryurek, and C. Shahabi. Utilizing real-world transportation data for accurate traffic prediction. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 595–604. IEEE, 2012.
- [16] B. Pan, U. Demiryurek, C. Shahabi, and C. Gupta. Forecasting spatiotemporal impact of traffic incidents on road networks. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, 2013.
- [17] D. Park, L. Rilett, and G. Han. Spectral basis neural networks for real-time travel time forecasting. *Journal of Transportation Engineering*, 125(6):515–523, 1999.
- [18] TomTom International. White paper - how tomtoms hd traffic and iq routes data provides the very best routing. Technical report, TomTom International, 2009.
- [19] L. Vanajakshi and L.R. Rilett. A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed. In *Intelligent Vehicles Symposium, 2004 IEEE*, pages 194–199, June 2004.
- [20] C. Wu, J. Ho, and D.T. Lee. Travel-time prediction with support vector regression. *Intelligent Transportation Systems, IEEE Transactions on*, 5(4):276–281, Dec 2004.
- [21] X. Zhang and J. A. Rice. Short-term travel time prediction. *Transportation Research Part C: Emerging Technologies*, 11(34):187 – 210, 2003. Traffic Detection and Estimation.

# Representing Uncertainty in Visual Integration

Bilal Berjawi  
LIRIS, UMR5205  
INSA de Lyon  
Lyon, France  
bilal.berjawi@liris.cnrs.fr

Élisabeth Chesneau  
EVS, UMR5600  
UJM  
Saint Etienne, France  
elisabeth.chesneau@univ-st-etienne.fr

Fabien Duchateau  
LIRIS, UMR5205  
UCBL  
Lyon, France  
fabien.duchateau@liris.cnrs.fr

Franck Favetta  
LIRIS, UMR5205  
ENSNP  
Lyon, France  
franck.favetta@liris.cnrs.fr

Claire Cuntz  
EVS, UMR5600  
Université Lumière Lyon2  
Lyon, France  
claire.cuntz@univ-lyon2.fr

Maryvonne Miquel, Robert Laurini  
LIRIS, UMR5205  
INSA de Lyon  
Lyon, France  
firstname.lastname@liris.cnrs.fr

**Abstract**—Multiple cartographic providers propose services displaying points of interests (POI) on maps. However, the provided POIs are often incomplete and contradictory from one provider to another. Previous works proposed solutions for detecting correspondences between spatial entities that refer to the same geographic object. Although one can visualize the result of the integration of corresponding entities, users do not have any information about the quality of this integration. In this paper, we propose a solution to visualize the uncertainty inherent to a spatial integration algorithm. We present an integration process that identifies three degrees of confidence for spatial and terminological integration results. A prototype has been implemented to present the benefits of our proposal in an use-case scenario. This work has been realized within the framework of UNIMAP<sup>1</sup> project.

**Keywords:** *Visual uncertainty; Spatial integration*

## I. INTRODUCTION

Location-based services (LBS) are daily used in various applications, and cartographic providers play an essential role in displaying points of interest (POI) such as restaurants, hotels, and tourist places. A POI can be defined as a geographic object that has a point geometric shape. A POI has spatial attributes longitude and latitude, and terminological (non-spatial) attributes such as name and type (e.g., restaurant, hotel). Some providers may supply additional terminological attributes such as address, phone number, Web site, customers' ratings, etc. A provider usually represents a POI on a map with a specific symbol or icon. Due to lack of completeness, noisy, inaccurate and contradictory data, it is interesting to propose solutions for detecting corresponding entities (i.e., which refer to the same POI) from different providers. This challenge aims at improving the quality and the relevance of information, which has a significant impact in tourist applications.

The integration of spatial information issued from different sources has been studied [10]. Earlier works so called "map

conflation" were specifically devoted to vector objects such as roads [23]. In the last decade, the integration problem mainly refers to the "entity matching" research domain, enhanced by a spatial aspect. The discovery of corresponding entities is performed either by exploiting only spatial information [26] or by computing and combining terminological similarities for selected attributes (e.g., name, type) [22]. Machine learning algorithms may be applied for tuning the parameters (e.g., weights) of a matching process [28]. When corresponding entities have been detected, an interesting use case aims at displaying a merged entity, i.e., to use a crafted algorithm to fusion the attributes' values of these corresponding entities. Such merging algorithms are not 100% confident. For instance, two corresponding entities may have a different location and the algorithm needs to determine the correct position. Similarly, the names or the phone numbers of two corresponding entities may differ, and the choice of the correct values relies on the merging algorithm. A merged entity may therefore include at different levels some uncertainties, which have to be presented to end-users [19].

In this paper, we are interested in visualizing the uncertainty resulting from the merging process of spatial entities. Our contributions can be summarized as follows: (i) identifying the "dimensions" which have to be taken into account for uncertainty, i.e., the POI type, the spatial attributes and the terminological (non-spatial) attributes; (ii) measuring the confidence level for each dimension as well as a global confidence score; (iii) modeling the visualization of a merged entity and its uncertainty; and (iv) implementing a prototype to demonstrate in a scenario the benefits for end-users.

The next section describes the related work, both in spatial integration and visualization. Section III provides a detailed explanation of our solution to represent and visualize various criteria about a merged entity. In Section IV, we demonstrate the benefits of our approach in a scenario, and we conclude in Section V.

<sup>1</sup> UNIMAP: <http://liris.cnrs.fr/unimap> (July 2014)

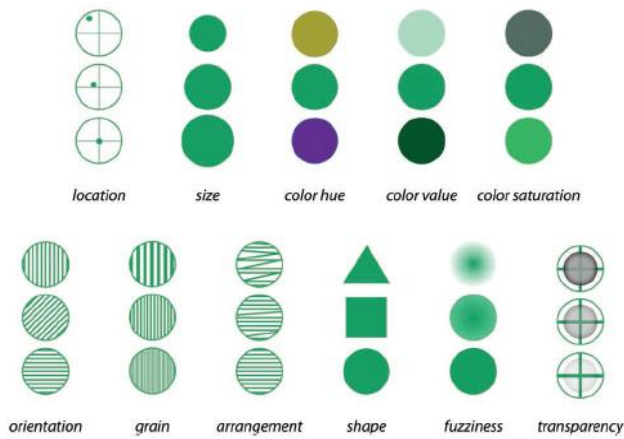


Figure 1. Visual variables proposed by [19]

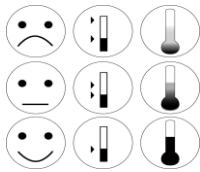


Figure 2. Smiley, Filled bar with Slider, and Thermometer icons proposed by [19]

## II. RELATED WORK

This section covers the existing works in two domains: the methods for integrating spatial data and the visualization of uncertainty in a spatial context.

### A. Spatial Integration

The same reality is described with a multiplicity of geographical information. This information growth rapidly over the Internet, some may be incomplete, inaccurate or contradictory. Integration of several sources of geographical information is necessary in order to update information that changes daily [13] or to produce a more complete and accurate information [8]. In [33], authors define three categories of imperfection that occurs when integrating several spatial data sources, namely (i) inaccuracy, which concerns wrong spatial information that do not correspond to reality, (ii) imprecision, which deals with spatial information that corresponds to reality but is not sufficiently precise and (iii) vagueness, which is about ambiguity of spatial information (e.g., boundaries heterogeneity). Geospatial integration has been widely studied under the term "map conflation" where two whole maps are integrated. Integration of maps consists in identifying the corresponding entities and to fuse them [6]. In [23], authors describe existing works in map conflation regarding their formats (raster and vector) and their criteria such as spatial data, terminological data and topological relationships between entities. Some works have been proposed in map conflation using points [24, 7, 31], lines [25, 11, 32] and polygons [1, 12, 20].

In [2, 26], the authors use only the spatial information (location) to detect the corresponding entities with a similarity measure based on probabilistic consideration. The probability

that two entities are corresponding is estimated using the Euclidean distance between them. In order to improve the quality of integration, some works propose to combine similarity measures that use spatial information with similarity measures that use terminological information to identify correspondences. In [27], three algorithms were proposed using a first similarity measure to filter the entities and a second to detect the corresponding entities. For example, a string similarity measure can be applied on the name of the POI, then for each pair of entities that are not considered as corresponding, the distance between them is increased. The final step is to apply a similarity measure on spatial information with the new distances. Note that increasing the distance between two entities lowers the probability that they will be considered as corresponding entities when we apply a similarity measure on spatial information. A variety of learning-based methods including logistic regression, support vector machines and voted perceptron has been proposed to find out how to combine and tune several similarity measures in order to identify the corresponding entities [28].

The "Theory of Evidence", also called "Dempster-Shafer theory" [29], combines an evidence measure of different sources and finds a degree of belief that takes into account all the available evidence. That is, a belief mass represented by a belief function, is associated to each evidence, then the Dempster's rule is used to combine them. The "Theory of Evidence" is proposed to integrate geospatial databases [22] and to match geospatial entities of several LBS providers [15].

Kang et al. propose a visual interface to detect the corresponding geospatial entities based on a neighborhood similarity [14]. It takes two sources of entities as input, then the user chooses a similarity measure to apply on terminological information or on spatial information. Detected entities are considered as potentially corresponding. Then each pair of entities are visualized on the screen. Their shared neighborhood of entities are placed between them and non-shared neighbors on the sides. Finally, the user has to make a decision for each pair to be considered as corresponding or not.

### B. Spatial Uncertainty Visualization

[30, 18] define nine categories of uncertainty paired with three "components" of geographic information: space, time and attribute. On this basis, [30] makes an empirical study to characterize the kind of visual signification that is appropriate for representing those different categories of uncertainty. The authors use a set of visual variables corresponding to the visual variables defined by [4, 21, 17]: Location, Size, Color Hue, Color Value, Grain, Orientation, Shape, Color Saturation, Arrangement, Clarity, Resolution and Transparency. Their symbol sets are points and for each visual variable, three degrees are specified coming from high to low certainty (Fig. 1). They add iconic/pictorial symbols to compare their efficiency according to abstract/geometric symbols such as Smiley, Filled bar with Slider, and Thermometer (Fig. 2). Two tests are realized to judge the suitability of different symbol sets for representing variation in uncertainty by manipulating one single visual variable for all the categories of certainty in all the components of geographic information or for one specific category of certainty (accuracy, precision,

trustworthiness) in each component of the geographic information.

### III. REPRESENTING UNCERTAINTY

This section covers our contributions for representing uncertainty in spatial integration. We first introduce an overview of our approach. Next we focus on the integration process, which produces confidence scores, and on uncertainty visualization on maps.

#### A. Approach Overview

A (semi-)automatic integration process does not achieve perfect results. Depending on data quality of providers, an integration process may have to deal with various kinds of uncertainty and to take decisions. In our geospatial context, the quality is strongly variable from one provider to another, and we need to take into account the uncertainty inherent to the process. Besides, this uncertainty should be represented, especially on a map. Our integration approach consists of three consecutive processes, namely "mediation process", "integration process" and "visualization process".

- **Mediation process:** it is in charge of processing and rewriting a spatial query. For each LBS provider, the initial query is rewritten to comply with the schema or model of each provider. In addition, the mediation process performs a blocking process, which reduces the set of returned entities based on the location (within a radius) and the POI type specified in the query. As an example, let us imagine that a user is interested in finding the hotels in Pittsburgh. This query may be rewritten as "accommodations in Pittsburgh" for a first provider, and as "hotels in Pittsburgh, PA" for another provider. The output of this process is a sets of entities returned from each provider to the mediator that are ready for the integration part. Note that the mediation is not further discussed in this paper, since the schema heterogeneity of the providers has been beforehand manually solved and that the blocking processes are performed using the providers' querying systems.
- **Integration process:** it aims at detecting and merging spatial entities which refer to the same POI (corresponding entities). It takes the sets obtained from the mediation process to produce a single set of entities, in which corresponding entities from different providers are merged into a single entity. Our integration process produces various confidence scores between the attributes of corresponding entities (see Section III.B). The lower the uncertainty, the higher the confidence levels. Note that any spatial integration system, which takes the same inputs, could be used in replacement.
- **Visualization process:** its main objective is the transformation of the confidence scores into visual representation of confidence levels (see Section III.C). In this process, the merged entities resulting from the integration are displayed on a map.

#### B. Integration and Uncertainty Computation

The challenges in entity integration traditionally deals with the selection of data and transformation functions to be used for merging. In our context, we can add the computation of relevant and useful confidence scores for spatial and terminological attributes. In this part, we describe a simple solution for detecting and matching corresponding entities and for computing confidence levels.

Many generic approaches for "entity matching" have been proposed [16]. Getting inspired by these generic approaches, we propose a simple entity matching process based on sophisticated similarity measures. The matching process is performed between all entities resulting from the mediation process. Given two entities from different sets, we compute confidence scores between their attributes. A score close to 0 means that two entities are totally dissimilar. Conversely, a score equal to 1 indicates that both entities are equivalent. The coordinates of two entities are compared according to the Euclidian distance. The shorter the distance between both entities, the closer to 1 the similarity value for coordinates is. All terminological attributes (e.g., name, phone) are compared using the Levenshtein measure. This measure is the most effective with regards to other string similarity measures [28]. Using several metrics to match the same attribute involves a new problem for combining the different similarity scores. When all the individual scores have been computed, we may also compute a global score. A weighted average is traditionally used for combining the individual similarity scores. A decision step is finally required to select the correspondences. Various methods such as a threshold or the top-K enables this automatic selection [3]. In our context, proposing the corresponding entities with the highest global score is sufficient. To select which attributes of corresponding entities should be merged, we apply statistics (mainly value frequency). This simple proposition of entity matching and merging aims at illustrating our uncertainty visualization solutions. Note that any integration algorithm, which takes the same inputs, can replace our proposition.

Concerning the output of the confidence scores, they are deduced from the similarity scores computed during the entity matching. The score computed between the coordinates constitutes the spatial confidence score. All terminological scores (between names, phones, etc.) are averaged to become the terminological confidence score. The global confidence score aims at evaluating the global confidence about a merged entity. For instance, the integration process produces a high score of spatial confidence when two providers locate the same POI at the same place and a low score of terminological confidence when two providers provide the user with totally different names, addresses, telephones, websites, etc. At the end of the integration process, corresponding entities have been merged and three confidence scores have been computed for each merged entity. The next step consists in visualizing these scores on a map.

#### C. Visualization of Uncertainty

Visualization of integrated information may be insufficient in various cases. For instance, a user needs to check original information when observing strange outcome from the

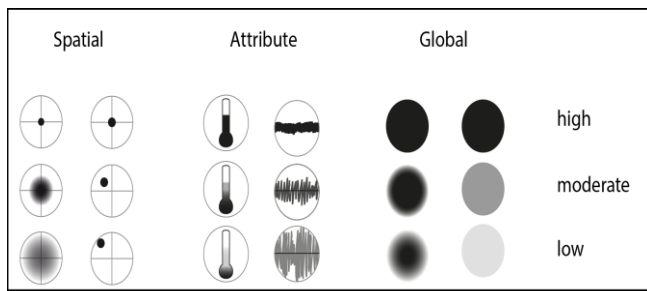


Figure 3. Visual variables chosen for our study

integration process. Therefore, the user requires to estimate himself the confidence of integration process visualizing (i) the spatial and terminological uncertainty for each integrated POI and (ii) the whole providers' source information. This requirement generates a large amount of information that might become an issue to visualize. To meet this requirement, in our approach, we first convert the spatial, terminological and global scores output from the integration process into three confidence levels (similar to the three uncertainty levels in [30, 19]): uncertain (low confidence level), moderately certain (medium confidence level), certain (high confidence level). The first range  $[0, 0.5]$  is associated to the uncertain level. The middle range  $(0.5, 0.75]$  includes the moderately certain values. And the certain level stands for highest values in the range  $[0.75, 1]$ . These ranges have been fixed according to experiments performed with similarity measures [9]. In the future, we intend to learn the best ranges for each level.

We are interested in monitoring uncertainty of two dimensions: the confidence level of spatial attributes (the spatial confidence score from integration process) and the confidence level of terminological attributes (the terminological confidence score from integration process). Moreover, to create a map easier to read and understand for a tourist, we propose to group these two dimensions of confidence to display one global confidence level. Then, a POI has a global (spatial and terminological) high confidence level when the data of the providers are consistent and complete between them. On the contrary, a POI has a global low confidence level when the providers are not consistent and/or not complete between them.

An analysis of the results obtained by [19] leads us to select the most relevant data useful in our context. Location, Size and Fuzziness variables are relevant to portray spatial uncertainty. Smiley, Filled bar associated with Slider and Thermometer are interesting to portray terminological uncertainty. Finally, Fuzziness, Location and Color Value are well suited to portray global uncertainty.

We define various cartographic proposals to portray confidence levels of POI that are oriented in two directions: first the choice of the visual variables, second the choice of the dimension(s) of the geographic information (the attributes) to display on the map.

#### 1) Visual Variables to Portray Confidence Levels

On the basis of conclusions made in Section II.B, we propose two alternative visual variables to portray the

confidence level of each dimension of geographic information (spatial, terminological, global). Fig. 3 illustrates them.

Concerning the spatial attributes, we decide to compare Location with Size associated to Fuzziness. We choose Location because it is intuitively connoted to space. We aggregate Size and Fuzziness. The taller the sign, the fuzzier the sign. We do this combination because independently, an order would be created between the signs with large or distinct signs seen before the others. This combination reduces this perception of order.

Concerning the terminological attributes, the proposals of [30, 19] have been investigated. For our application, Smiley is too connoted to a score relative to the quality of a POI obtained from the opinions of different users. Then if the smiley is happy, it will be interpreted as a good POI for the public (e.g., a "good" restaurant) and this is not what we want to represent. Concerning Filled bar associated with Slider, we think it is difficult to correctly perceive the differences between its three degrees because only one small element of the slider is modified. For the previous reasons, the Thermometer icon is selected and is compared to a new visual variable: Frequency, based on graphic representations created to show uncertain chaotic behaviors of signals in Electronics Science.

Finally, for the global confidence level (spatial and terminological attributes together), we choose to combine Fuzziness with Color Value. We eliminate Location because it is too connoted to the spatial dimension.

#### 2) Dimensions of the Geographic Information to Display on Map

Portraying whole uncertainty information may overload the interface. Our approach proposes instead to portray the confidence levels with a cartographic interactive application that gives the advantage to provide the user with only main confidence information and get more confidence details on demand opening a tool-tip to display complementary information. The user can also interact with the map (zoom in/out, move around, etc.). In such an application, various visualization strategies can be proposed depending on various confidence information we can highlight on the map. In the first two proposals, we make the assumption that spatial (respectively terminological) dimension of geographic information is estimated as the most important for the user. In this case, for each POI, we portray only the confidence level of this more significant dimension whereas the other one is shown in its tool-tip (Fig. 4).

In three other proposals, geographic information dimensions are both considered important for the user. In Fig. 5, spatial and terminological confidence levels are both portrayed on each POI using two signs. In Fig. 6, a global confidence level is displayed for each POI, corresponding to the confidence combination of spatial and terminological attributes. In this case, the tool-tip of each POI shows the spatial and terminological confidence levels. Finally, global, spatial and terminological confidence levels are all portrayed together for each POI (Fig. 7).

The next section illustrates some of our proposals by describing a use-case navigation scenario of a prototype we



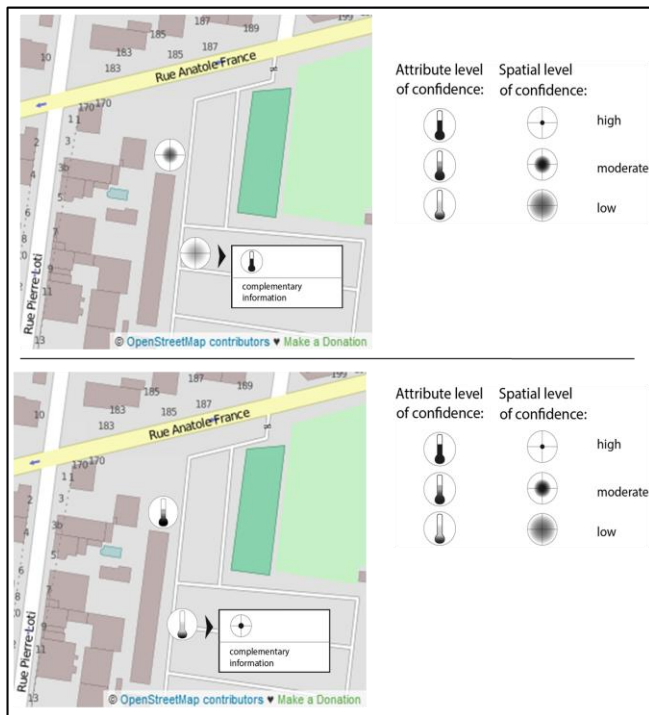


Figure 4. Spatial (top) or terminological (down) confidence level displayed on the map

have implemented. The proposals are Color Value for global confidence levels and in the tool-tip of each POI: Location, Size and Fuzziness to portray spatial confidence level and Thermometer to portray terminological confidence level.

#### IV. PROTOTYPE

Our proposal has been integrated in a LBS prototype. The POIs of this service are the result of the integration of the POIs from several LBS providers. This prototype implements the choice of solution presented above for visualizing uncertainty of integrated spatial data.

The prototype runs on an ad-hoc POI database that has been created collecting POIs of several types from three real providers using their Application Programming Interface (API). The integration process is pre-performed on the whole POI database and the prototype interface navigates through the result. The prototype interface is composed of three components as shown in Fig. 8: 1) POI types selector: a list that the user check/uncheck to display or hide, 2) legend: the visualization solution used to portray global, spatial and terminological confidence levels and 3) map inheriting OpenStreetMap background and features (zoom in/out, satellite/map view, etc). The user can choose two modes for the map, the former denoted as "Integrated mode" displays integrated POIs with their global confidence levels. The latter, denoted as "Source mode", portrays the POIs of the source providers of an integrated POI with full information.

When the user starts navigating, the prototype detects and centers the map at user location, and the "Integrated mode" is set by default. The user selects the POI types from the selector. All the POIs of the selected types that are near the user location

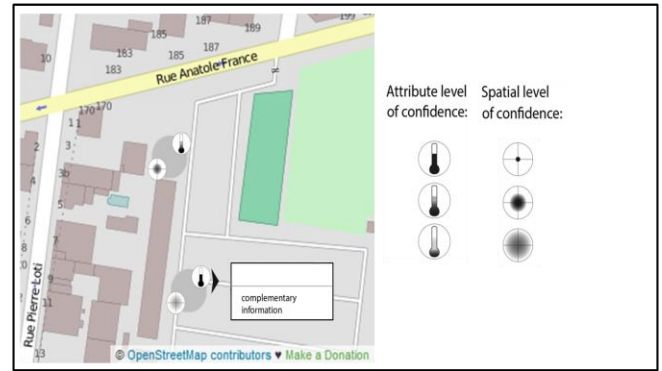


Figure 5. Spatial and terminological confidence levels are both portrayed

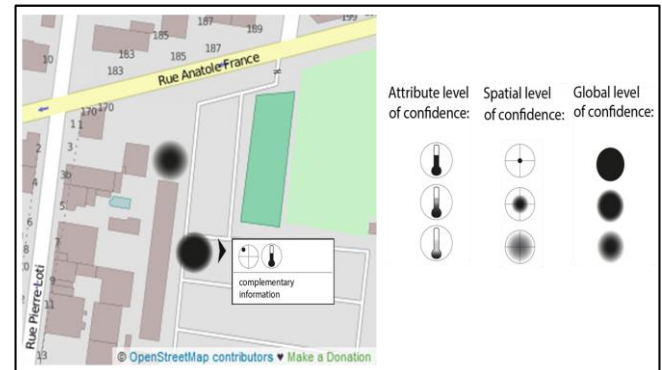


Figure 6. Global confidence level is portrayed, spatial and terminological confidence level are displayed in the tool-tip

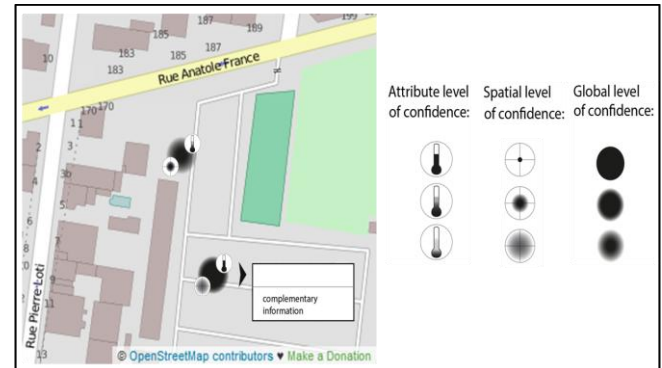


Figure 7. Global, spatial and terminological confidence levels are all portrayed together

are collected from the integrated dataset and displayed on the map. The map window of Fig. 8 illustrates the POIs of type Restaurant, Color Value variable indicating global confidence. Two deep colored restaurants have a high global confidence level (top and bottom), two light colored restaurants have a low global confidence level (in the center) and the three remaining have a medium global confidence level. The user can click on a POI to display the tool-tip that contains the full POI terminological information, spatial confidence and terminological confidence of the integration as shown in Fig. 8. At the right top corner of the tool-tip, the Thermometer icon indicates that the terminological confidence is medium while the Location, Size, Fuzziness icon indicates that the spatial confidence is low for the selected POI.

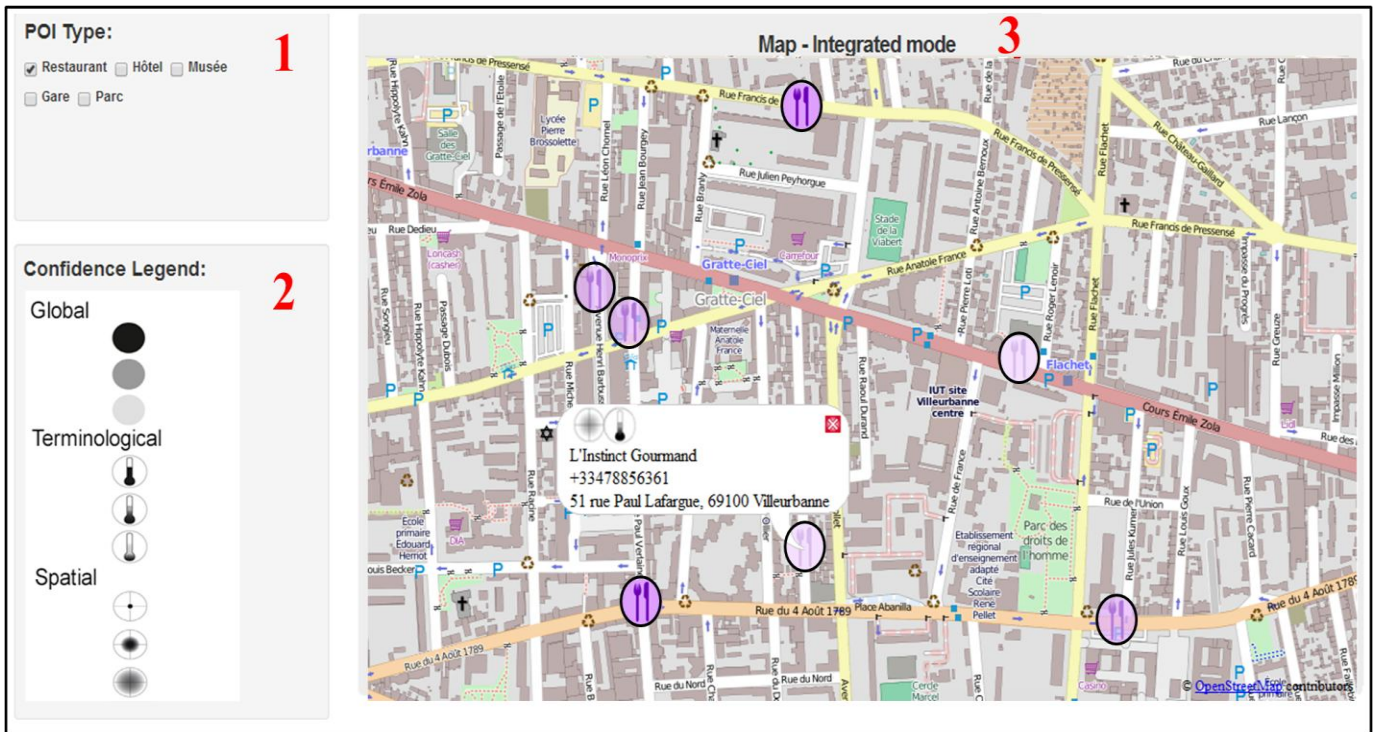


Figure 8. The prototype interface is composed of three components: 1) POI types, 2) legend and 3) map (here in Integrated mode).

	Provider 1	Provider 2	Provider 3
name	L'Instinct Gourmand	L'Instinct Gourmand	L'Instinct Gourmand
type	Restaurant	bar	RESTAURATION
address	51 Rue Paul Lafargue 69100 Villeurbanne France	51 Rue Paul Lafargue, Villeurbanne, France	51 rue Paul Lafargue, 69100 Villeurbanne, France
phone	+33478856361	04 78 85 63 61	04 78 85 63 61
site	undefined	<a href="http://www.linstinct-gourmand.fr/">http://www.linstinct-gourmand.fr/</a>	undefined
Distance	10.02 m	33.17 m	142.41 m

Figure 9. Comparison of terminological information offered by several providers for the same POI

As well, the user can check the source providers of an integrated POI by switching to the "Source mode" where all the integrated POIs are hidden except the checked one. In this mode, the user can consult the full POI information delivered by all the source providers. This mode shows the location of the integration result and of all source providers that the user can compare. The user can also check out terminological information of every source and compare them all. A table that contains POI terminological information delivered by each provider can be displayed (Fig. 9). Also, the distance between

each source POI and the integrated one is indicated at the bottom of the table for each provider.

## V. CONCLUSION

In this paper, we have proposed and studied different representations of uncertainty in a spatial integration context. Our approach is generic and the simple integration process that we have presented can be replaced. The integration process merges corresponding entities and produces confidence scores at spatial, terminological and global levels. These confidence scores are converted into confidence visualization solutions. Solutions have been implemented into an application prototype to demonstrate the feasibility and the benefits in a scenario.

One of our future objectives is to customize the visual representation and the navigation process according to user profiles. To reach this goal, we plan to test our proposals to select, for each dimension of the geographic information (spatial, terminological and global), the most efficient visual variables. Quantitative tests (e.g., A/B testing) are used to evaluate alternative choices while qualitative tests (e.g., user observation, interviews) explore how end-users (tourists) navigate with interactive maps. These tests should allow us to select the solution that is both best perceived and the most useful for the end-user according to her expectations. They could also demonstrate how such uncertainty representation is partly user-dependent. In that case, learning automatically the best representation for a new user could be an interesting challenge. For instance, a dynamic prototype which allows users to customize the mode of representation would allow us to evaluate the preferred solutions and to identify their

correlations with various criteria such as user profile and device type (e.g., computer, smart phone).

#### ACKNOWLEDGMENT

This work was supported by the LABEX IMU (ANR-10-LABX-0088) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). Special thanks goes to OpenStreetMap for providing the base maps.

#### REFERENCES

- [1] E. M. Arkin, L. P. Chew, D. P. Huttenlocher, K. Kedem, and J. S. B. Mitchell, "An efficiently computable metric for comparing polygonal shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no.3, pp. 209–21, March 1991.
- [2] C. Beeri, Y. Doytsher, Y. Kanza, E. Safra, and Y. Sagiv, "Finding corresponding objects when integrating several geospatial datasets". *ACM International Workshop on Geographic Information Systems*, 2005, pp. 87–96.
- [3] Z. Bellahsene, A. Bonifati, and E. Rahm, "Schema Matching and Mapping," Springer-Verlag, Heidelberg, 2011.
- [4] J. Bertin "Semiology of graphics: diagrams, networks, maps," 1983.
- [5] P. A. Burrough, "Principles of Geographical Information Systems for Land Resources Assessment (Monographs on Soil Resources Survey)," Oxford University Press, USA, 1986.
- [6] M. L. Casado, "Some basic mathematical constraints for the geometric conflation problem," *International symposium on spatial accuracy assessment in natural resources and environmental sciences*, 2006, pp. 264–274.
- [7] C.-C. Chen, S. Thakkar, C. A. Knoblock, and C. Shahabi, "Automatically annotating and integrating spatial datasets," In *Symposium on Spatial and Temporal Databases*, 2003, pp. 469–488.
- [8] M. A. Cobb, F. E. Petry, and K. B. Shaw, "Fuzzy spatial relationship refinements based on minimum bounding rectangle variations," *Fuzzy Sets and Systems*, vol. 113, no. 1, pp. 111–120, 2000.
- [9] W. Cohen, P. Ravikumar, and S. Fienberg, "A comparison of string metrics for matching names and records," *Workshop on Information Integration on the Web*, 2003, pp. 73–78.
- [10] T. Devoegele, C. Parent, and S. Spaccapietra, "On spatial database integration," *International Journal of Geographical Information Science*, vol. 12, no. 4, pp. 335–352, 1998.
- [11] Y. Doytsher, "A rubber sheeting algorithm for non-rectangular maps," *Computers & Geosciences*, vol. 26, no. 9, pp. 1001–1010, 2000.
- [12] M. Gombosoi, B. Zalík, and S. Krivograd, "Comparing two sets of polygons," *International Journal of Geographical Information Science*, vol. 17, no. 5, pp. 431–443, July-August 2003.
- [13] I. N. Gregory, "Time-variant gis databases of changing historical administrative boundaries: A european comparison," *Transactions in GIS*, vol. 6, no. 2, pp. 161–178, March 2002.
- [14] H. Kang, V. Sehgal, and L. Getoor, "Geoddupe: A novel interface for interactive entity resolution in geospatial data," In *International Conference on Information Visualisation*, 2007, pp. 489–496.
- [15] R. Karam, F. Favetta, R. Kilany, and R. Laurini, "Integration of similar location based services proposed by several providers," In *Networked Digital Technologies*, 2010, pp. 136–144.
- [16] H. Kopcke and E. Rahm, "Frameworks for entity matching: A comparison," *Data Knowl. Eng.*, vol. 69, no. 2, pp. 197–210, February 2010.
- [17] A. MacEachern, "How maps work," New York, London: The Guilford Press, 1995.
- [18] A. M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler, "Visualizing geospatial information uncertainty: What we know and what we need to know," *Cartography and Geographic Information Science*, vol. 32, no. 3, pp. 139–160, 2005.
- [19] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan, "Visual semiotics and uncertainty visualization: An empirical study," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2496–2505, December 2012.
- [20] A. Masuyama, "Methods for detecting apparent differences between spatial tessellations at different time points," *International Journal of Geographical Information Science*, vol. 20, no. 6, pp. 633–648, July 2006.
- [21] J. L. Morrison, "A theoretical framework for cartographic generalization with the emphasis on the process of symbolization," *International Yearbook of Cartography*, vol. 14, pp. 115–27, 1974.
- [22] A. Olteanu, "A multi-criteria fusion approach for geographical data matching," *International Symposium on Spatial Data Quality*, 2007.
- [23] J. J. Ruiz, F. J. Ariza, M. A. Urena, and E. B. Blazquez, "Digital map conflation: a review of the process and a proposal for classification," *International Journal of Geographical Information Science*, vol. 25, no. 9, pp.1439–1466, 2011.
- [24] A. Saalfeld, "A fast rubber-sheeting transformation using simplicial coordinates," *The American Cartographer*, vol. 12, no. 2, pp. 169–173, 1985.
- [25] A. Saalfeld. Conflation automated map compilation, "International Journal of Geographical Information System, vol. 2, no. 3, pp. 217–228, 1988.
- [26] E. Safra, Y. Kanza, Y. Sagiv, C. Beeri, and Y. Doytsher, "Location-based algorithms for finding sets of corresponding objects over several geo-spatial data sets," *International Journal of Geographical Information Science*, vol. 24, no. 1, pp. 69–106, 2010.
- [27] E. Safra, Y. Kanza, Y. Sagiv, and Y. Doytsher, "Integrating data from maps on the world-wide web," In *Web and Wireless Geographical Information Systems*, 2006, pp. 180–191.
- [28] V. Sehgal, L. Getoor, and P. Viechnicki, "Entity resolution in geospatial data integration," *ACM International Symposium on Geographic Information Systems*, 2006, pp. 83–90.
- [29] G. Shafer, "A mathematical theory of evidence, volume 1," Princeton university press Princeton, 1976.
- [30] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel, "A typology for visualizing uncertainty," In *Electronic Imaging*, 2005, pp. 146–157.
- [31] S. Volz, "An iterative approach for matching multiple representations of street data," In *Proceedings of the JOINT ISPRS Workshop on Multiple Representations and Interoperability of Spatial Data*, 2006, pp. 101–110.
- [32] M. Zhang, W. Shi, and L. Meng, "A generic matching algorithm for line networks of different resolutions," In *Workshop of ICA Commission on Generalization and Multiple Representation Computing Faculty of A Coruna University-Campus de Elvina, Spain*, 2005.
- [33] M. F. Worboys and E. Clementini, "Integration of imperfect spatial information," *Journal of Visual Languages and Computing*, vol. 12, no.1, pp. 61–80, February 2001.



# Proceedings for Distributed Education Technologies





# Smarter Universities

## A Vision for the Fast Changing Digital Era

Mauro Coccoli  
DIBRIS  
University of Genoa  
Genoa, Italy  
mauro.coccoli@unige.it

Angela Guercio  
Dept. of Computer Science  
Kent State University  
Kent OH, USA  
aguercio@kent.edu

Paolo Maresca  
DIETI  
Federico II University  
Naples, Italy  
paolo.maresca@unina.it

Lidia Stanganelli  
Università Telematica  
eCampus  
Novedrate (CO), Italy  
lidia.stanganelli@ecampus.it

**Abstract**—In this paper we analyze the current situation of education in universities, with particular reference to the European scenario. Specifically, we observe that recent evolutions, such as pervasive networking and other enabling technologies, have been dramatically changing human life, knowledge acquisition, and the way works are performed and people learn. In this societal change, universities must maintain their leading role. Historically, they set trends primarily in education but now they are called to drive the change in other aspects too, such as management, safety, and environment protection. The availability of newer and newer technology reflects on how the relevant processes should be performed in the current fast changing digital era. This leads to the adoption of a variety of smart solutions in university environments to enhance the quality of life and to improve the performances of both teachers and students. Nevertheless, we argue that being smart is not enough for a modern university. In fact, universities should better become smarter. By “smarter university” we mean a place where knowledge is shared between employees, teachers, students, and all stakeholders in a seamless way. In this paper we propose, and discuss a smarter university model, derived from the one designed for the development of smart cities.

**Keywords**—Smart cities; smart applications; collaborative systems; technology enhanced learning.

### I. INTRODUCTION

At times, technological innovations have contributed to the creation of neologisms by introducing novel buzzwords such as, e.g., *micro*, *cyber*, *virtual*, which are used to identify the latest cutting-edge solutions. As an example, let us consider the prefix “e-”. The massive adoption of Internet and web-based solutions has suddenly given birth to e-mail, e-commerce, e-banking, e-learning, and many other modern terms. In many cases, the “e-” has been replaced by the suffix “2.0” to move the attention to a further evolutionary step of the same product. Now we have entered the smart-*something* era, in which the prefix “smart” is attached to devices with computing and/or network capabilities. Moreover, such devices offer some form of *smartness* since they are easy to use and designed to improve users experience in common operations. Hence, we make a daily use of smart-phones, smart-TVs, smart-fridges, and so on. Riding this wave, the prefix smart has also been applied to places (e.g., smart-city, smart-building, smart-museum), and concepts (e.g., smart-work, smart-power, smart-grid). In this context, the design of a smart-system should

follow the human-centered design approach and exploit all available technologies to improve sustainability, environmental friendliness, reliability, mobility, and flexibility. In conclusion, smart systems and smart solutions are green, robust, personalized, responsive, interactive, and adaptive as well as accessible anytime, anywhere, from any device, according to the ubiquitous Internet paradigm.

In this scenario, we consider the concept of smart-university too. First of all, we notice that a commonly accepted definition is lacking in the literature. In particular, a tentative interpretation is given in [1], describing smart university as “*a platform that acquires and delivers foundational data to drive the analysis and improvement of the teaching & learning environment*,” by retrieving sensor-data, and using linked (open) data and formalized teaching knowledge. However, this is a merely technological approach and we observe that technology is just one among the many variables to take into account. In fact, recent modifications in laws and policy, also driven by economics and market analysis, are influencing universities’ learning environments and processes as well. Moreover, social issues, more recent innovations, and enabling technologies have been changing the way humans learn and thus are reshaping the relationship between learners and teachers.

These changes must be reflected in the university organization, which is asked to supply high quality services in order to stay competitive in a global scenario. This leads to the need for many modifications including the way in which teachers should work and in the creation of new models of students’ evaluation and assessment. As an example, one of the main changes recently observed in university teaching is the decrease in the amount of time for face-to-face lectures and accordingly, an increase in the amount of time for individual study, which is carried on by students mainly over the Internet. This new independent learning ability must be empowered by supplemental resources such as, e.g., video lessons, and scheduling of individual learning activities followed by self-evaluation. Moreover, the globalization process has dramatically accelerated the dynamics in production techniques and methodologies, thus requiring a more flexible education model, able to react quickly to unexpected changes whilst maintaining a high level quality. In addition, there are many human factors affecting the whole educational process. Among these, one of the most influential factors is that today’s students

have different attitudes and learning styles, derived from the highly interactive world they live in. Furthermore, the advent of social-media has influenced the way people use their knowledge across a distributed environment in a new collaborative fashion.

To cope with this reality, technology is no longer sufficient. We suggest that a paradigm change is necessary to transform a smart university into a *smarter university*, hence more efficient, more effective, and with a higher participation of both students and teachers, collaborating to achieve the common objective of better learning. In this respect the smarter university offers rich, interactive and ever-changing learning environments by exploiting the suite of technologies and services available through the Internet, by empowering individuals' abilities and attitudes, and by encouraging them to interact and collaborate in a framework in which people are co-responsible for raising and appraising the inclination of everyone. To achieve this objective, in this paper we analyze the issues of the current reality and, finally, we propose a model for the smarter university.

The remainder of this paper is organized as follows: Section 2 is a review of smart education including users' perspectives and a look at current and future trends; in Section 3 we outline the issues and the challenges of a smart educational ecosystem with reference to technologies, competences, and processes; in Section 4 we propose the model of a smarter university. Conclusions follow in Section 5.

## II. SMART EDUCATION

Education in a smart environment supported by smart technologies, making use of smart tools and smart devices, can be considered smart education. In this respect, we observe that novel technologies have been widely adopted in schools and especially in universities, which, in many cases, exploit cloud and grid computing, Next Generation Network (NGN) services and portable devices, with advanced applications in highly interactive frameworks. Thus, we can say that smart universities are already here. Nevertheless, smart education is just the upper layer, though the most visible one, and other aspects must be considered such as:

- communication;
- social interaction;
- transport;
- management (administration and courses);
- wellness (safety and health);
- governance;
- energy management;
- data storage and delivery;
- knowledge sharing;
- IT infrastructure;
- environment.

In this respect, six key areas are identified [2] for the design of an *iCampus*, where “i” stands both for *integrative* and *intelligent*. Namely, these areas are: *learning, management, governance, social, health, and green*. Other researchers focus on the Knowledge Management (KM) aspects, stating that KM is the foundation of a smart university, since it is the corner stone to fulfill business goals. Reference [3] describes a smart university as composed of five entities: *smart people, smart building, smart environment, smart governance*, and, last but not least, a *knowledge grid*. Other works address only technological solutions by outlining a smart space based on the use of RFID (Radio Frequency IDentification) technology [4] or providing NFC (Near Field Communication) support [5], [6]. To carry on with the overview of enabling infrastructure solutions, we must mention cloud computing as a resource for improving efficiency, cost, and convenience in the educational sector. Traditionally, cloud computing has been a convenient tool used in research laboratories for the optimization of computing resources. Today an increasing number of educational establishments are adopting cloud computing for economic reasons too [7].

### A. Users' Perspective

From the users' perspective, smart education is mainly related to the use of mobile web technology, which fosters a new conceptual model of mobile education in which teaching and learning activities are performed using ubiquitous computing. As an example, we mention the model proposed in [8], which deals with issues related to the usability of mobile applications and communication power among students and teachers. The paper refers to the FRAME model [9] and also includes a rich literature review that covers important concepts for the definition of new conceptual models, such as: the Transaction Distance Theory (DTD) [10], the Social Information Processing Theory (SIPT) [11], mobile web usability principles, cloud computing, and mobile web technology.

### B. Current and Future Trends in Education

To conclude this brief analysis of related works, we mention the reports published by the New Media Consortium (NMC)<sup>1</sup>, which contain interesting outlooks on trends in education and a perspective timeline for their adoption. For example, the 2012 edition [12] forecasted the success of mobile apps and tablet computing within one year or less, game-based learning and learning analytics adoption was indicated on a two to three years horizon, while gesture-based computing and the Internet of the Things (IoT) on a four to five years horizon. We still need time to see whether these estimates will be fully realized or not. In the 2013 edition [13], the main focus was on the success of Massive Open Online Courses (MOOC) in one year and new issues are 3D printing and wearable technology on a four to five years horizon. The current 2014 edition [14] highlights the key trends that are driving changes in higher education in the next years. Among these is mentioned the growing ubiquity of social media, the integration of online, hybrid and collaborative learning environments, the rise of data-driven learning and assessment, the shift from students as consumers to students as creators (a shift that will mark the

<sup>1</sup> <http://nmc.org>

Identify applicable sponsor/s here. (*sponsors*)

definitive evolution of online learning). It is worth noting that in this work, the authors indicate the “*low digital fluency of faculty*” and the “*lack of rewards for teaching*” as challenges to be solved.

This general scenario and these last statements confirm the thesis at the basis of this paper: there is a need for the change of the current model of the modern university from smart to smarter.

### III. ISSUES AND CHALLENGES IN SMART EDUCATION

According to the above situation, the university faces challenges to cope with novel learners’ needs, and to provide a seamless integration between the education production system (i.e., the education of the future workforce) and jobs, firms, industries, and organizations, which are requesting a multi-disciplinary education with complementary competencies and skills ranging from humanities to technologies. It is not unusual to find job searches demanding experts in humanities with technology skills, as well as technicians with good communications skills. This reflects the need for a novel profile of professionals and experts, which are required to have “*T-shaped*” skills, where the *T* denotes a solid basic knowledge on a specific, even narrow, area, complemented by a wide, yet shallow, knowledge of common topics.

Unfortunately, today most graduates only have a deep knowledge in a specific discipline. In fact, on the one hand, humanities courses disregard education in technology and, on the other hand, technological curricula (e.g., engineering, computer science, and physics) do not take into account social matters and writing skills. This prevents young graduates from having immediate success in the global market, which is focused on services, human interaction, and activities performed in teams. In this scenario we need to foster collaboration among all the parts involved in education. Students should have the opportunity of interacting with companies and industries during their studies, so as to be able to orient themselves towards specific applications. Moreover, the tight collaboration between the university and the business world can provide multiple advantages, such as having classes based on up-to-date programs and the most advanced technologies, and facilitating the technology knowledge transfer from university research laboratories straight to industry.

To overcome these problems, we can rely on most recent evolution in Information and Communication Technologies (ICT), which offers technological infrastructures, suited services, and platforms. Internet, grids, and cloud computing provide technologies for the restructuring of traditional education environments by supporting the more effective paradigm of a university with distributed resources: a university in which everyone (i.e., students, teachers, IT and administrative personnel, etc.) is provided with a rich set of functionalities that help to perform her/his activities through ubiquitous devices either via mobile services or in an unstructured environment. In addition to this, the Web 2.0 has among its most common functionalities, the ability to enable people to fully exploit communication. Therefore it supports the implementation of the paradigm of collaborative work and collaborative learning. From the students’ point of view,

collaborative learning can promote the generation of communities of practice and group activities, which can represent an advance look at the organization of the work field they will encounter in their forthcoming job, which, probably, will be asynchronous and geographically distributed.

#### A. Smart Technologies

In an ecosystem where learners, teachers, technicians and administrative personnel work together to improve the quality of their job results, technology plays a fundamental role, along with the individuals involved in the process. The most important points that we have identified are the following:

- The popularity of social network sites is dramatically changing interpersonal communication and the adoption of social media and networks in education is highly debated, especially in the community of pedagogy [15]. Social networks can become an alternative to traditional learning environments, and leverage a new generation of Learning Management Systems (LMS), especially for independent learning and communities of practice. Furthermore, social network sites are naturally related to the use of portable devices, thus fostering innovative mobile learning strategies as well.
- Cloud computing is changing software distribution and usage and its adoption deeply impacts on technological solutions. From the e-learning perspective, cloud and distributed computing support virtual laboratories [16] and extensive simulations, thus reducing machinery costs. Moreover, a mobile cloud-computing environment is the enabling technological framework for achieving effective functionalities for learning in mobility [17], [18].
- The use of smart devices fostered a new dimension in learning, especially from the students’ side, due to the continuous availability of network resources and advanced software programs. Modern mobile devices, especially tablets, make available a large number of applications in education. However this introduces problems from the providers’ side, especially due to privacy and security issues [19].
- Contactless technologies (e.g., Bluetooth, Quick Response Code, RFID, NFC) and modern network solutions are giving rise to novel interactive and immersive environments, some based on the IoT paradigm. Using interconnected smart objects results in multiple possibilities to design technology-enhanced learning activities [20]. Moreover, IoT can be exploited to deliver services to students within the university boundaries based on their position. In addition, location-based services coupled with contactless technology can be used to implement augmented reality systems.
- The evolution of the Web towards the Semantic Web (SW) is changing information retrieval techniques and content management. SW technology influences the way knowledge is organized and distributed. As an example, consider the adoption of a linked data-based

infrastructure in an Educational Semantic Web [21] to achieve a more efficient discovery of learning objects [22]. From an educational point of view, we mention an activity based on the use of semantic wiki [23] and more details on the scenario of SW in e-learning can be found in references therein.

- Peer-to-peer (P2P) has changed computer-to-computer communication and implements, at a technical level, the basic principle of sharing and collaboration. P2P is at the basis of file sharing systems, which can be used to circulate documents among people involved in the learning process, and to foster collaboration among peers by generating student-to-student relationships, thus facilitating the collaborative-learning model [24], [25], [26].
- The wide availability of network connectivity and its broad diffusion is stimulating interactions among people and fostering the adoption of collaborative paradigms. In education as well, collaborative learning is assuming new aspects, especially through the adoption of specific collaborative environments designed to make the most of students' teams effort (see, e.g., [27], [28], [29] and references therein).

We emphasize that smart education solutions must deal with most, if not all, of these topics. For example, social networking involves the use of mobile devices and introduces mobile learning, which relies on a cloud-computing distributed system. Moreover, using personal devices in an interactive environment implies the use of networked objects within the IoT, and so forth.

### B. Smart Competences

In today's historical context, there is an additional force accelerating the evolution process of education, that is the requirement of new competencies in the ICT area. This force is generated by the advent of smart cities, which results in the creation of new jobs for new professionals. An expertise is a proven ability in applying knowledge, skills and attitudes to achieve measurable objectives. The European countries have been discussing in recent years the need for the creation of such competencies, and have produced: (i) a document titled "European Competence Framework (ECF)", (ii) a website for the publication of the workforce [30], and (iii) a tool for building such competencies. The proposed approach is similar to the one already experienced in the European universities scenario with the introduction of the ECTS<sup>2</sup> (European Credit Transfer and Accumulation System). The ECTS represents a sort of "currency" of knowledge in a given area and provides a measure of the contents acquired, guaranteeing the portability across European universities.

In the same way, ECF represents the currency for the relevant competencies, allowing large public and private companies, small and medium-sized enterprises, local and state administration offices, career schools, workers, and professionals to share the same language for supply and

demand. Many vendors and big players too are currently mapping their certifications on the ECF requirements. ECF provides them with a tool to access a standard reference set of skills to apply for recruitment, placement, career paths, training, certification, etc. In other words, ECF provides a common European (and national) language to express the ICT skills.

This means that a person with a *bouquet* of skills can be identified through standard professional profiles that will be recognized across the European countries and, therefore, portable in terms of job opportunities. Hence, a smart university should have tools similar to the ECF framework to build educational profiles and, consequently, curricula and courses that both adhere to the standards required by the scientific and professional communities (e.g., IEEE, ACM), as well as to the platforms and agreements identified by the institutions. Some examples of platform creation are beginning to appear [31] but, at present, are mainly limited to experiences in single universities and, therefore, not ready to be exported. We need platforms that ensure the construction of courses of studies whose contents are fresh and relevant. Unfortunately, the construction of course contents does not follow this process very often and this situation increases the gap between education and companies' needs.

### C. Smart Processes

Smart courses and smart universities are like two sides of the same coin: one cannot exist without the other. In other words, smart courses can promote smart universities and smart universities can promote smart courses. From the organizational standpoint a university becomes smart when (i) it contains less bureaucracy, (ii) facilitates the creation of smart human social capital, and (iii) supports innovative investments fostering synergy between teaching and research. These aspects of university life have been measured over and over inside each university system. However, those measures have not always been an effective tool for modifying the process model fast enough. Therefore they did not help universities to evolve in a reasonable amount of time. Those metrics have not often helped in maintaining fresh and updated course contents, in building networks of supports of local and non-local companies or in creating new companies, and in promoting the direct participation of companies in terms of input acquisition to use in educational choices. Those measures have not increased the participation in scientific and social responsibilities of new and existing enterprises.

To conclude, a smart course can help turn a university into a smart-university with its ability to form talents and, indirectly, to develop innovations that can become the startup for something new. A smart course creates jobs since it promotes a vision that is not limited to the simple acquisition of knowledge, but aims to create culturally qualified personnel by anticipating users' demands.

## IV. THE MODEL OF A SMARTER UNIVERSITY

The starting point to define the model of a smarter university is represented by a shared vision among the various stakeholders (i.e., teachers, students, administration, nonprofit organizations, research institutions, citizens, industries, etc.)

<sup>2</sup> [http://ec.europa.eu/education/tools/ects\\_en.htm](http://ec.europa.eu/education/tools/ects_en.htm)

derived from the analysis of the territory (i.e., industrial reality, colleges and universities in close distance, types of schools, etc.), from the type of cultural areas that require intervention, and, above all, from the university strengths that can be applied in the territory in terms of social, economic, and cultural instances (both real or virtual).

During this preliminary phase of “opinion mining”, which can also be conducted using a “jam session” within social network sites (e.g., Facebook, Twitter, etc.), the active participation of citizens is fundamental. Their involvement must start at an early stage so as to ensure the consensus throughout the implementation of the entire operational plan and to guarantee that the plan is derived from a real need expressed by the territory. This procedure resembles the process of consensus building adopted for the development of a smart city.

Based on the above consideration, the definition process of a smarter university is derived from the model commonly used for the design of smart cities [32]. The proposed model is sketched in Figure 1 and the relevant components and their roles are described in this section.

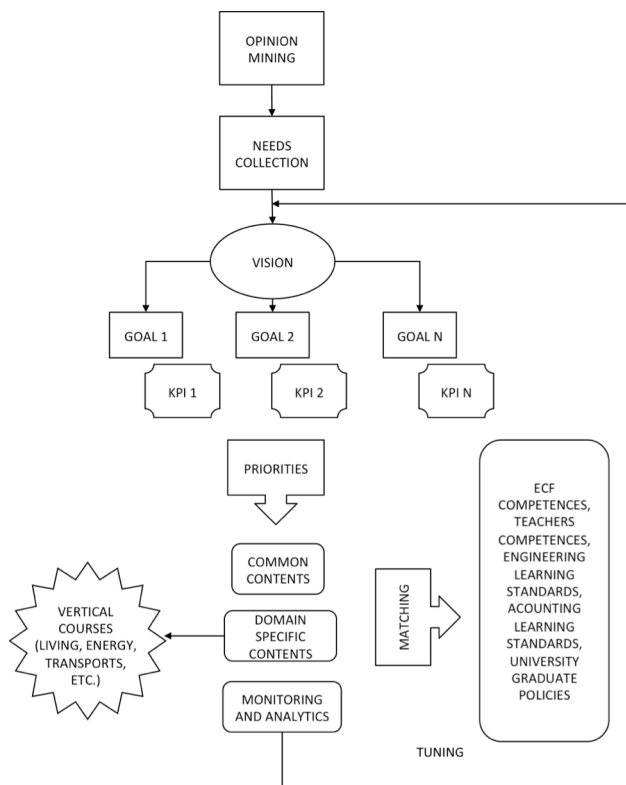


Figure 1. The model of a smarter university

- **Opinion mining** – The first step of the process is collecting different opinions, which will be later organized and structured. This approach is similar to the one used by the Coventry City Council in an experiment conducted by IBM. The trial aimed to produce the vision of a smart city and started by using

participation from the bottom up<sup>3</sup>. This step is a very delicate one and should be treated with particular care, because the vision of a smarter university produced will have a medium-long term impact on the future of an entire area. At the same time, the same vision must produce a return even in the short term. Research organizations and industries in the considered area, due to their competencies, should also participate in this phase and help drawing a clear roadmap.

- **Needs collection** – The second phase of the proposed model corresponds to an in-depth analysis of the needs emerging from the area, the communities and the organizations. In this step, the collected views are organized and structured according to their sources (stakeholders). These views are then translated into specifications and constraints of the system. The system generated in this phase is complex and has multiple input variables (both dependent and independent) as well as many constraints. A simulation of the system model at this stage is a challenge since it requires the study of a non-linear system, which provides a complex solution, assuming that a feasible solution exists. Nevertheless, the production of a simulation should not be dismissed.
- **Vision** – The presence of multiple variables and constraints encourages the creation of a “strategic” vision that must be translated into clear objectives, ambitious yet realistic. The objectives should clearly state the expected achievements in the medium and long term. Therefore, the strategic vision is structured with measurable objectives, a set of goals  $G = \{Goal_1, Goal_2, \dots, Goal_N\}$ , whose reach is measured through the set  $K = \{KPI_1, KPI_2, \dots, KPI_N\}$  of relevant Key Performance Indicators (KPIs). The goals are to be constantly monitored and the associated KPIs allow measuring the degree of achievement of each goal. This activity of monitoring and measurement is accompanied by a parallel corrective activity used to steer the objectives of the system, if necessary. A metric plan similar to the Goal/Question/Metrics [33] can be used at this stage.
- **Priorities** – The above mentioned objectives are then ordered as a two-dimensional array according to their priority and their measure of urgency. Specifically, higher priority goals are implemented first; lesser priority and less urgent ones follow. Goals with low priority and non-urgent goals are excluded. At this point, it is necessary to train dual thinker individuals (i.e., students). Dual thinkers have an ample and complete knowledge in one sector, but a general and broad knowledge in multiple domains. To achieve this objective the model identifies, on one hand, contents common to multiple domains and, on the other hand, contents specific to each domain in consideration.
- **Common contents** – The model extracts common contents, knowledge and skills that an individual must

<sup>3</sup> <https://www.collaborationjam.com>

have in multiple scientific areas, corresponding to the transverse part of T-shaped people.

- **Domain specific contents** – The vertical part of the T is represented by the knowledge and the skills that individuals must possess in a specific domain. For these domain-specific contents suited *vertical courses* are identified. In practice, at this stage the creation of multi-domain contents should be taken into account as well. These multi-domain contents may not be directly derivable from the previous step and might not be even vertical. For example, let us assume we are interested in forming the non-existent figure of an e-leader. In which domains should we seek the knowledge and skills for such a figure? Probably, those domains will include ICT, business, and leadership. In other words, this step is crucial for the formation of T-shaped skills, for both the vertical part and the completion of the horizontal one of the T shape.
- **Competences, standards and policies** – On the right side of the model depicted in Figure 1, the competencies are described in the ECF and teachers' skills are taken into account. In regard to the ECF, it is worth noting that in Italy the government agency, "Agency for the Digital Italy", not only has adopted the ECF 3.0 standard but it has also made it UNI 11506. This means that all the regional digital agencies must align to this standard. Italy is one of the first countries in Europe to adhere this revolutionary process. This adoption promotes uniformity, and produces a revolution, as well as it establishes a higher rigor, in the definitions of the competencies accepted at a European level. In addition to this standard, we consider the *de facto* or *de jure* required standards in terms of courses and contents of each major (e.g., engineering, economics), as well as the learning-related standards offered by the variety of university employees. Finally, the policies required by university management are taken into account. When this phase is completed, it is possible to evaluate the vertical paths oriented towards specific application areas.
- **Matching** – One of the most challenging parts of the model is the task of matching the choices with the needs. The complexity is not only due to the presence of requirements imposed by outside the system, but also due to the feasibility of the actions defined in the higher parts of the model. For example, it may happen that the professional figures identified by the model cannot be realized simply because there are no teachers with appropriate knowledge available. The point here is not only a simple question of supply and demand, which is an obvious role of a university. A smarter university wants also to look at the future of the professions and be ready to build them with a model that performs an appropriate matching accompanied by a corrective activity, where required. For example, assume we need to create the imaginary professional figure of an e-leader. This professional figure might have 3 types of skills: the first in ICT, the second in

Business and the third in Leadership. The activity of matching in this step should be able to check the depth of the gap between the profession we desire to build and what can be accomplished realistically with the conditions and the constraints of the system (skills of the teachers, companies in the territory, prior experiences, specific associations, etc.).

- **Monitoring and analytics** – The proposed model provides an abstract representation of a vision. Its application to specific situations and environments requires the implementation of a specific structuring of each part and such a process that can be very complex. It must also be said that this model requires the help of various tools such as forecasting, simulation, data collection and analytic tools. The monitoring and data collection must be implemented in such a way that it initially performs analysis on the data collected during the trial implementation of the process and successively uses the result of the analysis to make adjustments on the process model to adapt to the needs of the system.

The presented vision of a smarter university is the vision of the future university, which responds to the students' needs in a sustainable, social and technological way. "Being smart" should not be confused with "being digital". The ICT infrastructures are the means, not the end, enabling a set of services that affect deeply the life of the university. For example, the availability of broadband as a resource is essential to ensure that business is more competitive and to reduce the digital divide between citizens; the availability of low-frequency short-range technologies represent an essential resource for enabling the development of the IoT [34]. Similarly, Wi-Fi technology provides a valuable tool for reaching high-frequency strategic areas such as laboratories, libraries, meeting points, and so on. Of course, wherever the interconnection of these technologies is available, it might be possible to have an effective and timely monitoring, as well as a constant update, of each student's part of the vision.

#### A. Monitoring and Analytics Tools

Since the effectiveness of a university is a complex mix of several variables related to funding, programs, staff, demographic studies, curriculum, class size, classrooms types and availability, educational content, campus reachability and livability, and so on, when these data are distributed throughout offices across campuses, it is hard to have an integrated vision unless there is an organized sharing of digital data. In such a situation, no prediction is possible.

The availability of analytic tools that put together these variables and provide warnings to students, administrators and teachers is a useful way to overcome the delicate phase of monitoring. An effective monitoring can detect beforehand problems such as students' absence and dropout. For example, according the 2013 ANVUR (the Italian Agency of the Evaluation of the University System) "Report on the status of the Italian university system" [35], between 10%-30% of students leave university prior to the completion of their degree. Of 1,762,719 Italian students, 40,4% remains enrolled beyond the expected time for completing the degree, and in a



year are unable to reach the minimum quota of 60 credits. These are the students that fill the long list of unemployed individuals. Problems like this, however, could be diagnosed at an early stage by the monitoring phase of the model with the help of analytic tools. The individuals at risk could be reintegrated with the help of proper tutoring. However, this can happen only if the “intelligent” system is able to detect such individuals. For example, the Department of Education of Tennessee used an intelligent system, developed by IBM, for the management of students at risk of an upper secondary school in the Hamilton County. The analytic tool used collected all the existing data and, with the help of teachers, redesigned individual plans, customized curricula, and ad hoc tasks that brought the percentage of students able to complete their degree from 8% to 80% in one year [36]. This monitoring phase and the following analytic activity is even more important in a smart university system whose mission is to prepare young adults by shaping and providing them with multiple skills that will let them better compete in a market that is more and more service-oriented.

The newly required skills will need new teaching methodologies. Interaction and collaboration tools will be at the core of these methodologies. These new methods are important for two reasons. First, they transform a traditional course into a more interactive one that aims to increase concept retention in students. Second, they help developing and shaping the skills of the XXI century that will change the way in which students learn and work. Some of these skills are: problem-solving skills; adaptability; technology management; economics; as we observe in many companies’ job searches. Another example of an intelligent experience is the Teachers TryScience<sup>4</sup> initiative, supported by IBM. It is a project for the creation of a web site for teachers that provides free and engaging lessons, along with teaching strategies and resources, and events at the New York Hall of Science (NYSCI), designed to spark K-12 students’ interest in Science, Technology, Engineering and Math (STEM).

### B. Efficiency

In addition to this, the model of a smarter university assumes the existence of an efficient university administration open to continuous changes. This is a real critical point across multiple university systems. Each university should be able to control all its parts from a “single virtual automated room”, thanks to analytic and forecasting platforms that should help in managing the risks, the financial exposures, and anything else. A few initial experiments that support our vision have been set in place by some universities that attempt to be smarter. The results are encouraging. For example, the University of California has developed a risk management system that involves 10 campuses and 5 medical universities. By forecasting, managing and scheduling the risks, and by using simulations of activities, the participating universities have reduced by 39% their expenses, and reduced the insurance costs by 167 million dollars since 2006. North Carolina State University uses a cloud computing system to manage all the 250,000 students belonging to the entire educational system of the state public schools, including colleges and universities.

This offers to the state of North Carolina and to the university system a considerable savings in terms of cost of infrastructures.

Last but not least, a smarter university must be efficient even in term of energy. Smarter universities set in place plans for sensor-guided automated lights that turn on/off when necessary. Data analytics are run to detect deficiencies or problems that can be tuned to save money in electricity bills. Those savings can be later reinvested in teaching and new cloud based infrastructures. This is what Kent State University, Tulane University, Syracuse University, and other American universities have done: a significant step towards the creation of a sustainable territory.

## V. CONCLUSIONS

Currently, universities can be regarded as smart universities, as they profitably use available technologies to improve their performance and to enhance the quality of their graduates. Such smart universities act in the context of smart cities, which offer smart services and applications to their citizens to enhance their quality of life. Despite this, there is still space for improvement and universities should become smarter universities. In brief, the infrastructure and the model of the universities are already some how smart. To become smarter, we need a step forward in this direction from students, teachers, and all the other people involved in education. In a smarter university the ultimate technological solutions foster collaboration and cooperation among individuals.

In this paper we have outlined a process model where people’s needs are analyzed, and a set of goals are identified. Then the university acts to satisfy those goals, which represent current and/or future needs. Each goal is associated with a KPI so that the process can be monitored and tracked by means of suited analysis tools. In future work we will address how to improve this model and we will analyze and define accurate indicators for a better evaluation and assessment of the whole process.

## REFERENCES

- [1] T. Roth-Berghofer, “Smart university, the university as a platform, available at <https://smartuniversity.uwl.ac.uk/blog/?p=100> last accessed April, 2014.
- [2] J.W.P. Ng, N. Azarmi, M. Leida, F. Saffre, A. Afzal, and P.D. Yoo, “The intelligent campus (iCampus): end-to-end learning lifecycle of a knowledge ecosystem,” Proc. of the 6<sup>th</sup> Int. Conf. on Intelligent Environments, 2010, pp. 332-337.
- [3] M. Owoc and K. Marciniak, “Knowledge management as foundation of smart university,” Proc. of the Federated Conference on Computer Science and Information Systems, 2013, pp. 1267-1272.
- [4] Aqeel-ur-Rehman, A.Z. Abbasi, and Z.A. Shaikh, “Building a smart university using rfid technology,” Proc. of the 2008 Int. Conf. on Computer Science and Software Engineering, 2008, pp. 641-644.
- [5] G.M. Miraz, I.L. Ruiz, and M.A. Gomez-Nieto, “How NFC can be used for the compliance of European higher education area guidelines in European universities,” Proc. of the First Int. Workshop on Near Field Communication, 2009, pp. 3-8.
- [6] M.V. Bueno-Delgado, P. Pavon-Marino, A. De-Gea-Garcia, and A. Dolon-Garcia “The smart university experience: an NFC-based ubiquitous environment,” Proc. of the 6<sup>th</sup> Int. Conf. on Innovative Mobile and Internet Services in Ubiquitous Computing, 2012, pp. 799-804.

<sup>4</sup> <http://www.teacherstryscience.org>

- [7] N. Sultan, "Cloud computing for education: a new dawn?," *Int. J. of Information Management*, vol. 30, no. 2, 2010, pp. 109-116.
- [8] A.S. Hashim and W.F.W. Ahmad, "The development of new conceptual model for MobileSchool," *Proc. of the 6<sup>th</sup> UKSim/AMSS European Symposium on Computer Modeling and Simulation*, 2012, pp. 517-522.
- [9] M.L. Koole, "A model for framing mobile learning," in *Ally, M. Ed., Mobile learning: transforming the delivery of education and training*, 2009.
- [10] M.G. Moore, "The theory of transactional distance," in M.G. Moore (Ed.), *Handbook of distance education*, pp. 89-101, Mahwah, NJ: Lawrence Erlbaum, 2007.
- [11] J.B. Walter "Interpersonal effects in computer-mediated interaction: a relational perspective," *Communication research*, vol. 19, 1992, pp. 52-90.
- [12] L. Johnson, S. Adams, and M. Cummins, "NMC horizon report: 2012 higher education edition," 2012.
- [13] L. Johnson, S. Adams Becker, M. Cummins, V. Estrada, A. Freeman, and H. Ludgate, "NMC horizon report: 2013 higher education edition," 2013.
- [14] L. Johnson, S. Adams Becker, V. Estrada, and A. Freeman, "NMC horizon report: 2014 higher education edition," 2014.
- [15] A.J. Daly, *Social network theory and educational change*. Harvard Education Press, 2010.
- [16] L. Caviglione, and M. Coccoli, "Enhancement of e-learning systems and methodologies through advancements in distributed computing technologies," in *Internet and distributed computing advancements: theoretical frameworks and practical applications*, 2012, pp. 45-69.
- [17] M. Coccoli, and I. Torre, "A review of mobile-learning in mobile cloud computing environment," *Proc. of the Int Conf. of Education Research and Innovation*, , 2013, pp. 3499-3505.
- [18] F. Niroshinie, W. Loke Seng, and R. Wenny, "Mobile cloud computing: a survey," *Future generation computer systems*, vol. 29, no. 1, 2013, pp. 84-106.
- [19] L. Caviglione. and M. Coccoli, "Privacy problems with Web 2.0," *Computer Fraud & Security*, 10, 2011, pp. 16-19.
- [20] G. Adomi, M. Coccoli, and I. Torre, "Semantic Web and Internet of Things supporting enhanced learning," *J. of e-Learning and Knowledge Society*, vol. 8. no. 2, 2012, pp. 23-32.
- [21] K. Krieger, and D. Rösner, "Linked data in e-learning: a survey," *J. of Semantic Web*, 2011.
- [22] B. Yoosooka and V. Wuwongse, "Linked open data for learning object discovery: adaptive e-learning systems," *Proc. of the 3<sup>rd</sup> Int. Conf. on Intelligent Networking and Collaborative Systems*, 2011, pp. 60-67.
- [23] M. Coccoli, G. Vercelli, and G. Vivanet, "Semantic wiki: a collaborative tool for instructional content design," *J. of e-Learning and Knowledge Society*, vol. 8, no. 2, 2012, pp. 113-122.
- [24] P. Maresca, A. Guercio, and L. Stanganelli, "Building wider team cooperation projects from lessons learned in open communities of practice," *Proc. of the 18<sup>th</sup> Int. Conf. on Distributed Multimedia Systems, Workshop on Distance Educations Technologies*, 2012, pp. 144-149.
- [25] A. Guercio, P. Maresca, and L. Stanganelli, "Modeling multiple common learning goals in an ETC-plus educational project," *Proc. of the 19<sup>th</sup> Int. Conf. on Distributed Multimedia Systems, Workshop on Distance Education Technologies*, 2013, pp. 122-128.
- [26] L. Caviglione, M. Coccoli, and V. Gianuzzi, "Opportunities, integration and issues of applying new technologies over e-learning platforms," *Proc. of Int. Conf. on Next Generation Networks and Services*, 2011, pp. 12-17.
- [27] T. Arndt and A. Guercio, "Evaluating student attitudes on ubiquitous e-learning," *Proc. of the 7<sup>th</sup> Int. Conf. on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2013, pp. 98-101.
- [28] M. Coccoli, P. Maresca, and L. Stanganelli, "Enforcing team cooperation: an example of computer supported collaborative learning in software engineering," *Proc. of the 16<sup>th</sup> Int. Conf. on Distributed Multimedia Systems, Workshop on Distance Educations Technologies*, 2010, pp. 189-192.
- [29] M. Coccoli, P. Maresca, and L. Stanganelli, "Computer supported collaborative learning in software engineering," *Proc. of Global Engineering Education Conference*, 2011, pp. 990-995.
- [30] CEN ICT Skills Workshop, "The what, how and why guide to the e-CF, E-CF 3.0," available at the URL <http://www.ecompetences.eu/cen-ict-skills-workshop>, last accessed April, 2014.
- [31] R. Gluga, J. Kay, and T. Lever, "Foundation for modeling university curricula in terms of multiple learning goals sets," *IEEE Trans. on Learning Technologies*, vol. 6, no. 1, 2013, pp. 25-37.
- [32] A. Carriero, and S. Camerano, "Smart-city progetti di sviluppo e strumenti di finanziamento," 2013.
- [33] V.R. Basili and H.D. Rombach, "The TAME project: towards improvement-oriented software engineering," *IEEE Trans. on Software Engineering*, vol. 14, no. 6, 1988, pp. 758-773.
- [34] A. Sissions, "More than making things, a new future for manufacturing in a service economy," London, UK, The work foundation, Lancaster University, 2011.
- [35] ANVUR, "Rapporto Del Sistema Universitario e della Ricerca 2013," available at <http://www.anvur.org/attachments/article/644/Rapporto%20ANVUR%202013%20il%20sistema%20universitario%201.1.pdf>, last accessed April, 2014.
- [36] IBM, "Education for a smarter planet: the future of e-learning" available at <http://www-935.ibm.com/services/us/gbs/bus/html/education-for-a-smarter-planet.html>, last accessed April, 2014.

# Interaction with Objects and Objects Annotation in the Semantic Web of Things

Mauro Coccoli and Ilaria Torre

Department of Informatics, Bioengineering, Robotics, and Systems Engineering  
DIBRIS – University of Genoa  
Genoa, Italy  
{mauro.coccoli, ilaria.torre}@unige.it

**Abstract**—This paper presents a model and an infrastructure for technology enhanced learning applications and discusses issues in using Semantic Web, Linked Data, and Internet of Things. In particular, we apply the novel paradigm of the Semantic Web of Things to enhance information management and users' experience in a setup borrowed from a previous work, which exploits contactless technologies in a museum exhibition to enable visitors to interact with the artworks exposed and receive relevant information. According to the Linked Data principles, the content delivery system has been empowered by introducing suited semantic annotations to the objects. This enables “things” in the real world to be linked to their corresponding software images in the Web of Data and, in turn, it enables users to discover new facts about the objects and the available knowledge about them. Moreover, the enrichment of the real world objects with education-oriented annotations, which specify how to use the sensorized objects in educational activities, makes possible to match the objects features with learning tasks and users characteristics, thus providing personalized services and improving the learning experience.

**Keywords**—*technology enhanced learning; semantic web; linked open data; Internet of things.*

## I. INTRODUCTION

The investigation of the Semantic Web (SW) capabilities is a challenging item for future Internet applications and is a hot research topic too. Nowadays, a minor part of available web resources are based on the SW paradigm and, currently, it is difficult to find data in standard formats, due to both data providers and web designers attitudes. Furthermore, in most situations, the description of relationships between data is still lacking. However, programmers can rely on a full-featured suite of SW technologies for writing and querying the Web and this outlines a scenario ready for semantics, where relationships between data are expressed clearly and information can be linked accordingly. In particular, the Linked Data (LD) principles provide a publishing paradigm in which structured data can be easily exposed, shared and linked. Sharing common languages, publishing methods and tools, as well as ontologies, enables semantic search engines to effectively retrieve information associated to such linked data and provide advanced services to users [1]. In this scenario, embedding semantic annotations into real world objects, locations and events and making them available over the Web is a fascinating

perspective that goes under the name of Semantic Web of Things (SWOT). Such a technological solution can be exploited in many fields of application and it is very attractive from the viewpoint of technology-enhanced learning. In fact, by interacting with objects, learners can acquire knowledge and abilities in a natural fashion and find relations with other objects and data, driven by logical connections. Moreover, as objects are part of a SWOT framework, they contain information that can be further enriched with new data, retrievable from the Internet by means of suited queries. Besides, the possibility for learners to interact with things in the surrounding environment can be exploited in the perspective of the *experiential learning* [2], that is the process of learning and making meaning from direct experience. In this fashion, SWOT can become an educational platform and a methodology that takes the benefits of both inductive and deductive educational approaches, the former coming from the interaction with objects and the latter coming from the semantic network of concepts relating to objects.

Specifically, in this paper we describe an application based on the use of contactless technologies in a museum exhibition, which makes possible the interaction between users and the art works exposed therein [3], [4]. In shifting the same setup to an educational scenario, we have added semantic annotations to the objects and we have used this setting to provide augmented and personalized information, according to the framework that we described in [5]. For the annotation of artworks we have adopted the Europeana open data model [6] and we have extended it with properties for educational purposes from LRMI (Learning Resource Metadata Initiative) specification. Semantic annotation can thus be exploited by users to investigate on related objects in the Web of Data such as, e.g., getting additional information about the author of a specific piece as well as on the historical period, and so on, based on both the links included within the linked data and the specific learning goals and tasks.

Summarizing, the objective of this paper is describing a application scenario for significant experimentations with SWOT in education. Currently, the state of the art in this research field lacks of tangible results and the contribution of this project is providing a working proof of concept to push the adoption of such technologies in this area.

The remainder of the paper is organized as follows: in Section 2, related works are presented with reference to SW and the use of SW and Internet of the Things (IoT) in education; Section 3 describes the reference scenario selected for the sample application; Section 4 deals with the annotation process; Section 5 reports the interaction workflow with objects. Finally, conclusions are presented in Section 6.

## II. RELATED WORK

Given their relevance, coverage and popularity, we introduce this section by mentioning a set of ontologies whose triples are often linked by other datasets. One of the most common is DBpedia<sup>1</sup>, which adds meanings, expressed through links, to the information included in Wikipedia, the renowned free encyclopedia. Another relevant ontology is YAGO, developed within the yago-naga project [7], based on a larger dataset, including Wikipedia and other databases like Wordnet<sup>2</sup> for a clean lexical taxonomy, Geonames<sup>3</sup> for geographical information, and WordNet Domains<sup>4</sup> to the aim of achieving a better categorization. The third example that is worth to mention is the Freebase<sup>5</sup> project, which is described as a *“community-curated database of well-known people, places, and things”*. Essentially, it is an open collection of data and connections between them, in which topics are organized by type, and people can freely connect pages with links and semantic tagging.

The W3C website<sup>6</sup> provides information and comparisons about RDF triple stores, storage managers and semantic tools to support the use of structured data, and thus transforming them into an effective and more powerful source. In particular, a useful benchmark analysis is available, about storage managers for storing large-scale RDF data sets.

The current situation of SW applications is discussed in [8], which copes with discoverability, interoperability and efficiency of available public SPARQL end-points. However, this topic is continuously evolving and a glance on future applications is given in [9], which focuses on the problem of browsing and searching the SW from mobile devices and wearable computers too.

To conclude this general overview on the SW, a featured list of applications based on the use of SW technology, including both use case and case studies, is maintained by the W3C<sup>7</sup>. Among these, we mention the British Broadcasting Company (BBC) web site, which is one of the most successful applications of SW technologies and one of the first organizations to use LD.

### A. Semantic Web and Education

The current Web is a fundamental source of information in both school and universities and can be even the only resource such as, e.g., in some lifelong learning projects. It can greatly

benefit from the adoption of solutions based on the Semantic Web infrastructure and, especially, LD can empower technology-enhanced learning [10]. This results in an Educational Semantic Web [11] where information can be managed, shared and retrieved in a more effective way.

Beside the traditional use of SW technologies to annotate learning materials following a top-down approach, a more recent approach concerns the annotation of user-generated content (UGC). Reference [12] discusses the possibility to annotate UGC with contextual metadata, showing the benefits of classifying digital content that emerge during learning processes and activities. This approach, based on merging semantic and social web to improve the retrieval of education resources is analyzed also in [13]. The authors propose a specific architecture for sharing and retrieving educational resources by using a dataset/repository based on the LD principles.

Following the same approach of combining semantics and social web, but with a focus on the authoring phase of learning activities, semantic technologies have been used in [14] to support collaboration. The author presents a Semantic Web-based authoring tool developed with the aim of facilitating the teachers in planning collaborative learning scenarios, according to consolidated learning models.

Improvement in discovery and retrieval of learning resources are reported in several projects. In [15] the focus is on the first issue, while others stress in particular the concept of interoperability, which enhances the possibility of retrieval. Our approach goes along the same line. In [16], for example, the SW technologies are considered as the basis for the construction of e-learning resources and to enable cross-platform sharing and interoperability, through effective querying and positioning. In the mentioned approach, querying should be based on an ontology layer, to improve the retrieval and sharing of learning resources, and to attain consistent comprehension.

Furthermore, a tentative listing of the benefits deriving from a semantic approach in mobile learning environments is presented in [17], with a focus on integrating mobile devices with web-based services. In particular the authors investigate the development of a mobile ontology designed for a standard data exchange format among the mobile nodes in a wireless environment, with more institutions.

To conclude this overview on the main trends in SW for education, we mention the efforts to exploit the huge amount of public open data (present and forthcoming). Beside the interest in open archives, multimedia open data are currently attracting attention for their suitability to implement hypermedia activities and to pursue personalization in their design. For example, one can annotate educational video resources as in [18], where the authors annotate video using standard vocabularies including Friend-of-a-Friend (FOAF), the Dublin Core<sup>8</sup>, and the Timeline<sup>9</sup> ontology. Annotations, stored in a Sesame RDF repository, are used to improve video search and retrieval.

---

<sup>1</sup> <http://dbpedia.org>

<sup>2</sup> <http://wordnet.princeton.edu>

<sup>3</sup> <http://geonames.org>

<sup>4</sup> <http://wndomains.fbk.eu>

<sup>5</sup> <http://freebase.com>

<sup>6</sup> <http://www.w3.org/RDF>

<sup>7</sup> <http://www.w3.org/2001/sw/sweo/public/UseCases>

---

<sup>8</sup> <http://dublincore.org>

<sup>9</sup> <http://purl.org/NET/c4dm/timeline.owl>

Among the more recent initiatives concerning open data, we finally mention the Linked Up<sup>10</sup> project, whose sub-title is “*Linking Web Data for Education*”, aimed to exploit the vast amounts of public, open data available on the Web, published by educational institutions. A result of this project is the definition of a scalable approach for a more efficient generation of structured dataset topic profiles [19].

### B. Internet of Things and Education

The development of education strategies based on the use of the IoT paradigm is the next big thing in learning systems. Presently, IoT is a matter of investigation for an ever-increasing number of disciplines and the relevant applications are on-the-edge solutions for a variety of situations, including educational issues. Accordingly, the term IoT appears in the 2012 edition of the yearly survey on the impact of current and future technologies on education [20], published by NMC<sup>11</sup>. In this report, the authors foresee the spread adoption of IoT as an educational technology within the year 2017, highlighting the relevant benefits too. Such benefits refer to both the teaching/learning activities and the management of educational resources, including building and courses, as well. In fact, IoT can improve efficiency in managing energy resources and, for example, in remotely controlling laboratory machineries. The mentioned report gives also hints on how to exploit networked objects and content related to them, by presenting the example of an Allosaurus’ skeleton. Considering bones as distinct objects, each one could contain its related history: where and when it was discovered, its position in the overall skeleton, the temperature it is being stored for a better preservation, and more about the physical object. Generally speaking, an IoT enhanced learning framework, enables the users-to-objects interaction and this feature can be exploited effectively for the delivery of content personalized with respect to users characteristics, environment variables and current state. Many examples of reference architectures for learners to interact in an IoT environment are reported in the literature, e.g., [21], and motivations for the adoption of a learner centered approach can be found in [22]. A comprehensive analysis of the scenario of IoT in educations is reported in [23] and references therein, where the authors investigate the evolutionary trends in e-learning, pushed by the Semantic Web technology and the IoT, showcasing a number of relevant applications. Another example is described in [24], which reports about a university course offering an IoT based learning infrastructure allowing to teach IoT technologies through activities that include collaborative and collective programming of real world sensing applications.

In addition, the use of IoT technologies is considered the enabler for the realization of ubiquitous learning environments, in which learning demand and learning resources are everywhere in a seamless integration of life, study and work. In this framework, ubiquitous computing systems support the interactions between human and machines, allowing man-to-man, man-to-machine, and machine-to-machine data exchange and communications [25]. More recent works, envisage the use of IoT for the realization of the so-called living labs. With

reference to the concept of iCampus [26] it acts as a playground for people who work, study and live therein [27]. In this perspective, owing to the pervasive networking technology, the iCampus and the IoT work together for the delivery of learning content, driven by environment sensors. Hence, the campus itself can be used for teaching or research, and corresponds to the Living Lab model<sup>12</sup>, a real-life test and experimentation environment where users and producers co-create innovation.

### C. Semantic Web of Things and Education

The combination of SW and IoT gives rise to the SWOT, where semantic annotations are mapped to some real world objects and places. According to this, and to the above considerations, the adoption of a semantically enriched Web of Things becomes a desirable feature for most information systems, and especially in education. Many researchers are investigating this field and relevant architectures as well as frameworks are proposed (e.g., [28], [29], and [30]). However, at the best of the authors’ knowledge, there is a lack of mature and concrete experiences of educational applications of the SWOT, if we exclude applications concerning *Smart Campus*, which are focused on smart buildings rather than on educational features (e.g., [31]). To confirm this statement, we also observe that the already mentioned project Linkedup has launched a challenge for people to “*design and build innovative and robust prototypes and demos for tools that analyze and/or integrate open web data for educational purposes*”. Within this call-to-action, in the first (May - June, 2013) and in the second phase (November, 2013 - February, 2014) no one applied for a SWOT project. Instead, most of the proposals were about techniques and tools for personalization of educational processes and/or contents.

## III. INTERACTION WITH OBJECTS

In this section we describe the application scenario of a “smart education museum” that we used as a testbed to implement the LD principles and apply the SWOT paradigm. Specifically, we make reference to a previous work [4], where the interaction with objects is based on Radio Frequency IDentification (RFID) tags and readers. Nevertheless, the architecture is independent from the technology adopted and can be implemented with a variety of solutions, such as Near Field Communication (NFC) or Quick Response Code (QR-Code). More in details, the contactless technology is used to drive to custom information, which is duly stored in a multimedia repository. Figure 1 (partially borrowed from [4]) depicts the basic reference scenario, as well as device-to-environment interactions. The basic components of the architecture are a set of tags properly installed in the museum, used to trigger the delivery of contextualized (*where is the user?*) and personalized (*who is the user?*) information.

According to this, the basic entities involved in the system are:

- (i) contactless hardware,
- (ii) mobile devices,

<sup>10</sup> <http://linkedup-project.eu>

<sup>11</sup> <http://nmc.org>

<sup>12</sup> <http://openlivinglabs.eu>

- (iii) a suited communication network,
- (iv) the content management system and related modules for personalization and delivery, including web server capabilities,
- (v) the database to feed the system with proper information.

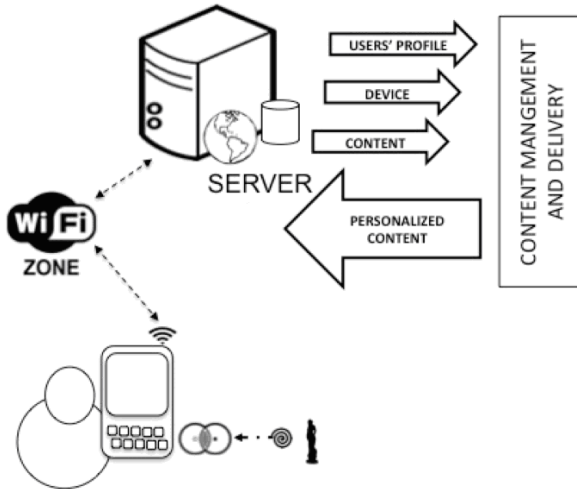


Figure 1. The reference architecture of the content management and delivery system.

To clarify the roles of the above components, let us consider the use-case of a visitor in a museum approaching a statue in the exhibition. The flow is the following:

- (i) the Point of Interest (POI) is “sensorized” with a proper tag enabling contactless interactions;
- (ii) the user is equipped with a mobile device compliant with the chosen contactless solution, i.e., has a camera for reading QR-Code or has a NFC/RFID chip to act as a reader. Moreover, the device has a suited software configuration, e.g., a Web browser or a specific application is installed;
- (iii) the interaction with the POI triggers the user’s device to send a request through the available network, e.g., IEEE 802.11 or 3G/4G connection;
- (iv) based on the specific POI, the server receives and processes the request, thus delivers multimedia content, keeping into account personalization issues, hence creating ad hoc content based on a user profiling (e.g., language and age), and on the features of the device (e.g., do not send videos if the device has no play-back capabilities);
- (v) the information is stored and managed by a suited database, which receives queries from the above mentioned server.

Put briefly, a 3-ple (USER\_ID, DEVICE\_ID, ITEM\_ID) is sent to the server at every interaction and to this aim, a suited information architecture was developed, based on an ad hoc

ontology reflecting the museum’s organization. The system was developed in such a way that it can act as a generic data provider, regardless the specific implementation, also to support third party applications or new functionalities.

Owing to this, we were able to modify the information layer of the system, while maintaining the overall IoT functionalities. The new information architecture addresses the SWOT requirements. Specifically, it is designed according to LD principles. Figure 2 displays the main modules that will be described in the forthcoming Sections 4 and 5.

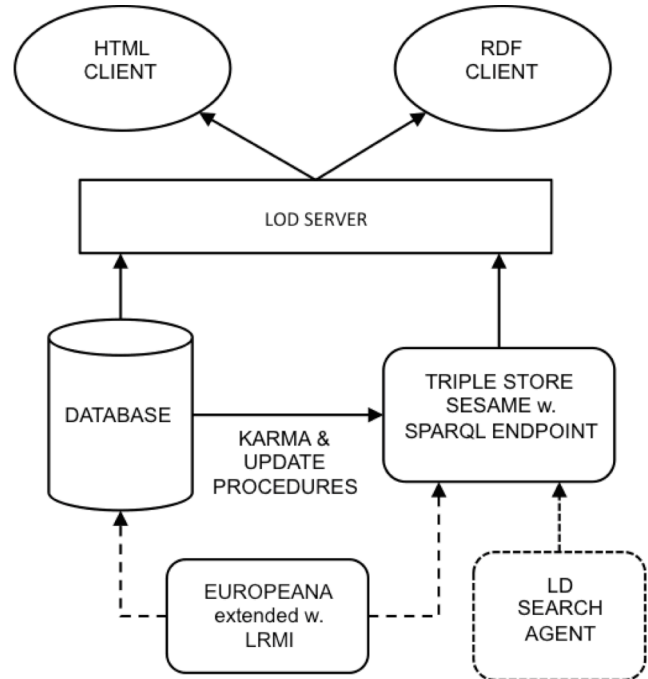


Figure 2. The information architecture.

#### IV. OBJECTS ANNOTATION

The annotation of physical objects according to the principles of LD enriches such objects with a digital representation of them, available on the Web. Thus, as outlined in the framework described in [5], the Web is going to become an extension of reality, where boundaries between physical and virtual objects are more and more faint. As packaging is often mixed with products, web data associated with objects is going to fade into objects. Annotations can be used to describe the features of the physical object, to provide information about its provenance, its location, its history and so forth. Therefore, in an educational scenario, enriching objects with annotations enables learners to get such data and to improve their knowledge.

The basic idea of using SWOT in this context is that, with reference to experiential learning [2], key drivers for education are: the physical interaction with the learning objects, the natural access to knowledge about the object, and the discovery of knowledge related to the object.



The SWOT paradigm seems particularly suited to cope with these challenges. In fact, accessing data about an object by interacting with it makes learning an active and natural task. Moreover, annotating POIs by following LD principles guarantees that: (i) POIs can be linked to other data about them in the LOD cloud (by means of RDF links to related objects and concepts), (ii) the retrieval of related data about POIs in the LOD cloud is more accurate, thus improving the discovery of knowledge about the object. The use of shared vocabularies to annotate objects further fosters the precision of retrieval, by reducing error risk in ontology matching.

Following these principles, we shifted from an ad hoc ontology to another one that exploits shared vocabularies and combines annotations, aimed to enrich the description of the artworks' characteristics with annotations specifically crafted for educational purposes.

The annotation process has been accomplished in three steps:

- a) definition of a data model for representing POIs;
- b) mapping the relational database to the data model;
- c) RDF data generation and enrichment with links to related data.

#### A. Data Model for Representing POIs

For the definition of the data model to use, we identified a set of ontologies, which could be candidate to represent POIs also addressing educational requirements: CIDOC CRM<sup>13</sup>, a reference conceptual model for cultural heritage resources, defined in 2006 as an ISO standard; EDM<sup>14</sup>, the data model for Europeana, the online platform for collecting and accessing European cultural heritage resources; the OpenArt ontology<sup>15</sup>, a data model that follows an event-driven approach, suited to trace events that occur over a specific time period (it originates from a project of Tate Britain art gallery for publishing a searchable corpus of materials relating to the arts in early modern Britain). Besides, other popular vocabularies that we also considered are specifically suited to describe agents (e.g., FOAF), places (e.g., GeoNames), and concepts (e.g., Simple Knowledge Organization System Reference - SKOS).

Finally, we decided to adopt the EDM model and to extend it with specific schemas for representing educational resources. The main reasons that led us to this decision are the following: (i) EDM is a general data model for cultural heritage objects which incorporates other standard vocabularies, some of them mentioned above; (ii) it is a well-documented data model; (iii) several mapping tools to other models are available; (iv) there is a vast repository of collections using such data model.<sup>16</sup> This point is very important for it is extremely useful for enriching objects since the use of a shared schema reduces ontology alignment problems when looking for related resources to be

included in the RDF file; (v) the classes of Europeana are properly suited for the IoT paradigm, in fact they distinguish a class that identifies the physical object (e.g., a painting, a book, etc.) from a class that provides one or more digital representations of the object (e.g., text, audio, video, quiz, etc.). These two classes, respectively `ProvidedCHO` and `WebResources`, are related one to each other through the `Aggregation` class. The `edm:aggregatedCHO` property, relates each instance of the `ore:Aggregation` class to a physical POI, while the `edm:hasview` property, relates each instance of the `ore:Aggregation` class to its digital representation. EDM uses specialization of classes and properties, also allowing different levels of granularity, as well as the representation of specific relationships such as, similarity between objects, part-whole relation, derivation, and others. The possibility to relate a physical POI to a variety of digital representations is useful in education, since it offers several options for filtering related resources based on the cognitive style and attitudes of individual learners.

Then, in order to enhance the educational purpose, we extended the EDM schema with the LRMI<sup>17</sup> specification. LRMI derives concepts and inspirations from models such as IEEE, LOM, and Dublin Core and it has been accepted and published as a part of Schema.org, the collaboration between major search engines, i.e., Google, Bing, Yahoo!, and Yandex, to annotate web resources. In our data model, we introduced a new class, `Audience`, with its LRMI properties and, in addition, we extended the `edm:WebResources` class, providing it with LRMI properties mapped from the `schema:CreativeWork` class.

As a consequence, for the new `Audience` class, we have three new properties:

```
schema:audienceType; schema:educationalAudience
schema:educationalRole
```

while, for the `WebResources` class, we have four new properties:

```
schema:educationalUse, schema:interactivityType
schema:learningResourceType and
schema:typicalAgeRange
```

#### B. Mapping the Relational Database to the RDF Data Model

One can choose among different techniques, to extract LD from a relational database. There are on the fly solutions, that use LD wrappers (e.g., D2R Server<sup>18</sup>, a tool for transforming relational databases to RDF), and there are solutions based on the migration to a triple store (e.g., Sesame, Virtuoso). In both solutions, when transforming the relational database into RDF, usually, some additional operations of mapping become necessary. Typically, mapping the data to RDF is done by writing rules in specialized languages such as R2RML<sup>19</sup> and D2RQ<sup>20</sup>, or by writing scripts in languages such as XSLT<sup>21</sup>, Python or Java. To achieve higher query performance, in our project we opted for the second choice, that is storing RDF data

<sup>13</sup> <http://cidoc-crm.org>

<sup>14</sup> <http://europeana.eu>

<sup>15</sup> <http://dlib.york.ac.uk>

<sup>16</sup> Several repositories are available using EDM data model: the Europeana collections, the Musical Instrument Museums Online (MIMO), CARARE digital content from archaeology and architecture domain, HOPE content, etc.

<sup>17</sup> <http://lrmi.net>

<sup>18</sup> <http://d2rq.org/d2r-server>

<sup>19</sup> <http://www.w3.org/TR/r2rml>

<sup>20</sup> <http://d2rq.org>

<sup>21</sup> <http://www.w3.org/TR/xslt>

in a Triple Store and maintaining the alignment of the two repositories through specialized synchronization procedures.

In order to support mapping and conversion to the store, we exploited KARMA<sup>22</sup>, a tool that can be used to map data in a variety of formats and from a variety of data sources (e.g., databases, spreadsheets, XML) to structured RDF. It has been used in several projects; just to mention one, it has been exploited to map the Smithsonian American Art Museum (SAAM) to the Europeana EDM ontology. KARMA provides a graphical user interface to model the source data according to the classes and properties defined in the target data model. Through learning techniques, it identifies possible mapping between the database schema and the ontology classes and then uses the ontology to suggest semantic types and relationships between them. In addition, specific relationships are created, based on “manual” mappings. As a further step, in order to fit the target data model, we extended our relational database with classes and properties within EDM and LRMI. In particular, we introduced new relations about artistic movements and we defined two new tables, about educational purpose and target audience. Through these, we can describe the type and the educational level of every web resource related to a physical POI. Finally, according to LOD principles, we had to assign an HTTP URI to every instance of the classes. After this pre-processing, KARMA was able to generate the RDF documents. Given some redundancies about code, a post-processing was necessary to produce the final RDF.

As a Triple Store we used Sesame<sup>23</sup>, a popular platform for managing RDF. It supports SPARQL query language, offers a Java GUI and provides a SPARQL end-point. According to the benchmark analysis on Large Triple Stores on the W3C<sup>24</sup>, it is scalable up to 70M triples. It runs on a Tomcat server and the platform includes Openrdf-Sesame, the actual Triple Store, and Openrdf-Workbench, the application that provides a web interface for managing RDF documents and querying them.

### C. RDF Linking

As discussed in the introduction to Section 4, the basic idea for publishing LD and making them accessible while interacting with physical objects, is exploiting a natural way to discover the world lying “behind” the object. In other terms, the objective is to consider the museum artworks as a gateway to the Web of Data related to such objects, thus favoring processes of experiential learning, informal learning and also edutainment. To accomplish this goal, a further step is necessary and deals with the actions to support the retrieval of resources related to POIs. These actions can be carried out in two stages of the objects’ lifecycle. The first one is the generation of RDF links to be included in the RDF description of POIs (e.g., the generation of Identity Links that connect the object to the URI of another object, using the `sameAs` property). The second one is the dynamic retrieval of data linked to a POI while interacting with it. This option requires an agent that crawls the Web of LD, retrieves related resources and returns them to the requesting client in a proper fashion.

Note that Figure 2 displays this agent (named LD search agent) in dotted line, since it is still under development in our project.

The retrieval of related resources, both to edit the RDF description of POIs and to dynamically retrieve relevant data, entails managing the problem of ontology matching, which can be treated at different levels: the semantic match between classes’ names, properties, and values. The choice of using commonly adopted data models is effective in reducing at least the first type of matching problems, making more accurate the retrieval of related data that use the same data model. In this respect, it is worth noting that EDM has done this choice as well, by including different popular data models within its ontology. To improve metadata interoperability, EDM ontology is based on OAI ORE (Open Archives Initiative Object Reuse and Exchange) for the representation of aggregated web objects, Dublin Core for descriptive metadata, SKOS for conceptual vocabulary representation, CIDOC-CRM for event and relationships between objects. Moreover, specific EDM properties (i.e., `edm:isRelatedTo`, `edm:hasMet`, and `edm:hasType`), are used to enhance the interlinking challenge.

In our project, annotating objects with RDF links is managed by using a semi-automatic approach. First, we use a set of API to run SPARQL queries against the EDM end-point<sup>25</sup>. In addition, to extend the search, we exploit other tools such as the DBpedia Lookup Service<sup>26</sup> and the Linked Media Framework<sup>27</sup>. DBpedia Lookup Service is based on a Lucene index providing a weighted label lookup, which combines string similarity with a relevance ranking (similar to PageRank) in order to find the most likely matches for a given term. The Linked Media Framework is a server application that bundles together some key open source projects to offer advanced services for linked media management.

## V. CONTENT NEGOTIATION AND INTERACTION WORKFLOW WITH POIS

In this section we detail the LOD server settings and the interaction mode between users and POIs. To demonstrate the interaction workflow, a proof of concept is set up. This is composed by a LOD server, able to serve the requests of two types of clients: web clients that process HTML documents and clients that process RDF. Specifically, the LOD server is available at `http://130.251.47.112/dbxml`, while Sesame at `http://130.251.47.112:8080/openrdf-workbench`. As a client, we developed an Android application that receives an URI from QR-Code or RFID tag and lets users choose how to browse information about that POI: in a *standard* or *augmented view*. The standard view, generates an HTTP request, adding the HTTP header `Accept:text/html`. The augmented view generates an HTTP request, adding the HTTP header `Accept:application/rdf+xml`. In both cases, it sends also the user data stored on the application (as detailed in Section 3).

Considering the LOD server, different configuration options are available, according to LD principles. As a first

<sup>22</sup> <http://www.isi.edu/integration/karma>

<sup>23</sup> <http://openrdf.org>

<sup>24</sup> <http://www.w3.org/wiki/LargeTripleStores>

<sup>25</sup> <http://europeana.ontotext.com>

<sup>26</sup> <https://github.com/dbpedia/lookup>

<sup>27</sup> <https://www.w3.org/2001/sw/wiki/LMF>

option, the URI represents the physical object and a 303 redirection with content negotiation is used to return its digital representation: an HTML page describing a POI for standard view or RDF content for augmented view. An alternative method is delivering RDFa embedded in a HTML document. In our project we chose the 303 redirect method, with RDF documents generated by KARMA and subsequently annotated with RDF links, as described in Section 4. In order to set up this method we had to manage the URI redirection, named Content Negotiation, and the generation of the dynamic pages for the standard view and the augmented view.

As displayed in Figure 2, requests from standard HTML clients are redirected to a dynamic page that accesses the relational database and generates an HTML response; requests from RDF clients are redirected to a page that runs a SPARQL query against the Sesame Triple Store and generates an RDF response. In both cases, the dynamic pages retrieve the POI's ID from the HTTP request and perform the query. Furthermore, in order to make the URI cleaner, we have used the Apache `mod_rewrite` and we defined a rewrite rule in the `.htaccess` file.

The description above outlines the basic mechanisms to serve HTML clients and RDF clients. The former output is suited for human users. The latter is machine understandable. Given this standard context for LD applications [32], what makes the RDF version an augmented view for users in our project is the possibility to:

- (i) exploit RDF links in order to provide users/learners with augmented content from the LOD cloud;
- (ii) exploit semantic annotations related to educational features in order to pursue the goals sketched in Section 4.

To accomplish these goals,

- (i) the SPARQL query to the EDM end-point (Section 4) is expanded or restricted according to the user data that the Android application sends to the LOD server within the URI request;
- (ii) transformation rules are applied to the retrieved RDF triples in order to display them in a user-adapted layout.

For the query expansion and restriction, several algorithms are available. We are working on testing the algorithm proposed in [33].

As regards layout transformation rules, the current version of the Android application implements mapping rules between RDF tags and HTML tags, to display aggregations of statements concerning a single object.

## VI. CONCLUSIONS

In this paper, we have introduced a model and an infrastructure to implement technology enhanced learning applications, particularly suited in the field of informal learning. Specifically, we refer to an application addressed to enrich a visit to a museum but the general principles expressed can be transported to many other fields of application.

In the proposed scenario, learners can interact with objects of the real world, which assume a pro-active role in retrieving information over the Web. In fact, owing to the available IoT technology and to the Semantic Web architecture, they can gather information about themselves, to the aim of enriching the set of data associated and to adapt to the individual users' needs, for example selecting resources in different language. Such a scenario results in breaking the boundary of smart-environments, where objects, locations, and events can communicate each other and with people. With the newly acquired SW capabilities, *things* in the real world can access knowledge over the Internet, enhancing their *self-knowledge*.

Such features are appealing for a wide range of applications but they can be regarded as a revolutionary approach for instructional designers to rethink their lesson plans in the direction of creating more captivating lessons. This can be done through the adoption of a renewed collaborative learning paradigm, where teacher-to-student and student-to-student interactions are not the only possible interactions. Now people can communicate with surrounding objects too in a completely new fashion and teachers/students-to-objects relations are introduced.

In this paper we have presented an infrastructure for that to happen. The basics for this infrastructure include a proper architecture of the content management and delivery for the IoT (Section 3), a data model for physical object annotation (Section 4) and a LOD server and workflow definition for the interaction with physical objects (Section 5). Given this infrastructure, several components can be added. In the paper we mentioned a LD Search Agent for the retrieval of related objects in the LOD cloud and, furthermore, specific Personalization Agents can be introduced to improve the search of related objects and the adaptation to the learner features.

## ACKNOWLEDGMENT

This research has been partially funded by the University of Genoa, within PRA 2013 projects, Prot. N. 9563.

## REFERENCES

- [1] F. Shaikh, U.A. Siddiqui, I. Shahzadi, S.I. Jami, and Z.A. Shaikh, "SWISE: Semantic Web based intelligent search engine," Proc. of the Int. Conf. on Information and Emerging Technologies, 2010, pp. 1-5.
- [2] D.A. Kolb, "Experiential learning: experience as the source of learning and development," Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [3] L. Caviglione, M. Coccoli, and A. Grosso, "A framework for the delivery of contents in RFID-driven smart environments," Proc. of the IEEE Int. Conf. on RFID Technologies and Applications, 2011, pp. 45-49.
- [4] L. Caviglione and M. Coccoli, "Design of a software framework for the management and personalization of contents in smart museums," Int. J. of Software Engineering, vol. 5 no. 2, 2012, pp. 51-69.
- [5] I. Torre, "Interaction with linked digital memories," Proc. of the Int. Workshop on Personalized Access to Cultural Heritage, within UMAP, vol. 997, 2013, pp. 80-87.
- [6] A. Isaac, R. Clayphan, and B. Haslhofer, "Europeana: moving to linked open data," Information Standards Quarterly, vol. 24, no. 2/3, 2012, pp. 33-40.
- [7] G. Kasneci, M. Ramanath, F. Suchanek, and G. Weikum, "The YAGO-NAGA approach to knowledge discovery," SIGMOD Rec. vol. 37, no. 4, 2009, pp. 41-47.

- [8] C. Buil-Aranda, A. Hogan, J. Umbrich, and P.-Y. Vanden-busshe, "SPARQL web-querying infrastructure: Ready for action?," *Proc. of the 12<sup>th</sup> Int. Conf. on Semantic Web*, 2013.
- [9] A. Dessi, A. Maxia, M. Atzori, and C. Zaniolo, "Supporting semantic web search and structured queries on mobile devices," *Proc. of the 3<sup>rd</sup> Int. Workshop on Semantic Search Over the Web*, 2013.
- [10] S. Lohmann, P. Heim, and P. Díaz, "Exploiting the Semantic Web for interactive relationship discovery in technology enhanced learning," *Proc. of the IEEE 10<sup>th</sup> Int. Conf. on Advanced Learning Technologies*, 2010, pp. 302-306.
- [11] K. Krieger, and D. Rösner, "Linked data in e-learning: a survey," *J. of Semantic Web*, 2011.
- [12] M. Svensson, A. Kurti, and M. Milrad, "Enhancing emerging learning objects with contextual metadata using the linked data approach," *Proc. of the 6<sup>th</sup> IEEE Int. Conf. on Wireless, Mobile and Ubiquitous Technologies in Education*, 2010, pp. 50-56.
- [13] S.T. Konstantinidis, L. Ioannidis, D. Spachos, C. Bratsas, and P.D. Bamidis, "mEducator 3.0: combining semantic and social web approaches in sharing and retrieving medical education resources," *Proc. of the 7<sup>th</sup> Int. Workshop on Semantic and Social Media Adaptation and Personalization*, 2012, pp. 42-47.
- [14] S. Isotani et al., "A Semantic Web-based authoring tool to facilitate the planning of collaborative learning scenarios compliant with learning theories," *Computers & Education*, vol. 63, 2013, pp. 267-284.
- [15] B. Yoosooka and V. Wuwongse, "Linked open data for learning object discovery: adaptive e-learning systems," *Proc. of the 3<sup>rd</sup> Int. Conf. on Intelligent Networking and Collaborative Systems*, 2011, pp. 60-67.
- [16] L. Rui, and D. Maode, "A research on e-learning resources construction based on semantic web," *Physics Procedia*, vol. 25, 2012, pp. 1715-1719.
- [17] T. Ercan, "Benefits of semantic approach in the learning environment," *Procedia - Social and Behavioral Sciences*, vol. 28, 2011, pp. 963-967.
- [18] Hong Qing Yu, C. Pedrinaci, S. Dietze, and J. Domingue, "Using linked data to annotate and search educational video resources for supporting distance learning," *IEEE Trans. on Learning Technologies*, vol. 5, no. 2, 2012, pp. 130-142.
- [19] B. Fetahu, S. Dietze, B. Pereira Nunes, M. Antonio Casanova, D. Taibi, and W. Nejdl, "A scalable approach for efficiently generating structured dataset topic profiles," *Proc. of the 11<sup>th</sup> Extended Semantic Web Conference*, 2014.
- [20] L. Johnson, S. Adams, and M. Cummins, "The NMC horizon report: 2012 higher education edition," *New Media Consortium*, 2012.
- [21] G.R. Gonzalez, M.M. Organero, and C.D. Kloos, "Early infrastructure of an Internet of Things in spaces for learning," *Proc. of the 8<sup>th</sup> Int. Conf. on Advanced Learning Technologies*, 2008, pp. 381-383.
- [22] M.G. Domingo and J.A.M. Forner, "Expanding the learning environment: combining physicality and virtuality - the Internet of Things for elearning," *Proc. of the 10<sup>th</sup> Int. Conf. on Advanced Learning Technologies*, 2010, pp. 730-731.
- [23] G. Adorni, M. Coccoli, and I. Torre, "Semantic Web and Internet of Things supporting enhanced learning," *J. of e-Learning and Knowledge Society*, vol. 8, no. 2, 2012, pp. 23-32.
- [24] G. Kortuem, A.K. Bandara, N. Smith, M. Richards, and M. Petre, "Educating the Internet-of-Things generation," *Computer*, vol. 46, no. 2, 2013, pp. 53-61.
- [25] R. Xue, L. Wang, and J. Chen, "Using the IOT to construct ubiquitous learning environment," *Proc. of the 2<sup>nd</sup> Int. Conf. on Mechanic Automation and Control Engineering*, 2011, pp. 7878-7880.
- [26] J.W.P. Ng, N. Azarmi, M. Leida, F. Saffre, A. Afzal, and P.D. Yoo, "The intelligent campus (iCampus): end-to-end learning lifecycle of a knowledge ecosystem," *Proc. of the 6<sup>th</sup> Int. Conf. on Intelligent Environments*, 2010, pp. 332-337.
- [27] J. Chin and V. Callaghan, "Educational living labs: a novel Internet-of-Things based approach to teaching and research," *Proc. of the 9<sup>th</sup> Int. Conf. on Intelligent Environments*, 2013, pp. 92-99.
- [28] D. Pfisterer, et al., "SPITFIRE: toward a semantic web of things," *IEEE Communications Magazine*, vol. 49, no. 11, 2011, pp. 40-48.
- [29] M. Ruta, F. Scioscia, and E. Di Sciascio, "Enabling the Semantic Web of Things: framework and architecture," *Proc. of the 6<sup>th</sup> Int. Conf. on Semantic Computing*, 2012, pp. 345-347.
- [30] K. Kotis and A. Katasonov, "Semantic interoperability on the Web of Things: the semantic smart gateway framework," *Proc. of the 6<sup>th</sup> Int. Conf. on Complex, Intelligent and Software Intensive Systems*, 2012, pp. 630-635.
- [31] M. Nati, A. Gluhak, H. Abangar, and W. Headley, "SmartCampus: a user-centric testbed for Internet of Things experimentation," *Proc. of the 16<sup>th</sup> IEEE Int. Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2013.
- [32] T. Heath and C. Bizer, "Linked data: evolving the web into a global data space," *Synthesis Lectures on the Semantic Web*, Morgan & Claypool, 2011.
- [33] M.C. Lee, K.H. Tsai, and T.I. Wang, "A practical ontology query expansion algorithm for semantic-aware learning objects retrieval," *Computers and Education*, vol. 50, no. 4, 2008, pp. 1240-1257.

# Computer Tutors Can Reduce Student Errors and Promote Solution Efficiency for Complex Engineering Problems

Paul S. Steif, Matthew Eicholtz, Levent Burak Kara

Department of Mechanical Engineering  
Carnegie Mellon University  
Pittsburgh, PA USA

**Abstract**—An ability to solve complex problems, for which a variety of solution paths are possible, is an important goal in engineering education. While feedback is critical to learning, hand grading of homework rarely provides effective, timely feedback on attempts to solve complex problems. Such feedback is also unfeasible in distance education contexts. A technology, based on the approach of cognitive tutors, is presented as a generally applicable method of providing automated feedback on complex problem solving, with truss problems studied in engineering as an example. The tutor maintains a cognitive model of problem solving of the class of problems, and associates various solution steps with distinct skills or knowledge components. One can determine whether students learn individual skills by measuring the error rate as a function of practice. Prior work has shown that for many skills the error rate indeed decreases with practice. New insight into the tutor's effectiveness, pertaining to the efficiency of student solution paths, is presented in this paper. While no explicit feedback is given regarding solution efficiency, it is found that students using the tutor become more efficient with practice. Furthermore, more efficient paths are found to be associated with making fewer errors.

**Keywords**—concept inventory; interactive learning; measures of knowledge; pre-post tests; Statics; web-based courseware

## I. INTRODUCTION

The development of problem-solving skills is a cornerstone of engineering education. While some problems that students learn to solve are simple, utilizing a single concept or principle, more complex problems are undertaken even in lower division courses. Students may need to coordinate and organize several concepts and steps, and many pathways to correct answers may be possible.

It is recognized in general that learning of any new skill is promoted by timely and effective feedback [1-4]. The opportunity for feedback on complex problem solving traditionally occurs through grading of handwritten homework. With weeklong turnaround such feedback is virtually never timely, nor is it readily made *effective*. Solutions can vary from one student to another, and with an incorrect answer, it is laborious for graders to identify and communicate to the student how the solution deviated from a correct path. Further,

in a distance-education setting hand grading would be largely unfeasible.

This paper describes a technology that can provide students learning to solve complex engineering problems feedback on their efforts. The technology must be able to follow and assess student solutions for a variety of pathways pursued. To that end, we adapt the approach of cognitive tutors, which have been developed for computer programming [5], math [6-7], and other fields. Such tutors are based on a cognitive model for a learner encountering the chosen tasks, and so can potentially provide feedback for a range of solution pathways. There do not appear to be previous efforts to devise cognitive tutors to assist students with complex engineering problems. The feasibility of a cognitive tutor style approach to providing feedback on complex engineering problems has been demonstrated [8] through a tutor focusing on truss problems, which are commonly studied in mechanical and civil engineering.

In the present paper, we consider in more depth the solution path taken by students solving problems with the tutor. While the tutor gives feedback on individual errors, it does not prompt students, except for rare circumstances, to think about whether their overall solution strategy is efficient. However, one can speculate that more efficient strategies may lead to fewer errors and that, in the course of solving problems, students may discover such solution strategies on their own. Here, efficiency relates to the maximum number of unsolved, yet defined, variables at any given time in the solution path. Using this definition, we investigate whether solution efficiency changes with practice and whether higher efficiency is associated with lower propensity for errors.

## II. DESCRIPTION OF TUTOR FOR TRUSSES

Since use of the tutor is intended to ultimately lead to success in solving problems with paper and pencil, user interactions with the tutor should be as unconstrained as possible, provided the tutor maintains the ability to judge user work. Although progress continues to be made in computerized interpretation of completely freeform work, for example via writing with a stylus on a tablet [9-13], such technologies may be limited for the foreseeable future; we have therefore defined

---

Support provided by the National Science Foundation through grant DUE 1043241 and by the Department of Mechanical Engineering at Carnegie Mellon University.

unconstrained as still within the confines of a mouse and keyboard user interface.

Fig. 1 displays a typical truss problem as it would appear in a textbook. The problem consists of a set of pins (dark circles) and connected bars. There are specified forces (10 kN) and supports (idealized constraints that keep the pins in position). Fig. 2 shows a portion of a solution to the problem in Fig. 1; user input corresponding to such solution elements must be enabled by the technology. A portion of the truss (a subsystem) including point C has been singled out for attention, the unknown and known forces drawn on the diagram (a so-called free body diagram or FBD), and equations of equilibrium (imposing Newton's laws of motion) have been written. In solving truss problems, students select multiple portions of the truss, and for each subsystem draw free body diagrams and write equilibrium equations. Students must also organize the solving of equations and interpret results physically in terms of the original truss. The solver can choose any portion of the truss, write equations in any order, then choose any other portion, and so forth. The technology must grant the user latitude to pursue this large space of solution paths and still be able to judge and give feedback regardless of the path chosen.

Even within the confines of a mouse and keyboard user interface, there are a few additional intentional constraints on how closely students' action with the tutor mimic paper and pencil solving. First, to reduce the cognitive load [14-15] associated with exercising skills already mastered by the student, certain tasks have been offloaded to the tutor; for example, we removed the need to enter numbers into an electronic calculator to obtain numerical solutions. Second, motivated by the self-explanation effect [16] in educational psychology, that students who explained problems to themselves learn more, the tutor introduces selective highly targeted opportunities to make the student's thinking visible, thinking which is rarely visible in pencil and paper solving. Specifically, the tutor requests the user to designate each defined force as falling into one of several categories.

We assume that students using the tutor have learned about truss analysis through other means, such as lecture and textbook. The tutor focuses exclusively on helping students solve problems, allowing a solution process such as depicted in Fig. 2 to be conducted on the computer with as little constraint as possible, while maintaining the ability to interpret student work. Observations of student work and typical errors [17] solving truss problems have guided tutor design. The goal is to allow a student using the tutor to commit most, if not all, errors that are observed in pencil and paper solutions.

Based on an analysis of the required tasks to solve truss problems, informed by prior work on the concepts and skills needed in the overall subject in which trusses are taught [18], and typically observed student errors [17], the computer tutor limits users to the following actions:

- Any set of pins, members, and partial members can be chosen as a subsystem for further analysis.
- In the free body diagram of a subsystem, forces can be drawn only at pins or at the free ends of partial members. Forces are confined to lie along horizontal

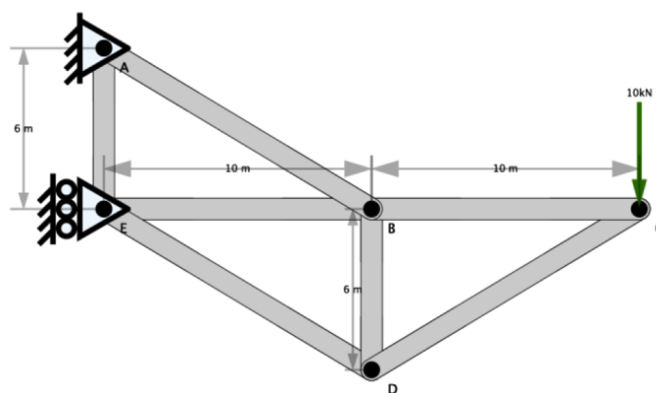


Figure 1. Typical truss problem, in which forces within the bars (members) are to be determined.

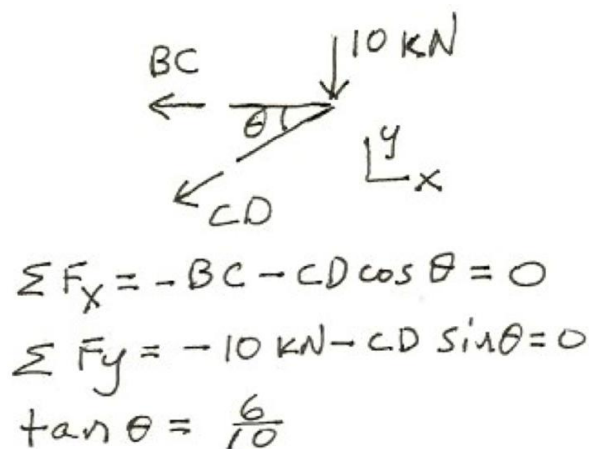


Figure 2. Portion of handwritten solution to problem in Fig. 1, showing the free body diagram of the pin at C and its connected partial bars, and associated equilibrium equations.

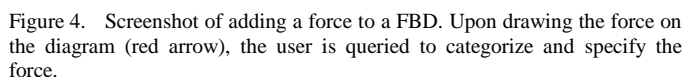
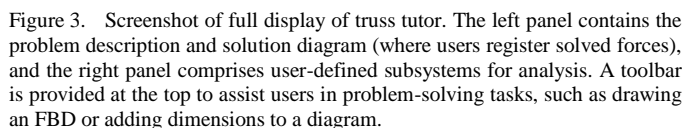
(x) or vertical (y) directions or parallel or perpendicular to members.

- For each subsystem, equations of force equilibrium along x- and y-axes, and equations of rotational equilibrium about any pin, can be written.

A screenshot of the tutor, with a problem partially solved, is shown in Fig. 3. The left half of the display contains a menu bar at the top and the problem diagram and statement. The problem diagram can be toggled to display the solution diagram, where support and bar forces that have been determined are registered by the student, as described below. The user chooses a subsystem for analysis by clicking on a set of pins, members and partial members, and then clicking on the draw (pencil) icon from the menu bar. The selected group of parts is added as another subsystem to the right half of the display, and would appear as one of the thumbnails depicted in Fig. 3. Clicking on a thumbnail expands that subsystem, allowing the user to draw its FBD and write its associated equilibrium equations.

In Fig. 4, we show a partially completed FBD with a new force being added to a half bar. In freeform solving of trusses, students draw arrows (for forces) and label those arrows with variables or numbers. With the tutor, we also require the user to





Requiring force categorization, together with the insistence on including partial members and pins in a subsystem, provides two benefits: (i) it helps students organize their thinking about the various forces in a way that can carry over to paper-and-pencil problem solving after tutor use and (ii) it establishes some clear bases for the tutor to recognize errors in student work, namely that applied and support forces can only act at pins, and internal forces can only act at the ends of partial bars.

Diagram of a truss structure with joints A, B, C, D, and E. A horizontal beam is supported by a pin at A and a roller at D. A vertical downward load of 40 kN is applied at D, and a vertical downward load of 30 kN is applied at C. A horizontal reaction force  $E_x$  acts to the right at E, and a horizontal reaction force  $A_x$  acts to the right at A. A window titled "Writing moment equation..." is open, showing a list of joints A, B, C, D, and E, with E selected. Below the window, a table shows the equilibrium equations:

	$\Sigma F_x = 0$	$\Sigma F_y = 0$	$\Sigma M = 0$
1	$\Sigma F_x = 0 \text{ (kN)}$		$E_x + A_x = 0$
2	$\Sigma F_y = 0 \text{ (kN)}$		$E_y - 40 - 30 = 0$

Figure 5. Screenshot of writing equilibrium equations for a given FBD. The user has already added consistent equations for the summation of forces in the  $x$ - and  $y$ -directions and is now trying to write a moment equation, which requires the selection of a moment center from the popup window.

with a specific subsystem. Admittedly, early in the subject in which trusses are taught, students do sometimes write down equations of equilibrium without specifying the subsystem or drawing its free body diagram. The design feature of the tutor, automatically associating equations with diagrams, is an instance where the trade-off between constraint and latitude seemed to argue in favor of constraint. The task of interpreting a large set of equations, each unassociated with a free body diagram, seemed overwhelming, with tutor errors likely.

We anticipate that a typical student would use the tutor for several hours over a period of a week or at most a few weeks, depending on the class. Therefore, the tutor must be easy to learn to use. Utilizing several rounds of user testing, we have sought to make its design as intuitive and simple as possible. However, some instruction in its use will inevitably be necessary. When students first start the program, the tutor appears with an example problem loaded, and on top of the tutor window there appears a window with a voice-over instruction video that addresses how to solve the example problem. The instruction video contains four phases, which deal with successive features of using the tutor. The video pauses after each phase, and prompts the user to go to the tutor window and carry out the portion of the solution just described in the video.

The tutor is a standalone application developed in C++; it can run on multiple platforms (Windows or Mac). While typical user interaction comprises mouse clicks and keyboard input, the tutor could be run on tablets or mobile devices using simple finger gestures, although we have presently not pursued this modality. For simplicity, we use the Qt toolkit<sup>1</sup> for graphics rendering, the Boost Spirit library<sup>2</sup> for parsing equations, and the SymbolicC++ library<sup>3</sup> for solving equations. All computation is done on the fly at interactive speed with no

<sup>3</sup> <http://issc.uj.ac.za/symbolic/symbolic.html>

apparent delay on a 2GHz single core desktop computer with 2GB RAM. Truss problems are created by hand and encoded as XML data. A separate XML file is generated to track solution progress; this data structure allows for easy export to the Pittsburgh Science of Learning Center DataShop tools [19] or MATLAB [20] for post-processing analysis. While specifically developed to complement statics instruction, our technology could be extended to other engineering disciplines provided a set of required learning concepts/skills is known and a cognitive model is established to enable timely, effective feedback; the specific cognitive model used in the present tutor is described below.

### III. JUDGING STUDENT WORK AND GIVING FEEDBACK

A key capability of the tutor is to judge student work and give feedback on it. The tutor does this by having a cognitive model for solving truss problems. The cognition in the tutor consists of the following algorithms corresponding to stages in the solution:

- **SUBSYSTEM:** An algorithm to determine if a group of pins, members, and partial members constitutes a valid subsystem.
- **FREE BODY DIAGRAM:** Given a valid subsystem, and any forces defined or determined up to that point, an algorithm for the allowable forces that can be drawn on the pins and partial bars of the subsystem. The FBD of a given subsystem is not unique; for example, if an internal force has been determined, the algorithm allows that force in a new FBD to be represented either as a determined force using the correct value, or as an unknown internal force, but the symbols should be consistent with the first definition.
- **EQUILIBRIUM EQUATIONS:** Given a valid FBD, an algorithm for the correct set of terms in the summations of forces along  $x$ - and  $y$ -axes and the rotational equilibrium equation about any pin. These summations include variables and constants and must be consistent with how forces appear in the FBD.
- **SOLUTION REGISTRATION:** Given a correctly determined support or internal force (from the equilibrium equations), an algorithm for the correct registration of that force in the solution diagram.

When to offer feedback on errors is a critical part of the tutor design. Prior research has shown that it is typically preferable to give immediate feedback [1], to ensure that the student associates the feedback with the action just taken. The tutor described here gives immediate feedback with the following caveat. Tutors for solving complex problems with limited constraints are distinct from most existing tutors: there is not a predetermined set of answers which users are expected to supply. The user is gradually adding elements of the solution on what is, in effect, a blank canvas. In contrast to the answer entered into a box, parts of the solution just added to the canvas, such as a force added to a free body diagram, may be tentative. It would be annoying and counterproductive to critique user work that is still tentative. On the other hand, if

errors accumulate too long and new work builds upon errors, judging new work becomes ambiguous.

The tutor balances these competing goals by identifying natural breakpoints at which each task can be viewed as completed and thus ready to be judged. The breakpoints are: (i) the subsystem is judged after the user has selected parts and clicked on the draw subsystem button; (ii) the FBD of the subsystem is judged after the user clicks to initiate the writing of the first equation; (iii) an equation is judged after the user types return while entering an equation or clicks to initiate the writing of a new equation; and (iv) the registered result is judged after the user has entered a result into the solution diagram and clicked OK. In each case, the student receives feedback that points out the error, with additional information to enable the user to fix the error and to learn why it is wrong, lessening the likelihood of repetition. Moreover, until the errors are corrected, the user cannot go on to the next stage of solution for the subsystem that has an error. Thus, it is unnecessary for the tutor to have algorithms to judge solution paths that build upon earlier committed errors. The student can pursue many different solution paths, but is halted on a chosen path until detected errors are corrected. Fig. 6 displays an overview of the tutor architecture, highlighting core modules involved in the cognitive model for providing feedback.

### IV. TRACKING EFFECTIVENESS OF FEEDBACK

As described in prior work [8], we studied the effectiveness of the tutor's feedback in helping students reduce the frequency of errors. To analyze the progression of learning quantitatively, the terminology, methodologies, and tools from the Pittsburgh Science of Learning Center DataShop were adopted [19]. The separate skills, or knowledge components (KC), with each task of the problem solving process were tracked separately. Whether a student correctly exercises the same KC at successive opportunities is monitored. Fig. 7 shows the percentage of students who erred as a function of opportunity for a group of students in a given cohort (e.g. a class of students taught by an instructor).

For each KC, we want to determine whether the error rate decreases with practice (a sign of learning), the error rate is always rather low (the particular skill is not difficult), or the error remains high or has no discernable pattern (feedback on errors appears to have little impact on future performance). Furthermore, a logistic regression model [21], similar to those used in other cognitive tutors, was applied to the data. The statistical model predicts error fraction according to the equation,

$$\ln[(1 - e_{ij})/e_{ij}] = \theta_i + a_j + b_j T_j \quad (1)$$

where  $e_{ij}$  is the probability of an incorrect answer by the  $i$ th student on opportunity  $T_j$  for using the  $j$ th KC. Note that  $e_{ij}$  can range from 0 to 1, and  $T_j$  takes on values of 1, 2, 3, and so forth, for the first, second, and third opportunity.

Fitting this model to data for a student sample yields the parameters in the statistical model. The parameter  $\theta_i$  captures the overall skill level of the  $i$ th student. The parameter  $a_j$ ,

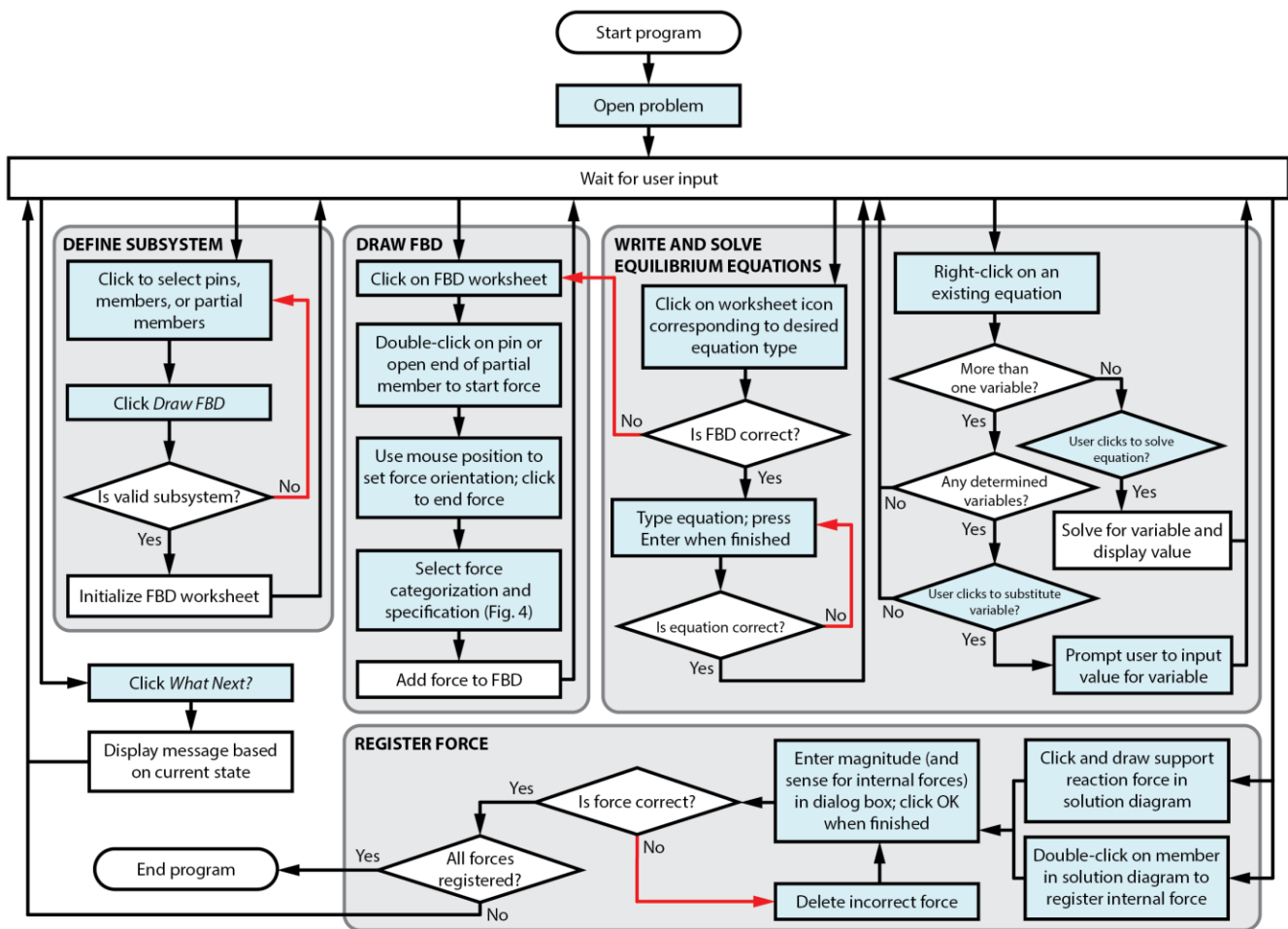


Figure 6. Flowchart of tutor processing routines. The core stages in solving truss problems are contained in gray boxes, tasks requiring user interaction are indicated by blue nodes, and user errors detected by the tutor are identified by red arrows. The parallel nature of the flowchart reflects the tutor's unique ability to allow users to explore multiple paths, only providing feedback at key junctures in the solution process. Some minor functions are not shown here for brevity; such tasks include zoom controls, adding angular dimensions to the problem diagram, and saving solution progress for future sessions or post-processing.

referred to as the intercept, reflects the initial probability of correctly applying the KC. The coefficient  $b_j$ , referred to as the slope, corresponds to the rate at which errors in using the  $j$ th KC decrease with successive opportunities to practice it. Prior work on the effectiveness of the tutor's feedback focused on the extracted values for  $a_j$  and  $b_j$ . For many of the skills, but not all, the values indicated that either a skill was relatively easy from the start ( $a_j$  high), or that the error rate reduced quickly with successive opportunities ( $b_j$  high). Relevant to the study described in the next section is the student skill parameter,  $\theta_i$ .

## V. SOLVING EFFICIENCY ACQUIRED THROUGH USE OF TUTOR

Explicit feedback given to users pertains to errors they make along the path they choose to pursue. As pointed out above, there are many paths that a user can choose. A path is described by the sequence of joints (each joint is a single pin and its connecting bars) that is chosen, and then, for each joint, the sequence in which the equilibrium equations are written. For each new joint, there are two or more unknowns, which correspond to the unknown forces in the connecting bars. But,

for each joint there are only two equations corresponding to equilibrium in the  $x$ - and  $y$ - directions. Thus, for each joint, the user can solve for at most two additional unknowns.

One strategy would be to choose joints at random, draw each FBD, and write down two equations of equilibrium for each. Then, one either substitutes into a solver for linear equations (not possible with the tutor and not available for students solving truss problems in exams), or one finds an equation that is solvable (only two unknowns), solves it and then substitutes sequentially into other problems. An alternative strategy would be to focus first on a joint that will be solvable, draw its FBD, write its equations, and then solve for the unknowns. The solved values can be used in analyzing the next chosen joint. We consider the latter approach a more efficient strategy. We hypothesize that, with it, one is less likely to get lost in the solution, and that one would also make fewer errors.

An instructor teaching students to solve truss problems may suggest that joints be chosen with an eye to introducing only two unknowns at a time. But, with the exception of one uncommon situation described below, the tutor does not directly promote a more efficient strategy or give feedback on

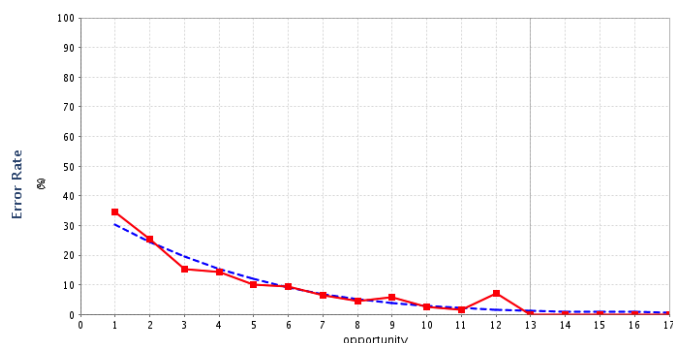


Figure 7. Percentage of students in error plotted as a function of opportunity (learning curve) on a skill (representing a determined support reaction) for which the error rate is initially high, but decreases with practice.

the strategy. However, for various reasons, for example because it is messy or frustrating to have too many unsolved equations, students may with time naturally gravitate towards a more efficient strategy. Certainly, another measure of the success of the tutor would be if students developed such an approach through using the tutor.

The efficiency of a solution path was studied by tracking the number of open (unsolved) variables. Variables are created when the user draws and labels undetermined forces on a FBD and remain open until the user registers a value for the force in the solution diagram. An efficient path, as defined above, should minimize the number of unsolved variables at any given time. The results presented here are based on data obtained from a class of 48 students enrolled in a regularly scheduled semester long statics course, 40 of whom completed all three of the problems studied here. Students used the tutor in lieu of a weekly homework assignment. Typical traces are shown in Fig. 8-10. User actions, as described earlier, include selecting a subsystem, drawing a FBD, writing an equilibrium equation, solving an equilibrium equation, and registering a determined force.

All problems can be solved with no more than three unknowns present at any time. Depending on the locations of the supports, one may have to analyze the entire truss prior to analyzing a joint; then there are three equilibrium equations (rotational equilibrium, as well as  $x$ - and  $y$ -forces). Therefore, a coarse estimate of efficiency relates to the maximum number of unsolved variables in a solution trace. We computed how many students solved each problem with no more than three unknowns present at any time. The efficiency rate, or fraction of students who were efficient, for the three successive problems is shown in Fig. 11. The results suggest that users, on average, did in fact become more efficient in the second problem compared to the first. (There is little change from the second to the third problem.) The difference in the efficiencies of the first and second problem is indeed statistically significant ( $z = 2.45$ ,  $p = 0.014$ ).

There is one rare circumstance when the tutor does promote efficient solving. When a user decides he or she is stuck, there is a button entitled *What Next?*. The message given is context dependent, and it often involves finishing work that has been started. However, if all initiated work is completed, the message tells the user to select a new joint, ideally one that has

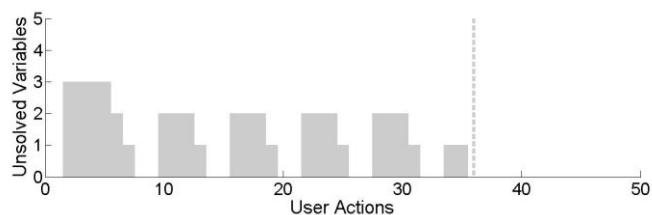


Figure 8. Example of an efficient solution path: the repeated sequence of steps (select subsystem, draw FBD, write and solve equations, register force) results in the "staircase" effect demonstrated here. The dotted line refers to problem completion.

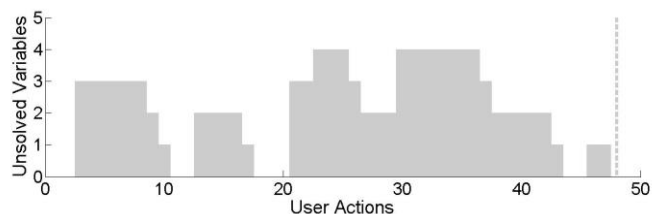


Figure 9. Example of an inefficient solution path: this user started in an efficient manner, but became inefficient after completing two subsystems. The dotted line refers to problem completion.

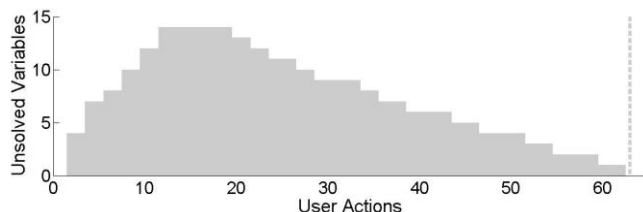


Figure 10. Example of an inefficient solution path: this user solves the entire problem in parallel fashion, drawing all possible FBDs before solving for any forces. The dotted line refers to problem completion.

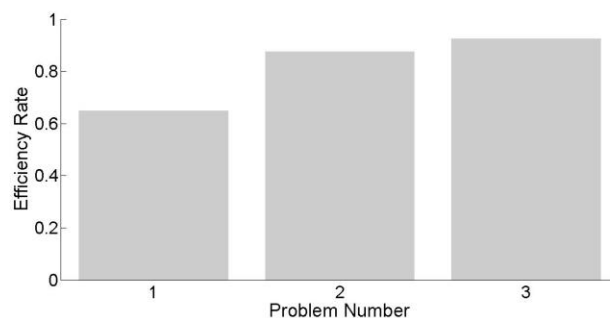


Figure 11. Fraction of students per problem who maintain no more than three unknowns throughout the solution path.

at most two unknowns. Of students in this sample, only eight clicked on the *What Next?* button, and of them only two received the particular message about choosing a new joint that is solvable, a strategy which is necessary, but not sufficient, to solving efficiently. Those two students received the message in the first problem; they were efficient in the second and third problems.

Finally, it was of interest to determine the relationship between choosing an efficient path and the propensity for errors, as measured by the overall student skill parameter,  $\theta_i$ , in the statistical model described in Section IV. Fig. 12 displays this relationship by plotting the individual ability of each student (skill parameter  $\theta_i$ ) against the number of problems they solved efficiently. The results demonstrate moderate positive correlation ( $r = 0.422$ ) and are statistically significant ( $p < 0.01$ ). Because the statistical model described in Section IV has only a single student parameter, we do not know whether individual students improve at different rates, even though for many skills we know that students improve with practice. Thus, the question of whether being coaxed to a more efficient path would reduce the number of errors a student makes – perhaps because the work space is less messy – cannot be answered without further study.

## VI. SUMMARY AND CONCLUSIONS

Complex problems that engineering students learn to solve often have multiple pathways to solution. It is difficult for human graders to provide effective formative feedback to handwritten solutions that are typically turned in as part of homework assignments. In this paper, we have described a technology, suitable for a distance education context, which can provide students learning to solve complex engineering problems feedback on their efforts.

By adapting the approach of cognitive tutors, we have developed a computer tutor that has a cognitive model of a student engaged in solving the problems of interest. The tutor interface permits the user to solve problems correctly following any pathway and to commit commonly observed errors. The cognitive model and the judicious timing of feedback give the user wide latitude to pursue different solution pathways, prevent new work from building upon previously committed errors, and still provide feedback on errors to enable students to complete most problems correctly.

As previously described, metrics for judging the effectiveness of feedback have been devised by viewing the solving of problems as a set of distinct skills or knowledge components. Actions by the student constitute opportunities to exercise different knowledge components; the effectiveness of feedback is quantified based on whether fewer students incur errors with successive opportunities on each knowledge component. The data from a cohort of students had been fit to a

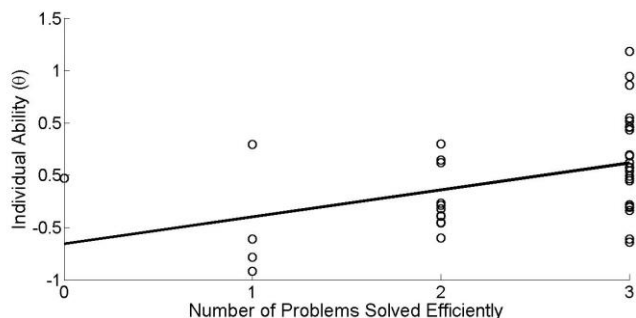


Figure 12. Correlation between individual ability and the number of problems solved efficiently. Each data point represents one student user.

statistical model that predicts the percent error as a function of opportunity. The model has several parameters, including initial difficulty and rate at which errors decrease with opportunity for each knowledge component. Fitting these parameters to the data provided evidence that the students do learn: that is, their propensity for errors decreases with practice. The statistical model also includes a single parameter for each student in the cohort, corresponding to the overall skill level.

The present paper has focused on a different aspect of student solutions using the tutor, namely on their overall solution path. By contrast to errors, upon which students receive feedback, the tutor does not give feedback to students on their overall solution path, nor does the tutor promote efficiency in the path. However, students may, through the additional effort needed when solving problems inefficiently, discover more efficient paths on their own. Furthermore, efficiency in the solution path may be associated with fewer errors.

To that end, a measure of efficiency involving the maximum number of unsolved variables at any instant was defined. It was found that the fraction of students who solved efficiently increased by a statistically significant amount from the first problem to the second problem, and held steady in the third problem. Furthermore, efficiency in solution was found to correlate significantly and positively with student skill, as measured by the overall skill level extracted from the statistical model. Thus, perhaps to reduce their workload, students do become more efficient with practice. Furthermore, higher efficiency is associated with making fewer errors. The possibility of causal relationships between efficiency and errors remains to be studied.

In summary, the technology described here has demonstrated that students can be given sufficient feedback, while working without an instructor, to complete complex problems, that their errors reduce with practice, that they become more efficient by virtue of their practice, and that such efficiency is associated with fewer errors. Moreover, because the technology maintains a detailed record of student work, and interprets that work meaningfully by way of a cognitive model, it enables the learning process to be studied and the tutor itself to be systematically improved with time.

## ACKNOWLEDGMENT

We thank Luoting Fu for his prime role in developing the tutor, and Jackie Yang, Jeremy Jiang, and Rebecca Piston for their assistance in development and initial testing of the tutor.

## REFERENCES

- [1] J. Hattie and H. Timperley, "The power of feedback," *Rev. Educ. Res.*, vol. 77, no. 1, pp. 81-112, March 2007.
- [2] J. R. Anderson, F. G. Conrad, and A. T. Corbett, "Skill acquisition and the LISP tutor," *Cogn. Sci.*, vol. 13, no. 4, pp. 467-505, 1989.
- [3] R. L. Bangert-Drowns, C.-L. Kulik, J. A. Kulik, and M. Morgan, "The instructional effect of feedback in test-like events," *Rev. Educ. Res.*, vol. 61, no. 2, pp. 213-238, Summer 1991.
- [4] A. T. Corbett and J. R. Anderson, "Locus of feedback control in computer-based tutoring: impact on learning rate, achievement and attitudes," *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 245-252, ACM, 2001.

- [5] J. R. Anderson, C. F. Boyle, and B. J. Reiser, "Intelligent tutoring systems," *Science*, vol. 228, pp. 456-468, 1985.
- [6] K. R. Koedinger, J. R. Anderson, W. H. Hadley, and M. A. Mark, "Intelligent tutoring goes to school in the big city," *Int. J. Artificial Intelligence in Education*, vol. 8, pp. 30-43, 1997.
- [7] K. R. Koedinger, "Toward evidence for instructional design principles: examples from Cognitive Tutor Math 6," *Proceedings of PME-NA XXXIII, Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, pp. 21-49, 2002.
- [8] P. S. Steif, L. Fu, and L. B. Kara, "The potential for computer tutors to assist students learning to solve complex problems," *Proceedings of the American Society for Engineering Education Annual Conference and Exposition, Indianapolis, Indiana, June 2014*.
- [9] L. B. Kara and T. F. Stahovich, "Hierarchical parsing and recognition of hand-sketched diagrams," *Proceedings of the 17th annual ACM symposium on User interface software and technology (UIST '04)*, New York, NY, pp. 13-22, ACM, 2004.
- [10] J. LaViola, "Advances in mathematical sketching: moving toward the paradigm's full potential," *IEEE Computer Graphics and Applications*, vol. 27, no. 1, pp. 38-48, January/February 2007.
- [11] J. Peschel and T. Hammond, "STRAT: a sketched-truss recognition and analysis tool," *International Workshop on Visual Languages and Computing*, Boston, MA, pp. 282-287, 2008.
- [12] W. Lee, R. de Silva, E. J. Peterson, R. C. Calfee, and T. F. Stahovich, "Newton's Pen: a pen-based tutoring system for statics," *Computers & Graphics*, vol. 32, no. 5, pp. 511-524, 2008.
- [13] L. Fu and L. B. Kara, "From engineering diagrams to engineering models: Visual recognition and applications," *Computer-Aided Design*, vol. 43, no. 3, pp. 278-292, 2011.
- [14] J. Sweller, "Cognitive load during problem solving: effects on learning," *Cogn. Sci.*, vol. 12, pp. 257-285, 1988.
- [15] J. Sweller, "Cognitive load theory, learning difficulty and instructional design," *Learning and Instruction*, vol. 4, pp. 295-312, 1994.
- [16] M. T. H. Chi, M. Bassok, M. W. Lewis, P. Reimann, and R. Glaser, "Self-explanations: how students study and use examples in learning to solve problems," *Cogn. Sci.*, vol. 13, pp. 145-182, 1989.
- [17] P. S. Steif, L. Fu, and L. B. Kara, "Development of a cognitive tutor for learning truss analysis," *Technical Report, Carnegie Mellon University*, 2013, unpublished.
- [18] P. S. Steif, "An articulation of the concepts and skills which underlie engineering statics," *34<sup>th</sup> ASEE/IEEE Frontiers in Education Conference*, Savannah, GA, 2004.
- [19] K. R. Koedinger, R. S. J. d. Baker, K. Cunningham, A. Skogsholm, B. Leper, and J. Stamper, "A data repository for the EDM community: the PSLC DataShop," In C. Romero, S. Ventura, M. Pechenizkiy, R.S.J.d. Baker (Eds.). *Handbook of Educational Data Mining*, Boca Raton, FL: CRC Press, 2010.
- [20] MATLAB, version 8.1.0.604 (R2013a). The MathWorks, Inc., Natick, Massachusetts, 2013.
- [21] K. L. Draney, P. Piroli, and M. Wilson, "A measurement model for complex cognitive skill," In P. Nichols, S.F. Chipman, and R.L. Brennan (Eds.). *Cognitively diagnostic assessment*, Hillsdale: Erlbaum, pp. 103-126, 1995.



# Digital Agenda and E-learning in Italian Public Administration

Milena Casagrande, Luigi Colazzo  
Department of Industrial Engineering  
University of Trento  
Via Sommarive, 2 - Trento  
{milena.casagrande, luigi.colazzo}@unitn.it

Andrea Molinari  
Department of Economics and Management  
University of Trento  
Via Inama, 2 - Trento  
Andrea.molinari@unitn.it

**Abstract—** In this paper we present a long experimentation conducted by the Trento Chamber of Commerce in the usage of e-learning and SCORM-based material in order to introduce companies to innovative practices, specifically the use of the Digital Signature and the Certified Electronic email. These two tools have become now mandatory in the Italian Public Administration, thus forcing not only this component of the society to pursue innovation through dematerialization processes. It also, and mostly, involves millions of citizens and companies that, from now on, will have the opportunity to interact with public bodies in a more efficient and effective way, through the use of these innovation. Neither the Public Administration nor most of the companies (especially SMEs) were ready to this radical change of using electronic communication. Thus an enormous effort of training will be needed to allow everyone to be aligned with concepts, tools, modalities and procedures to be used in order to take advantage of the enormous potential of citizen-to-government and business-to-government communication. In this paper we present our experience in the design, creation and delivery of distance learning material designed with the Chamber of Commerce of Trento, through the description of the project started in 2011 regarding the issues of dematerialization and digitalization of the relations between business and Public Administration.

**Keywords-** *e-learning, Public Administration, dematerialization, learning objects*

## I. INTRODUCTION

In 2005, many relevant and radical changes have been introduced by the Digital Administration Code (CAD) in Italian Public Administrations (PAs), with the precise objective of changing processes through dematerialization and ICT innovation. Consequently, most of the Italian citizens and organizations that have to deal with Italian PA, will have an impact on their own processes, not only because of the constant interaction with the PA itself for different reasons (from tax payments to requests). This revolution in Italian society will influence also the way citizens and companies will exchange information in the future.

The Italian Government, since 2005, is working on the “Codice dell’Amministrazione Digitale” (CAD - The Digital Administration Code), in order to modernize and digitalize processes of our Public Administrations (PA). Subsequently

integrated and amended, the CAD is now fully operational, and recently (December 2012) a new version with important novelties has been emanated. Despite its name, CAD applies to both private and public bodies, and it states a completely new scenario for the usage of ICT, in such a way that PAs can be more efficient and reactive in the relationship with their stakeholders. Among the many novelties, four are particularly relevant:

- a) the use of electronic signature for signing electronic documents;
- b) the use of certified emails;
- c) the use of digital protocol to track in/out movements of documents;
- d) rules about digital preservation.

Moreover, in the very near future (March 2015), any invoice sent to Italian PA will have to be an “electronic invoice”, following a precise XML schema and sent through specific, certified channels. The timing of implementation are certainly not instantaneous, but the change of mentality and especially organizational as well as technological and regulatory environment, will have a tremendous impact on Italian society. The Italian PA is so moving in this direction, where digital signature and PEC have already made a relevant progress in terms of implementation, while the other two themes are taking off. It is certain that the administrative procedures will be exclusively managed in the future through electronic transmission; interaction with citizens and businesses will be turning more and more in this mode, and this will profoundly change the Italian PA on one side, but will also have a deep impact on the lives of citizens and businesses. These elements radically change the scenario for ICT usage inside organizations, in particular for PAs. Besides a design, realization and reengineering of information systems and applications supporting such new modalities, reengineering of business processes and procedures will be needed. Nevertheless, public servants are used to perform their working processes without technologies. In case of inefficiencies, they are requested to contribute with their knowledge through the use of CAD prescriptions. These people are normally expert in their field, but probably not so expert in the specific domain of the CAD. The main expertise is hidden, not shared, not formalized, not

clarified: from tacit knowledge, PAs will have to extract this information and transform it into explicit knowledge. Finally, the Italian legislation stimulates the sharing and reuse of any educational resource produced by/for the PAs, in order to optimize investments and maximize uniformity of view about the topics.

In order to face this situation, an obvious need of training for millions of citizens, public and private employees or organizations is needed. Often, companies are reluctant to adopt innovative solutions, stopped by the cliché of new technologies capable only to create complex and critical processes, especially if this innovation stimulus comes from the PA. The question now is no longer in these terms: companies have to adapt to what the CAD indirectly requires them, or they risk not only to stay out of a process of modernization that paradoxically involves and is generated by the PA, but that is likely to have consequences in terms of fines and sanctions.

Real problems related to the impact of the CAD over the Italian society have always been clearly identified as a training problem. Organizational and technological solutions outlined by CAD (just think about the problems and technologies needed for the qualified digital signature) have clearly demonstrated the backwardness of large digital layers of Italian society. On the other side, most of the organizations were convinced that this innovation process is part of the Italian habit of proposing innovation and not realize it, thus ignoring the innovation coming.

Together with this initial approaches, it is clear that the processes of dematerializing documents is irreversible, and the best has yet to come

One of the points of reference in dealing with this issue of the relationship between CAD and businesses is certainly the Chamber of Commerce. The problem that we set in 2011 with the group work of the Chamber of Commerce of Trento, when it became clear the state of maturation of the CAD (which in the meantime, after the initial formulation of 2005, was consolidated and gave clear indications of the direction taken) was pretty simple. We wanted to help companies to improve certain processes without distorting them, bringing advantages in terms of speed, efficiency and cost containment through the application of CAD dictates.

The Chamber of Commerce of Trento, together with trade associations and the Province of Trento, used until 2011 some tools to raise awareness on the topic: sending printed material and newsletters, organizing dedicated events, publishing articles in newspapers or on magazines with dedicated spaces, opening specific spaces in the respective website, launching various initiatives and seminars. All these initiatives have reached, over two years, about 300 companies in our region. It seemed clear, however, the impracticability of reaching a larger number of organizations on schedule for the CAD deadline, and thus e-learning has naturally become the instrument in order to achieve:

- low costs and at appropriate times,
- a broad audience [8][6][3].

In this article we present experiment carried out by the research group of the Laboratory of Maieutics University of

Trento, together with the Chamber of Commerce of Trento and the Province of Trento, for the realization of a wide range of learning objects [4] [12] available for companies of Trentino and accessible through the e-learning platform developed by our group.

## II. LEARNING SOLUTIONS IN THE PA

Looking at the results presented in the European Digital Agenda (EDA) Scoreboard 2012, the needs of e-government seem to be very clear in Europe, keeping into consideration all the stakeholders involved. The European Digital Agenda (EDA) has the target of increasing regular internet use to 75% of the population, the confident projections of last year's scoreboard have been reviewed -- the 75% target will most likely be reached in 2014 and not in 2013, although still ahead of the EDA target year of 2015.

Even though the usage of eGovernment by citizens is stable at 41% with some significant progress in smaller countries, the main reasons of resistance for non-use of online public services by citizens are

- a) lack of need
- b) lack of trust
- c) lack of skills.

Nevertheless, digital agenda is pushing towards the adoption of ICT-related procedures and process re-engineering, citizens and firms will steadily increase to require eGovernment services (usage by firms has increased steadily from 76% in 2010 to 84% in 2011), and civil servants will be forced to adopt ICT not only for their ordinary work, but also to improve their processes according to ICT availability. This is a great effort, as the incomplete digitalisation of public services is an important barrier to an increasing eGovernment take-up. The initiative "EPSA Trends in Practice-Driving Public Sector Excellence to Shape Europe for 2020" demonstrated the need of integration, contacts and networks to be established among those Public Administrations that are able to showcase their achievements in terms modernisation, innovation and smart solutions for the Public Administrations facing with budget constraints, increasing demands from citizens and the need to at least maintain if not increase and improve their service delivery: the milestone for this is the availability of the EPSA learning platform.

Some market researches demonstrated that almost 90% of public sector Learning Departments plan to increase the use of TEL in order to meet cost reduction targets set in the Comprehensive Spending Review (CSR) (<http://www.brightwave.co.uk/>). While at the beginning of 2009 only 50% of L&D managers expected to increase their use of e-learning, 88% currently anticipate an increase, and more trainers expect to use collaborative learning techniques. In this sense, the current economic period of austerity is clearly having a deep impact on these considerations, with no much distinction between private and public sector.

The examples reported in market research are very clear in this sense. The City of Edinburgh Council Interactive Learning (CECIL) started four years ago providing online learning for all

employees, reducing training costs by more than £800,000. The conditions under which to provide these savings are well known: high volume training to be delivered in a short period of time.

Learning solutions for public sector have been promoted in the past in many different ways, with results that do not meet expectations. During recession periods like the one we are living, Technology-enhanced learning (TEL) has been indicated as a panacea for future learning and as a sort of “killer” of training in a classroom, but at the same time, many times in its history TEL has not been living up to the hype. The public sector expects to use e-learning to improve the level of service delivered to organizations despite budget cuts. This seducing metaphor attracted the public sector, fitting well with the push towards e-government. Reality showed a different picture. TEL has a lot of researches and application studies in the public sector, if we consider as “public sector” the educational institutions from primary schools to master degrees. However, if we exclude this relevant part of use cases, the traditional Public Administrations have not been investigated too much as possible application field of TEL-based training programs. There are many examples of training employees with TEL that seem to be pushed by the idea of cost saving, or by the idea of substituting missing educational paths (like for example, qualification of civil servants that have no degree in their curriculum).

Most of use cases are concentrated in year 2000-2006: different experiences coming from different countries, but not a unique vision of the specificity of public sector. So, in terms of state-of-the-art of TEL in public sector, the situation is not so clearly identified like in the private sector, or in the traditional educational system. Many researchers and many studies have concentrated their attention to the application of TEL to institutional training, supporting traditional training with technologies, methodologies and tools. On the contrary, little attention has been devoted in the past to the specificity of TEL inside Public Administrations and their employees, with their specificity and needs.

Public Administrations have been involved, together with their employees, in TEL applications mainly for lifelong learning projects, considering the public servants as persons that will averagely stay for a long period of time in the same workplace, and therefore as interesting case studies for lifelong learning researches and applications. Nevertheless, even if we exclude educational institutions from the application field, the public sector is a place where all theories, methodologies and tools studied and implemented in TEL could be profitably applied. Most of the new approaches and tendencies in TEL, like story-telling, MOOCs, gamification etc. could be applied to the large ecosystem of public bodies’ stakeholders. Public Administrations are particularly interested in TEL because they have recognized e-learning and web-based learning resources as fundamental elements of their training processes, mainly for the capabilities of delivering educational contents to participants over the Internet anytime and anywhere at competitive costs.

Nevertheless, most of these well-known problems have now been overcome: it therefore seems that the hesitant progress of TEL – always on the point of spreading extensively and then for some reason never really succeeding – should not have to continue any longer, given the disappearance of most of the

technological barriers. Other obstacles have been cleared, i.e., standardization and reuse of learning material. Almost any public institution has at least considered using one of the different available approaches, blended or full online, and tools available today can guarantee the service level required by Public Administrations: platforms like Learning Management Systems-LMS, technologies like videoconference, standards for learning metadata or objects like SCORM or LOM [11], LTSC [12]. Another effect of the maturity of TEL for the public sector regards the large amount of educational material that has been produced and that is now available under various forms. Several Learning Objects (LOs) have been created by various institutions, mostly available for free. These elements combined with reusability, cost reduction, optimization of time spent away from the workplace and a more modern view of the public body, all these elements have contributed to the creation of relevant expectation about TEL in public sector. In terms of factors that influence TEL adoption inside public bodies, some researches conducted in Public Administrations indicate that civil servants’ behavior respect to the application of e-learning was significantly influenced by the satisfaction, this in turn being affected by:

- job relevance;
- expectation confirmation;
- perceived ease of use;
- perceived usefulness;
- computer self-efficacy.

Other studies show that adult learners perceive positively the effectiveness of e-learning in the workplace.

### III. DIGITAL AGENDA AND THE ITALIAN DIGITAL ACT

Nowadays, resources wasting prevention is a must for every Public Administration, and the digitization of processes in order to replace (among the others) traditional paper-based procedures is an opportunity to contribute to this prevention. The need for process digitization, especially in Public Administrations, was recognized since the '80 and under different labels is one of the areas where ICT can produce evident advantages. In this field, the term “dematerialization” has been used to identify the progressive elimination of paper-based processes in favor of their digitization. Following this direction, since 1997 the Italian PA has undertaken the complex task of creating a legal and normative framework that facilitates digitization of processes through the modernization of the PA information systems. In 2005, the “Digital Administration Code”, (CAD - Codice dell’amministrazione digitale, D. lgs. 7.3. 2005, n. 82 as modified by D. Lgs. 30.12.2010, n. 23), defined very clearly and extensively what should have been done in order to create a “digital Public Administration”. It states a completely new scenario for the usage of ICT, in such a way that PAs can be more efficient and reactive in the relationship with their stakeholders. The CAD became effective for all Italian PA, but despite its name, it involves both private and public bodies, in the end all Italian citizens and organizations. It is intended to reestablish tidiness in addressing and setting rules for every aspect of the technological innovation introduced in the Italian

society. There is a strong commitment inside CAD towards dematerialization, and this is the point where our use case wants to intervene through e-learning. More specifically, the CAD provides legal validity to digital documents involved in processes and procedures conducted through the usage of ICT. Among the many novelties, four are particularly relevant:

- the use of electronic signature for signing digital documents. This graduates the probatory effectiveness of the different type of electronic signature, specifically the “advanced electronic signature”, that plays a leading role on the Italian market today in certifying the identity of document’s subscriber/s;
- the use of certified emails. The certified email (Italian acronym PEC) provides citizens and organizations with legally valid electronic documentation of the sending and delivery of electronic documents to certified receivers, and conversely, receivers have legal evidence about who is the sender of a certified email;
- the use of digital protocol to track in/out movements of documents, thus allowing not only the certification of these movements across the Information System, but also providing a way to centralize the storage of documents officially sent and received by the organization;
- rules about digital preservation of electronic documents. The topic is a specific instance of the larger problem that today worldwide organizations are facing. According to one of the many, well-known definitions (<http://www.ala.org/alcts/resources/preserv/defdigpres0408>), digital preservation “combines policies, strategies and actions that ensure access to digital content over time”. In the Italian CAD, this problem regards every single organization that undertakes the CAD prescriptions. It is the logical and inevitable closing item of the full, automated process of digital communication. Nevertheless, it introduces a lot of technical issues that heavily impact the day-by-day activities of the organizations: ICT infrastructure to preserve contents, methods to search and retrieve digital objects etc., but most of all, competencies and people specifically devoted to and skilled on digital preservation.

These four elements are clearly revolutionizing the Italian Public Administrations’ processes, allowing new scenarios for the interaction between PAs (G2G), PAs and citizens (G2C), and PAs and companies (G2B). Specifically, these innovations have an enormous, unexplored (so far) potential in the re-engineering of the PAs processes, thus allowing obvious but crucial advantages, like relevant savings, flexibility in managing processes and relationships, velocity in fulfilling requests for services, openness respect to a procedure (especially in front of citizens), measurability of results and performances of individuals and public organizations(KPIs).

Besides a design and realization of applications supporting such new modalities, a deep learning intervention should be performed in order to take advantage of the opportunities created by the CAD:

- new rules, new systems and new procedures must be acquired by all civil servants acting inside PAs;

- reengineering of business processes and procedures will be needed.

Notoriously, the problem for the Italian PAs is not in the lack of people able to optimize and improve processes. The lack is in the capabilities of Italian PAs and citizens to metabolize the innovations that digitization and dematerialization offer, because of an endemic resistance to changes. The following table presents the situation of PEC adoption in some of the most “efficient” Italian provinces, where PAT is in the top positions (source: DIGITPA, 2013)

Chamber of Commerce	N. Of companies	With PEC	%
Cuneo	24.843	22.569	91%
Sondrio	6.543	5.920	90%
Bolzano	21.494	19.351	90%
Forlì	19.834	17.526	88%
Trento	22.297	19.632	88%
Pordenone	12.054	10.544	87%
Prato	16.644	14.461	87%
Bergamo	47.636	41.036	86%
Belluno	6.992	6.023	86%
Mantova	17.649	15.170	86%

Figure 1. Table 1: adoption of Italian certified emails among large corporations and SME

Over a national average of 74% of adoption of certified email, PAT has the 88% of organizations that formally adopted this tool. But if we look at individual companies, the situation radically changes (Table 2).

Chamber of Commerce	Individual Enterprises	With PEC	%
Firenze	53.909	2.399	4%
Prato	16.506	704	4%
Livorno	18.458	781	4%
Rimini	19.944	834	4%
Teramo	21.643	853	4%
Crotone	12.276	482	4%
Ancona	26.885	1.047	4%
Pistoia	18.227	699	4%
L'Aquila	17.455	664	4%
.....	.....	.....	.....
Trento /PAT→ ranked #103	29.423	359	1%

Figure 2. Table 2: adoption of Italian certified emails among individual enterprises

Considering that individual enterprises most of the times correspond to individual citizens, it is clear that Italy has a serious problem in the diffusion of CAD digital agenda. Besides, PEC is for sure the easiest of the four elements promoted by CAD, because of the similarities with a traditional e-mail box management.

Another interesting element of analysis regards the situation of the traffic of certified email during the last six years (source: DIGITPA, 2013). This clearly shows an increasing adoption of CAD tools, that could be a sign that few organizations/individuals are using these tools, but they are using in progressive and extensive ways. Interesting to note that from 2007, PEC domains and mailboxes have increased respectively of 27 and 45 times, but the number of PEC messages have increased of just 4 times. This is a clear effect of the compulsoriness of CAD prescription, but not representing an intimate and convinced adoption.

Italy has now the legal framework, and ICT provides the technological tools for a take-off in the adoption of CAD prescriptions. The missing link now is a pervasive training initiative devoted to individual users and organizations about the tools introduced by CAD. The promotion of this cultural growth of Italian citizens and organizations is clearly in charge of the PAs.

YEAR	Tot.Domains	Tot.Mailboxes	Tot.Messages	YEAR	Tot.Domains	Tot.Mailboxes	Tot.Messages
2007	42.369	618.165	116.376.864	2007	100%	100%	100%
2008	111.244	1.147.208	218.477.050	2008	263%	186%	188%
2009	295.220	3.943.160	253.098.716	2009	697%	638%	217%
2010	578.258	11.518.079	327.476.760	2010	1365%	1863%	281%
2011	840.404	17.797.879	324.125.539	2011	1984%	2879%	279%
2012	1.164.829	28.297.727	459.662.512	2012	2749%	4578%	395%
Grand Total	3.032.324	63.322.218	1.699.217.441				

Figure 3. Table 3: the diffusion of PEC domains, mailboxes and messages in last six years (DIGITPA, 2013)

#### IV. CREATING LEARNING OBJECTS FOR THE ITALIAN PA

Our group has a long experience in e-learning applications, since 1998 specifically in Public Administration area. We currently are delivering e-learning initiatives with many public and private partners, among which the Autonomous Province of Trento (~12.000 employees), Trento Chamber of Commerce (~55.000 individual and small/medium Enterprise' users), Trentino Development Agency (~1.000 users/companies), the Academy of Commerce and Tourism (~1.000 users/companies) and of course the University of Trento (~ 15.000 users). Through our experience of e-learning in PAs, we have proven that excellent results could be obtained not only using e-learning as a substitute of traditional classroom, but also involving and motivating the public servants directly in the creation of the learning objects. These results can be obtained if some elements are available during the creation of the educational path:

- a common methodology for gathering, formalization and distribution of requisites, avoiding their uncontrolled production;
- a different approach respect to traditional F2F interaction with teachers/experts, less formal and "boring" respect to PowerPoint-like presentations;
- a virtual place with a strong collaborative connotation, where to share ideas and results with other colleagues, in order to compare different views and interpretation of the process/procedure, especially from a legal point of view;
- the availability of consultants on legal topics, able to provide straight-to-the-point suggestions: what is needed is more a consultant than a teacher;
- a way to see their activities recognized and rewarded, at least in terms of reputations: trivially, a formal way to recognize the work done in e-learning material production.

Many learning packages and training initiatives have been started after the advent of the CAD, but no tangible results are visible yet. As a direct experience, we decided to specifically develop e-learning SCORM packages regarding all the CAD topics, together with the Chamber of Commerce. We have produced 14 hours of learning objects explaining in details the two tools that are mandatory for companies today, i.e., digital signature and certified email. We launched this initiative in 2011, together with PAT over the Trentino territory, also

involving the various professional associations. In PAT, over approx. 55.000 enterprises (public and private large corporations, SME, individual companies), approximately 700 participated to this joint initiative PAT-Chamber of Commerce-University of Trento, with the fruition of this e-learning material completely free-of-charge. In order to interpret this data correctly, consider that most of public servants are used to use our platform for e-learning activities, and most of the professional associations delivered courses to their companies using this modality: e-learning could not, therefore, be considered the responsible of these results.

After this, we tried to change the approach: in the last months of 2012 we experimented a prototypical approach to e-learning creation, for some associations related to the local Chamber of commerce, regarding three processes:

- authorizations to the opening of new hotels in touristic areas;
- registration of a new company to the Local Registry of Enterprises;
- cancellation of a company from the Local Registry of Enterprises.

The material, this time, has been created by our experts, but with a deep interaction with the public servants that are in charge (and know very well) the above processes. Preliminary results are clearly showing that this change of gear, is producing significant appreciation from end-users. After having tested other methods, like the creation of learning objects by academic experts, it is clear from our empirical evidences that learning objects created by people taken from the inside of the organization with a high reputation, are by far more appreciated and trusted, most of all for their capabilities in proposing a re-engineering of the whole process.

We are therefore injecting a new paradigm of delivering training to public servants, citizens and companies by radically changing the subjects that produce, and the methods of production, sharing and use of the e-learning material. Nevertheless, it is clear that the level of complexity in process re-engineering caused by CAD is perfectly manageable by these "prosumers" when supported on purpose by experts that act more as consultants than teachers.

The biggest obstacle has of course occurred in the design phase, when turning materials provided by experts into learning objects, both in terms of estimating the effort of our development group to be shared with the Chamber of Commerce of Trento, and for the quantification of the rewarding to be recognized to authors of the materials and the learning objects. This is the typical case of industrial production of learning objects where a "teacher" should be recognized a number of hours by far superior to the total duration of the learning objects produced. We solved this issue by adopting the cost for creation of online learning objects presented in [5][10]. The two themes "Digital Signature" and "Certified Electronic Mail" were treated separately, to give different ways to adapt the training needs to any pre-existing knowledge or skills already present in the organization. For each item we provided three distinct levels of learning, adaptable to the needs and to specific questions. The material is in fact accessible either by following a sequential

approach proposed to the user, and in terms of an application-driven approach that allows the user to immediately identify topics of interest.

The levels of meta- organization of content are three:

1) "informative" level, aimed at disseminating the knowledge base on the issues, and then answer the questions with respect to the sense and meaning of the two instruments in acting professionally, by clearly expressing the benefits and points of attention resulting from its use (as well as a stimulus in the questions raised during the training sessions prior to the draft e-learning and during telephone follow-up triggered by the Chamber of Commerce on a sample of 50 firms )

2) "usage" level, designed to provide practical tools for the two topics, with the use of tutorials that guide you step by step in the activation of the instrument, in its use, in the verification of correctness of the operations. At this level, we used recorded sessions of interaction with the software tools to digitally sign and send certified emails;

3) "in-depth" level, devoted to users or potential users of the two instruments who wish to become more aware of the implications deriving from their usage, particularly in normative terms. The normative part was in fact reduced in scope during the design of learning objects described above.

## V. CONCLUSIONS

Up to now, more than 700 companies have benefited from the material made available on the platform and have gradually reported needs compared to having other types of content available. For this reason, from 2012-2013, the Chamber of Commerce has decided to make available to other communities some specific topics, content related to the aspects that collected more questions or more errors. The Chamber of Commerce, due to the very positive feedbacks received both for the approach and the contents, has decided to invest in e-learning in the creation of video lessons related to the common practice of "balance sheet management" and "Business Register". The Chamber of Commerce is considering how to reuse the material produced on the occasion of blended learning, which could further encourage companies to participate. We are together conducting further actions to raise awareness of companies, including through the evaluation of a partnership with the Association of Public Accountants that are the prime reference in the report by the Public Administration for many small and medium entrepreneurs. The study presented is a clear confirmation of the characteristics used to promote the use of e-learning to situations as training/outreach otherwise unobtainable, even if it has the specificity that we consider interesting to be replicated in other contexts:

a) we paid particular attention to the micro-design of contents, scheduling single items in a very fine-grained way;

b) we involved users in a virtual community closed to only members and key stakeholders, in order to increase the quality of the final results;

c) we used a platform designed structurally to the idea of managing a virtual community, rather than simply manage a classroom or a training event [7];

d) we enhanced the remuneration of the actors involved (teachers, content experts, e-tutor) using a cost model specifically designed for e-learning by enhancing the evaluation of the time spent in producing learning objects effort;

e) we enabled a perspective of lifelong learning, imagining a path that includes the migration of many of the training activities, information and update of the Chamber of Commerce of Trento towards these tools. This clarified the benefits that have always been professed by e-learning researchers, but which now seems even more indispensable for any Public Administration involved in the digital modernization

## REFERENCES

- [1] [1] Adelsberger H. (2003), "The Essen model: a step towards a standard learning process," <http://citeseer.ist.psu.edu/515384.html>, 2003.
- [2] [2] Boyd D. (2006), Friends, friendsters and mySpace Top 8: writing community into being on social network sites. *First Monday* 11(2), December.
- [3] [3] Stamatiou A., Tsihrantzis G.. "Education and Assessment of Civil Employees in e-Government: The Case of a Moodle Based Platform." *interaction* 4: 5.
- [4] [4] Brooks, C. and McCalla G. (2006), Towards flexible learning object metadata. *International Journal of Continuing Engineering Education and Life-Long Learning*, 2006. 16(No.1/2): pp. 50 - 63.
- [5] [5] Casagrande M., Colazzo L., Molinari A. (2013), Estimating the effort in the development of Distance Learning paths, ICE-B - 10th International Joint Conference on e-Business and Telecommunications, Reykjavik (Iceland), 29-31 Jul 2013
- [6] [6] Colazzo, L., Molinari, A., & Villa, N. (2011). Formal and informal lifelong learning in a virtual communities platform. *Lecture Notes in Computer Science including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*.(2011), Volume: 6537 LNCS, Pages: 291-300
- [7] [7] Colazzo, L.; Molinari, A.; Villa, N.(2009), "Collaboration vs. Participation: The Role of Virtual Communities in a Web 2.0 World," *Education Technology and Computer*., ICETC '09. International Conference on , vol., no., pp.321-325, 17-20 April 2009
- [8] [8] Moore, M. G & Kearsley, G. (2012). *Distance education: A systems view of online learning*. Belmont, CA: Wadsworth, Cengage Learning
- [9] [9] Rennie F., Morrison T.. *E-learning and social networking handbook: Resources for higher education*. Routledge, 2012.
- [10] [10] Casagrande M., Molinari A., Tomasini S. (2010), Progettazione e intervento nell'e-learning per la Pubblica Amministrazione: dalla sperimentazione all'analisi dei costi. *Je-LKS, Journal of E-Learning and Knowledge Society*, rivista della SIE-L, Società Italiana di e-Learning. Vol. 6, N. 1, Gennaio 2010. ISSN: 1826-6223
- [11] [11] LOM: IEEE Learning Object Metadata standard <http://ltsc.ieee.org/wg12/>
- [12] [12] LTSC. (2011) IEEE Standard for Learning Object Metadata, 1484.12.1-2002, IEEE LTSC, 2002.
- [13] [13] Rheingold H.(2000) *The Virtual Community: Homesteading on the Electronic Frontier*, The MIT Press; revised edition edition (November 1, 2000)



# Castor: Designing and Experimenting a Context-Aware Architecture for Creating Stories Outdoors

Fabio Pittarello and Luca Bertani

Università Ca' Foscari Venezia  
Mestre (VE), Italia  
{pitt, lbertani}@dsi.unive.it

**Abstract**—This work describes the design and the experimentation of Castor, a novel architecture for creating, editing and delivering engaging geolocalized context-aware stories outdoors. Castor is one of the first architectures that enable the direct creation of structured stories in-situ, rather than the simple gathering of material, and that use an extended set of context dimensions (i.e., environmental and social context) for augmenting the emotional engagement of the story listeners. The architecture was experimented in an educational project lasted two semesters, performed integrating the traditional educational path of a class of children aged 7, with the full collaboration of their teachers. Our architecture demonstrated to have an important role for bridging the gap between the structured classroom learning and the outdoor experience, engaging the children for obtaining interesting results in terms of acquisition of new skills, collaboration and social inclusion.

*Context-awareness; Education; Learning; Storytelling*

## I. INTRODUCTION

This work presents a novel software architecture, named Castor (i.e., Context-Aware STORytelling), for supporting all the phases of creation, editing and delivery of context-aware stories on the field. Compared with the available literature, the value added by Castor is the possibility to create stories compliant with a novel story model, where the narration is driven by the context conditions, matched with those ones specified by the story author for the delivery of the story. While other storytelling systems enable only the simple gathering of materials on the field for creating the narration elsewhere or limit the use of the context to the location, Castor enables the direct creation of narrations in-situ and uses an extended set of context dimensions (i.e., the weather conditions, the season, the time of the day and the number of listeners), for augmenting the engagement of the listeners. The architecture was tested in the context of an educational project that involved for two semesters a class of 19 children aged 7 in which the traditional educational path was integrated with the modern mobile technologies. The educational project was performed with the constant collaboration of the teachers. During the experimentation the class learned how to build, modify and listen to engaging context-dependent stories with the help of different tools, in classroom and outdoors. All these activities were composed in a smooth educational path that showed the positive role of our architecture for the acquisition of new skills and for bridging together the work done in the classroom and outdoors. The availability of a mobile platform permitted us to bring a structured learning experience outdoors, allowing to

capture easily the creative activity of children and enabling the acquisition of literacy skills. The novelty of Castor lays also on its social dimension, being conceived as a social repository and publishing platform for context-aware narrations that can be accessed by pupils for sharing new stories and listening to the available ones. Children involved in the experiment learned how to build stories, starting from the traditional model to the context-aware model, and then shared their creations with their fellows, that listened to the narrations in the locations where they had been created. The analysis of the classroom and outdoors activities showed another benefit deriving from the use of Castor: an improvement of the levels of collaboration and social inclusion outdoors. The teachers confirmed this significant improvement, considering also the behavior of the children during the ordinary classroom work. This represented a confirmation of the role of the environment for learning and the importance of having tools, like Castor, capable of bridging the gap between structured learning and outdoor experience. The rest of the work is organized as follows: Section II will discuss the related work; Section III and Section IV will give an overview of the story model and of the software architecture; Section V will present the user interfaces of Castor; Section VI will present the five phases of the educational project; Section VII will discuss the findings; Section 8 will draw the conclusions.

## II. RELATED WORKS

There are a number of models and architectures for computer-enhanced storytelling [1]. Most of them rely on the analysis of narratology theories, which study the structure of stories. This project relies on a story model based on the work of the Italian researcher Cesare Segre [2], chosen because of its generality and suitability to different literary genres and adapted to interactive storytelling. The derived software architecture is compliant with the “drama manager approach”, where a software architecture controls the narration on the basis of the story model and of the narrative choices of the author [3], opposed to the “autonomous agents approach”, where a set of software agents influences the evolution of the story [4]. We made this choice because we were more interested, as most academic researchers involved in the development of storytelling systems for educational purposes [5], to implement a system supporting the children creativity and expressivity rather than to generate stories with a higher - but probably less interesting - number of choices and endings. Storytelling has always been a powerful means in the educational curricula, not only for developing literary skills,

but also for improving the interest of children for other educational domains. Personal computing has been used in the last years to support the creation of stories in the classroom, permitting children to collaborate at various levels, in classroom or even from remote locations [6]. The collective authoring of stories composed by a whole class is one of the key points of the work proposed by Di Blas et al. [7], that focuses also on the integration with the children curricular activities and the inclusion of pupils otherwise marginalized. In recent years the rapid evolution of mobile technologies has permitted to support outdoor activities, such as fieldtrips [8]. Researchers have considered different educational domains, including history, geography and science. For example, in Ambient Wood [9], children provided with mobile devices explore a wood enhanced with ubiquitous technologies for gathering ecology data. Halloran et al. [8] [10] designed technology-enhanced fieldtrips for supporting the learning of literacy skills. The children were involved in the exploration of a historic English country building, Chawton House, supported by mobile devices for listening to the content prepared by curators and teachers and for gathering their own content (e.g., audio and photographic snapshots) at specific locations. They used also paper and pencils for fixing their thoughts. A peculiarity of this work is the integration with the classroom work: the children collected content and ideas outdoors, but wrote their stories the following day, when they returned to their classroom. Hansel et al. [11] take advantage of the Mobile Urban Drama [12] platform for designing an environmental drama where storytelling and study of natural sciences are mixed. The students listened to the narration delivered by mobile devices, but at the same time were asked to perform different assignments, involving sketching, collecting soil tests and taking pictures. As in [8], the students completed their work in classroom, in this case producing reports and presentations on the basis of the materials collected. The authors of the study underline the importance of learning outdoors, that stimulates the practical intelligence instead of the theoretical school intelligence [13] and the use of mobile devices that contribute to bridge the gap between symbol manipulation and contextualized reasoning. Both the [8] [11] approaches offer a very structured experience in terms of the content delivered to children, for inspiring the creation of stories or for stimulating the gathering of data. For what concerns the creation phase, they are more focused on the gathering of materials and on the solution of assignments rather than on the creation of narrations in situ. Fails et al. [14] propose a complete mobile system supporting both the creation and the listening of stories in-situ. The approach is focused mainly on the collaboration of children and uses the surrounding environment as a stimulating scenario for engaging them, even though the different parts of the narration have a loose coupling with distinct locations. Compared to previous literature, the project described in this work reserves a keen attention to the integration of classroom and outdoors activities, and provides a set of tools for the direct creation of the story in-situ, geolocating precisely all the story locations. Context-awareness concerns the capability of computer systems to log different dimensions of the context (i.e., according to Dey et al. [15], “any information that can be used to characterize the situation of an entity. An entity is a person, a

place or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves”) and to use it for guiding and adapting the user experience. While the dimensions of the context are several and include the location, the environmental conditions, the user profile and history of previous interactions, the social conditions and the technological features of the devices (e.g., the width of the screen or the network bandwidth), only a part of them have been extensively used in context-aware architectures. Concerning the applications, in most cases the knowledge of the context has been used for delivering content and for lowering the cognitive load of the user rather than for other purposes. There are several implementations related to different interaction paradigms, including the hypermedia [16] [17], the web and the mixed reality. The implementations involve also different domains. As far as the cultural heritage is concerned, there are museum guides [18] or educational games for enhancing the visits of scholars to archaeological sites [19]. As far as tourism is concerned, Medina [20] uses multiple dimensions of the context, including the location, the device, the user profile, the network and the time for triggering the presentation of hypertextual information. The iLand platform [21] delivers to mobile users geolocalized content related to the oral culture and traditions of the island of Madeira. In [22] the user history is used for enabling proactive presentation of content in a virtual fair. The available implementations for the educational domain often limit the use of the context to the location [8] [11], for delivering appropriate content and augmenting the user engagement. The importance of the location for the user engagement is underlined in [10] also for the authoring phase: the curators that prepared the content for the children involved in the exploration of the Chawton House stressed that they could make a rich and detailed audio recording related to a specific location only because they were there. Engagement and stimulation of the user emotions are becoming important parameters of the user experience as the interaction paradigms shift from the traditional working environment and melt with the everyday life. For expressing the attention to these new aspects, Picard coined in 1995 the term “affective computing”, for denoting “computing that relates to, arises from, or influences emotions” [23]. As far as storytelling and education are concerned, StoryFaces [24] is an interesting project that tries to augment the emotional engagement of the children permitting them to record facial expressions that become part of the narration. We share with the research related to affective computing the interest for the design of computing systems that permit to influence the emotional engagement of users. That is the reason why in our work we designed a software architecture that uses not only the location, but also an additional number of context dimensions for augmenting the user engagement, with the final goal of obtaining educational benefits.

### III. THE STORY MODEL

For the whole project we refer to a story model adapted from the work of the Italian researcher Cesare Segre [2] and described in [25]. The model is based on the identification of the “stages” where the narration happens and the association of narrative fragments of the story to these locations. A story is composed by one or more ordered stages. The author of the



Figure 1. Visual representation of context dimensions and values

story, during the composition of the stages, decides the values of the context parameters for the delivery of the narration, that include the time, the season, the weather and the number of listeners. Fig. 1 shows the iconic representation of the context parameters used in the project. For example, the author may decide that a given story may be delivered when the user enters the locations of the stages, but only at twilight, in autumn, when is cloudy and when alone. The aim is to augment the emotional impact of the narration, associating the delivery of the content only to those context conditions that can increase the involvement of the listeners.

#### IV. THE SOFTWARE ARCHITECTURE

The client-server architecture for the creation and the delivery of stories has been designed starting from the story model summarized above. The architecture, represented in Fig. 2 at a very high level of abstraction, features a web-based application server for managing the different phases of the story's lifecycle. All the stories, created by different authors, are stored in a database connected to the server. Three different clients (i.e., two tablet apps and a web application) connect to this server for managing the creation, the editing and the delivery of the narration. The first client from the left is a tablet app that enables the story author to create the story outdoors and communicate to the server all the story data, including the GPS location of the story stages. The story author in this phase specifies also the contextual parameters for the delivery of the narration. The second client is a web application, accessible from any standard web browser, that enables each author to retrieve the data of his stories created outdoors, to modify and finally to publish the narration. The third client is a tablet app for delivering the stories in the locations where they have been created. The app allows to search all the stories available in the repository and published by their authors, starting from the ones that are near to the listener's current location. The context values indicated by the authors during the creation of each story are matched with the current values retrieved from the app (i.e., the location, retrieved through the GPS embedded into the device, and the other context values, retrieved from a set of web services). By default only the stories characterized by a positive matching can be delivered. Most of the implementations were realized using web technologies, for porting them to different systems. The adoption of the Phonegap framework [26] enabled us to implement the tablet apps as web applications embedded in native shells. Phonegap, available for most of the modern mobile platforms, enabled also the access to the hardware of the mobile devices, such as

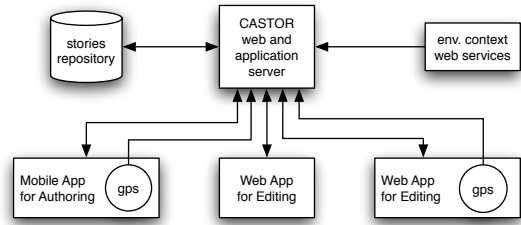


Figure 2. High-level representation of the software architecture.

the GPS, the camera and the microphone that ordinarily can't be accessed by a web application. The adoption of this platform permitted to couple a rapid development of the client interfaces and the ease of delivery for different platforms and devices.

#### V. THE VISUAL INTERFACES

All the interfaces of the system have been designed and implemented using a user-centered approach, discussing and progressively refining them through a constant dialogue with the teachers of the class that was involved in the project.

##### A. The mobile app for authoring on the field

This mobile app enables the author to create a story on the field from scratch. The app resulted from an iterative development cycle shared with the teachers, characterized by a special attention to the use of lexical terms and visual representations easy to understand. The final interface (see Fig. 3-1) permits to select different types of narrative structures and to compose incrementally the different stages that are part of the narration. The main interface (see Fig. 3-2) allows to specify all the components for a given stage: the audio content, the images of the location and of the characters, the context values for the delivery. The selection of the icons representing the context values for the delivery of the stage content is associated to funny audio fragments (e.g., the sound of the rooster for the selection of the dawn), for augmenting the user engagement. The application accesses the tablet camera for gathering the required images: the story cover, the authors' photo, the characters' representations and the photo of the locations associated with the stages. The GPS location is automatically logged when the user takes a snapshot of the stage location. The composition process is iterated for each stage of the story, until the end of the narration. The application features also a map view, based on Google Maps, that permits to visualize the recorded stages as a set of flags connected by arrows and to preview the recorded content.

##### B. The web application for editing the stories

While the first app was meant for capturing the creativity of content creation outdoors, the web application was designed for giving the users the opportunity to improve the quality of the recorded content in classroom or at home. This web application is accessible from any HTML5 web browser and permits the registered authors to access his own stories and preview them on a map, where the stories' stages are represented as numbered flags connected by arrows. A simple interface permits to edit all the components of the stories, including the context values for the delivery of the narration, and to publish the final version of the narrations.

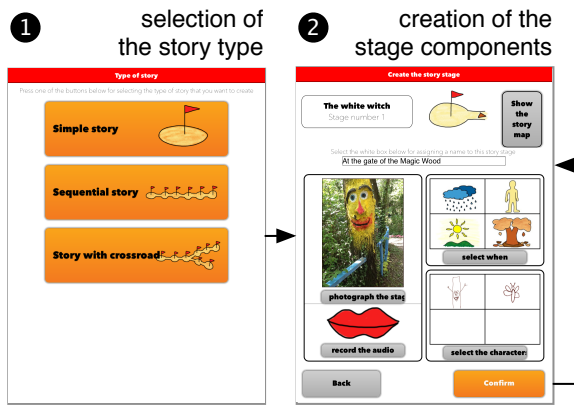


Figure 3. The visual interface for authoring the stories on the field: screen (1) displays the interface for selecting the story type; screen (2) displays the main screen for authoring each stage of the story.

### C. The mobile app for listening on the field

This novel mobile app permits to listen to the stories in the locations where they have been created, guiding the users to the different stages and delivering the appropriate content. For improving the engagement of the listeners, by default only the stories matching all the current context conditions (location, weather, time of the day, season, number of listeners) are available. An interface widget can however be used for making available also the stories matching only the location of the listener. This latter possibility was introduced in the context of the educational experience, for having a feedback by the children about how they perceived the importance of the different context dimensions for the delivery of the narration. When a story is selected, the interface (see Fig. 4-1) guides the listener towards the first stage, using a visual arrow pointing towards the location, a numerical label showing the distance and an audio signal whose frequency augments as the user gets near to it. Content is automatically delivered when the user enters the stage: the narration starts, while the visual interface displays the image and the characters associated to the stage (see Fig. 4-2). The process is iterated until the last stage of the story is reached. At the end of the story the users are invited to assign a score and to record an audio comment.

## VI. THE EDUCATIONAL PROJECT

The software architecture was experimented in an educational project, articulated in five phases. The teachers, rather than adopting our methodology and system, were deeply involved in the design process, since the early discussions with them about the context-aware story model to the design of the interfaces for the creation and the delivery of the narrations. They fully participated to the experimentation, leading the first phase of the project and collaborating to all the other ones. They modeled their educational work for enabling a smooth integration of the new educational concepts with the traditional ones (e.g., the traditional story model that usually is not driven by the context), respecting the timing but widening the scope of the traditional curriculum. From the teachers we received precious indications about the skills of the children and what we could expect from them in terms of learning. We must however point out that the introduction of novel educational

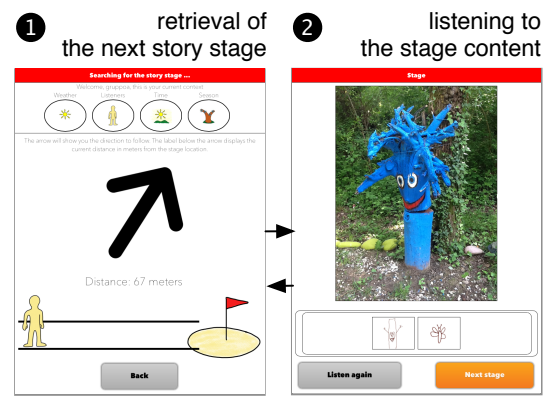


Figure 4. The visual interface for listening to the stories on the field: screen (1) displays the interface with the visual cues for retrieving the next stage; screen (2) shows the interface displayed to the user when he reaches the stage.

concepts, the integration of digital and traditional technologies, the combination of classroom and outdoors learning produced a new mix that surprised even the teachers, especially in terms of creativity, collaboration and social inclusion. More details are available in Section VII.

### A. Writing stories in the classroom

In the first phase, carried out during the two initial months of the first semester, the 19 children aged 7 learned to compose and illustrate stories in the traditional fashion, using paper and pencils. The children - supported by their teachers - learned the basic components of the story structure, focusing on the locations and on the characters. They were encouraged to describe the different dimensions of context that characterized the story, as an exercise preliminary to the introduction of the novel story model in the next phase. The class was organized in 6 groups (2 groups of girls, 3 groups of boys and 1 mixed group) that were maintained for the entire project. Stories were created both in the classroom, where the children worked in groups, and at home, where the children created the stories alone.

### B. Composing stories in the classroom with visual tools

In the second phase the children were guided to compose a story, compliant with the context-aware story model, starting from: a poster representing an inspiring scenario; a set of adhesive visual icons representing the different types of story stages and the different values of the context dimensions; a set of adhesive blank sheets for writing the content and drawing the characters associated to the stages. The goal was to give to the children a structured vision of the story components, permitting them to get acquainted with the set of symbols (see Fig. 1) that they would have used also in the following phases of the project. Each group worked autonomously in the classroom with a different copy of the scenario, composing stories with different structures, including the simple single stage story, the story based on a sequence of stages and the story with a branching structure. The children composed each story positioning on the poster (see Fig. 5) the stage icons associated with the locations of the narration and the related components (the written content, the drawings of the characters and the icons representing the values of the four context conditions). Overall the children worked in the classroom for





Figure 5. Composing a story in the classroom

about eight hours partitioned in two different days. This activity was very useful for giving the children a deeper comprehension of the story structure, in particular for what concerned the splitting of the narration content in stages. Besides, the children learned the concept of context-awareness related to the delivery of content. Most groups associated to the stages context values compliant with the mood of the narration, aimed at underlying its emotional quality.

### C. Composing stories outdoors with a tablet

In the third phase the children created the stories outdoors, using the first tablet application. We chose a park near to the school site as the inspiring scenario for the creation of stories. The park featured a small lake, a green area for sports and a wood named “Magic Wood” because a local artist decorated its trees and stones with fairy representations. The outdoor sessions were performed in two different days in the middle of December. Two groups worked in parallel in each session, that included an initial briefing and visit to the park area, the creation of a story with a single stage, the creation of a sequential story and an individual post-test questionnaire. Each group worked independently for the creation of the stories, supervised by one researcher and one teacher. Each group was provided with a tablet, paper and pencils. We established a support protocol that graduated the help from a simple invitation to examine more carefully the interface to practical demonstrations. In most cases however the simple invitation to consult the interface more carefully was sufficient for letting the children to proceed with their task. Each group worked for about two hours and was able to complete the task. The number of stages for the sequential stories varied from group to group, ranging from two to four stages. Overall, 12 stories were created in different locations of the park area (see Fig. 8).

### D. Editing stories in the classroom with a web application

The fourth phase was carried out in the computer lab of the school after a few months from the outdoor experience, in the second semester. The goal of this phase was to let the children to listen to the stories created outdoors and to improve the quality of narrations. We wanted also to check if the return to a more formal environment would have caused significant effects on the stories' structure and on the children's behavior. Each group worked for about two hours, assisted by one researcher, for editing the stories authored outdoors. At the beginning there was a short briefing for each group, aimed at describing the goal of the session and the basic functionalities of the interface. The children worked with a laptop that was alternatively used



Figure 6. Story authoring at the Magic Wood

by each component of the group. The children were initially asked to browse the content of the stories, focusing on the quality of the different components authored outdoors. After that, the children were showed how to edit the different audio and visual components of the stories. Finally the children were invited to focus on the context parameters for the delivery of narration, inviting them to confirm or to change them at their pleasure. At the end of the experiment the children were asked to fill in a short questionnaire, focused on the ease of use of the interface. Children were asked also how they would have improved it with additional features. The questionnaire included both open and closed questions based on a 3-points Likert scale.

### E. Browsing the stories outdoors with a tablet

We came back to the park at the end of the second semester, for offering the children the possibility to listen to the stories that had been created by their fellows (see Fig. 7). The session was divided in two parts: in the first part the groups could select only the stories matching all the current context conditions; in the second part they were given the possibility to access also the stories matching only the location. The support protocol included an initial invitation to look more carefully at the components of the interface, a verbal explanation and finally a practical demonstration of the features. At the end of each story the groups were invited to vote it and to record a comment about the quality of the narration. At the end of the session, the children were asked to fill in a questionnaire, including both open and closed questions.

## VII. FINDINGS AND DISCUSSION

The exploratory study provided interesting findings that were collected and evaluated through the direct observation, questionnaires filled in by the pupils, discussions with the teachers after the authoring sessions and the analysis of the stories created by children. The findings are summarized in relation to the three phases that involved the use of the tablet apps and the web applications. For the authoring sessions at the Magic Wood and in the classroom the analysis was focused on the acquisition of skills, the cooperation between children and the role of the context for augmenting the children engagement. For the final listening session, because of the different nature of the task, the analysis was less focused on the acquisition of new skills and more on the children engagement and the evaluation of the quality of the narrations.



Figure 7. Heading to the next stage of the story

TABLE I. AUTHORING STORIES OUTDOORS - SCORES ASSIGNED TO EACH GROUP (FROM A TO F) USING A 5-POINTS SCALE (1=WORST, 5=BEST).

Parameter	A	B	C	D	E	F
	M	F	M	F	M	FM
a1. Independence in story creation	5	4	4	4	4	4
a2. Complexity of stories created	5	4	4	5	4	4
a3. Organization of stories in stages	4	3	5	4	5	4
a4. Operative indep. with the device	4	4	4	4	4	4
b1. Group coop. in the story creation	3	3	4	4	4	4
c1. Emotional use of the context	5	4	5	4	5	4
c2. Authoring in the story locations	5	4	4	4	5	5

TABLE II. AUTHORING STORIES OUTDOORS - POST-TEST QUESTIONNAIRE. SCORES BASED ON A 3-POINTS SCALE (1=LOW, 3=HIGH).

Parameter	mean
a1. difficulty to create a story with a single stage	1,15
a2. difficulty to create a sequential story	1,52
c1. apprec. of the chance to specify the stage loc. in-situ	2,84
c2. apprec. of the chance to specify the weather	2,68
c3. apprec. of the chance to specify the time	2,57
c4. apprec. of the chance to specify the season	2,78
c5. apprec. of the chance to specify the social context	2,68
c6. apprec. of the chance to specify the context, overall	2,73

#### A. The authoring sessions at the Magic Wood

Table I and Table II display the evaluation grid, filled in collaboratively by the researchers and the teachers, and the children questionnaire. Table I displays the scores assigned to each group (from A to F) using a 5-points scale (1=worst, 5=best). The composition of groups is represented as follows: "M" for males, "F" for females and "FM" for mixed groups. Table II displays the scores assigned by the children in the post-test questionnaire at the end of the authoring sessions. The values express the mean calculated from the scores assigned by each child using a 3-points scale. The following discussion evidences the most interesting results of the fieldtrip, derived from these tables and from all the other sources cited above.

*a) Satisfying acquisition of skills for the creation of context-aware stories.* The children acquired a good knowledge of the story model and successfully applied it to the creation of stories. They created the stories with a good level of autonomy, selecting the theme, the locations, the characters and all the other components of the stages. The support of the teachers was limited to some hints related to the narrative content rather than to the structure of the narration, as shown by the parameter evaluating the children independence (Table I-a1). The complexity of the stories



Figure 8. A map of all the stories authored outdoors. The stages of a given story are represented by numbered flags connected by arrows.

created, measured in terms of stages and content, was satisfying (Table I-a2), given the time constraints for finishing the work. The organization of the story in stages and the association of the stages with the different locations was generally satisfying (Table I-a3). The children questionnaire (Table II-a1,-a2) is coherent with the researchers' evaluation, showing modest levels of perceived difficulty by the children for both the single stage and the sequential story, with a modest increase of difficulty for the latter one. Although most children didn't have previous experiences with the use of multi-touch devices, they generally didn't have problems in using the application (Table I-a4). When in difficulty, in most cases an invitation to examine more carefully the interface, reading all the labels for understanding how to proceed, was enough. The mobile app worked as a trait d'union between the symbolic system of icons representing the components of the story, already used in the classroom, and the practical context, providing engagement but also a structured experience. We had a confirmation about this issue when, in the few cases where there were slight differences between the icons used in classroom and outdoors, the children complained that the app interface wasn't "correct". In this sense the experiment confirmed the role that may have mobile devices, coupled with good design, for a valuable integration of symbol manipulation and practical experience [11].

*b) Improved levels of collaboration and inclusion.* We registered good levels of collaboration between the components of the groups (Table I-b1). More interestingly, comparing the activities done in the classroom and outdoors, we registered a higher degree of collaboration for those groups that were less collaborative in the classroom work. Besides, the experiment showed also higher levels of inclusion for those children that in the ordinary classroom work seemed to be less integrated with the rest of the class. These results were confirmed also by the teachers, which compared the children behavior outdoors and the ordinary classroom activities during the year.

*c) Emotional use of the context dimensions.* In most cases the children selected, for the delivery of the content,



environmental and social context parameters that were coherent with the content and that aimed to augment the emotional engagement of the listeners. We noticed the highest levels of attention for the context in the groups composed by boys, but also the other groups chose carefully the context values in relation to the quality of their narration (Table I-c1). For example, the groups that chose, as theme of the story, fairy tales populated by gnomes and ogres often specified dramatic contextual parameters such as “winter”, “twilight” or “storm” for the story delivery. Groups narrating scary stories required often that the narration should have been delivered to solitary listeners. Although the interface permitted to specify the delivery of the story also for all the values of the context conditions, only few groups decided to take advantage of such “neutral” choice, preferring an emotional use of the context. The result was confirmed by the answers to the children questionnaire, measuring high levels of appreciation for the possibility to choose the different dimensions of the context for the story delivery (Table II-c1 to c6).

d) *In-situ authoring as a stimulus for cognitive elaboration.* In spite of the fact that the park offered a number of benches and tables, the groups, after the initial exploration phase, chose to author the content in the locations chosen as story stages. Table I-c2 expresses, according to the observation of teachers and researchers, the relation between the creation of content and presence of the groups in the associated locations (i.e, the higher the number, the closer the distance). We noticed also that, during the creation of the content, many children kept on looking around the place, touching the objects that were part of the environment and trying to catch interesting details to add to the narration. Again this result is a confirmation of the importance of practical intelligence related to the discovery and to the manipulation of the elements of the location [11].

e) *Good levels of engagement.* The direct observation showed that the children were very engaged during the experiment. They were excited to create the stories outdoors and dedicated all their energies to this task, in spite of the low temperatures and the unfair season that we had to choose for the experiment. Children were engaged also by the multimodal features of the authoring environment, such as the audio feedback associated with the selection of the different values of the context, and by the multiple analogue and digital means to gather content. Overall, coupling outdoors and digital tools seemed to boost the children creativity and engagement.

#### B. The authoring sessions in the classroom

The web application received positive feedbacks for its ease of use and permitted to all the groups to obtain improved versions of the stories composed outdoors. We noticed however that coming back to the classroom resulted in a minor level of engagement. In particular the males were less involved and focused on the critical revision of the stories. Most revisions were focused on the technical improvement of the content rather than to the creation of new one. Besides, the children felt the relation between the content and the context for the delivery as less urgent, because in a number of cases (13

cases out of 48) they modified the values for listening to the stories shifting towards more inclusive values (i.e. “always”).

TABLE III. LISTENING TO STORIES OUTDOORS – P.-TEST QUESTIONNAIRE.

Parameter	mean
a1. easiness to find the locations of the stages	4,1
a2. helpfulness of the visual hints	4,0
a3. helpfulness of the audio hints	3,8
b1. engagement in listening to stories in the classroom	3,7
b2. engagement in listening to stories outdoors	4,8

#### C. The listening sessions at the Magic Wood

We registered 30 listening sessions, derived from the selection of the 12 available stories. As stated at the beginning of the section, this session was not characterized by the acquisition of new skills, aside from learning the new browsing interface. Therefore we focused our attention on measuring more carefully the engagement and evaluating the quality of the narrations. We mapped the findings derived from the direct observation, the scores assigned by the children to the stories after having listened to them and the post-test questionnaire (see Table III) to the six factors that according to O'Brien et al. [27] define the engagement.

- Perceived usability: we considered the answers to the following issues: easiness to find the locations (Table III-a1, mean 4.1), helpfulness of the visual hints (Table III-a2, mean 4.0), helpfulness of the audio hints (Table III-a3, mean 3.8);
- Felt involvement: we considered the engagement declared by the children listening to stories outdoors (Table III-b2, mean 4.8);
- Focused attention: we measured the distance between the children that carried the tablet and the other components of the groups, using the four space zones defined in [28]; during the listening phase for 121 measures out of 144 the distance was close to the edge between the intimate and the personal distance (0.45 m.) and only in 5 cases the distance reached the edge between the social and the public distance (3.6 m.);
- Aesthetics: we considered the scores assigned by the children to the stories after the listening sessions (mean score 3.4);
- Novelty: the answers to the open questions evidenced that a number of children were surprised by the ability of the system to guide them to the locations of the stage; we may infer that the application functionalities were perceived as novel at least by a part of the users;
- Endurability: we didn't measure directly this parameter but, because O'Brien et al. [27] relate high levels of endurability to high levels of perceived usability and felt involvement, we assume good levels of endurability for our application.

The analysis of the six parameters suggests a good result in terms of engagement for the children participating to the listening experience. The parameter associated to the felt involvement received a particularly high score (mean 4.8), that

can be also compared with the lower score assigned by the children to the session in the classroom (mean 3.7), when they listened to and edited the stories with the web application. In a related open question, the children associated the outdoor engagement to the fact of being in the places where the stories were supposed to happen and of moving in the real environment, following the hints of the application. Finally, the analysis of the scores assigned by the children to stories shows that the mean score of stories matching all the user context conditions (8 stories for 19 listening sessions) was higher than that one of stories matching only the location (4 stories for 11 listening sessions): respectively 3.8 and 2.7. Resuming, also the listening sessions displayed the positive effects of Castor for what concerned the engagement and the quality of the experience in an educational context.

## VIII. CONCLUSION

In this work we discussed the design and the experimentation in an educational context of Castor, a social platform for context-aware storytelling. The mobile platform implemented a novel story model meant to improve the engagement of the listeners through the use of an extended set of context dimensions. We used this platform in the context of an educational experience for checking if the possibility of proposing a structured experience outdoors and the use of an engaging story model would have produced results in terms of learning or other interesting parameters. The study produced significant results in terms of acquisition of new skills. Compared to traditional classroom activities, the analysis of outdoors activities revealed higher level of engagement, creativity, cooperation between children and social inclusion. The availability of a digital tool for creating stories, installed on a mobile device but modeled on the structured experience designed for the classroom, permitted to bring a structured learning process outdoors, taking advantage of the benefits given by this more informal environment in terms of collaboration and inclusion. Concluding, while the learning benefits of Castor deserve further investigations, we may infer from the experimental study that this architecture had a positive role on bridging the gap between the educational experiences performed in a formal classroom context and outdoors.

## REFERENCES

- [1] M. Cavazza and D. Pizzi, "Narratology for interactive storytelling: a critical introduction." in Proc. of TIDSE 2006. Heidelberg: Springer, 2006, pp. 72–83.
- [2] C. Segre and T. Kemesy, Introduction to the analysis of the literary text, Advances in semiotics. Indiana University Press, 1988.
- [3] A. Lamstein and M. Mateas, "Search-based drama management," in Proc. of the AAAI-04 Workshop on Challenges in Game AI. AAAI Press, 2004.
- [4] S. Sanchez, O. Balet, H. Luga, and Y. Duthen, "Autonomous virtual actors," in Proc. of TIDSE 2004. Heidelberg: Springer, 2004, pp. 68–78.
- [5] F. Garzotto, P. Paolini, and A. Sabiescu, "Interactive storytelling for children," in Proc. of IDC '10. New York, NY, USA: ACM, 2010, pp. 356–359.
- [6] T. Göttel, "Reviewing children's collaboration practices in storytelling environments," in Proc. of IDC '11. New York, NY, USA: ACM, 2011, pp. 153–156.
- [7] N. Di Blas, P. Paolini, and A. Sabiescu, "Collective digital storytelling at school as a whole-class interaction," in Proc. of IDC '10. New York, NY, USA: ACM, 2010, pp. 11–19.
- [8] J. Halloran, E. Hornecker, G. Fitzpatrick, M. Weal, D. Millard, D. Michaelides, D. Cruickshank, and D. De Roure, "The literacy fieldtrip: using ubicomp to support children's creative writing," in Proc. of IDC '06. New York, NY, USA: ACM, 2006, pp. 17–24.
- [9] Y. Rogers, S. Price, G. Fitzpatrick, R. Fleck, E. Harris, H. Smith, C. Randell, H. Muller, C. O'Malley, D. Stanton, M. Thompson, and M. Weal, "Ambient Wood: designing new forms of digital augmentation for learning outdoors," in Proc. of IDC '04. New York, NY, USA: ACM, 2004, pp. 3–10.
- [10] M. J. Weal, E. Hornecker, D. G. Cruickshank, D. T. Michaelides, D. E. Millard, J. Halloran, D. C. De Roure, and G. Fitzpatrick, "Requirements for in-situ authoring of location based experiences," in Proc. of Mobile-HCI '06. New York, NY, USA: ACM, 2006, pp. 121–128.
- [11] F. A. Hansen, K. J. Kortbek, and K. Grønbaek, "Mobile urban drama for multimedia-based out-of-school learning," in Proc. of MUM '10. New York, NY, USA: ACM, 2010, pp. 17:1–17:10.
- [12] F. A. Hansen, K. J. Kortbek, and K. Grønbaek, "Mobile urban drama — setting the stage with location based technologies," in Proc. of ICIDS '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 20–31.
- [13] Resnick, "Learning in school and out," Educational Researcher, vol. 16, pp. 13–20, 1987.
- [14] J. A. Fails, A. Druin, and M. L. Guha, "Mobile collaboration: collaboratively reading and creating children's stories on mobile devices," in Proc. of IDC '10. New York, NY, USA: ACM, 2010, pp. 20–29.
- [15] A. K. Dey, "Understanding and using context," Personal Ubiquitous Comput., vol. 5, pp. 4–7, January 2001.
- [16] P. Brusilovsky, "Adaptive hypermedia," User Modeling and User-Adapted Interaction, vol. 11, pp. 87–110, March 2001.
- [17] M. Perkowitz and O. Etzioni, "Adaptive web sites," Commun. ACM, vol. 43, pp. 152–158, August 2000.
- [18] S. Gordillo, G. Rossi, and F. Lyardet, "Modeling physical hypermedia applications," in Proc. of the 2005 Symposium on Applications and the Internet Workshops. Washington, DC, USA: IEEE Computer Society, 2005, pp. 410–413.
- [19] C. Ardito, P. Buono, M. F. Costabile, R. Lanzilotti, and T. Pederson, "Mobile games to foster the learning of history at archaeological sites," in Proc. of VLHCC '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 81–86.
- [20] F. Garzotto, P. Paolini, M. Speroni, B. Pro'Il, W. Retschitzegger, and W. Schwinger, "Ubiquitous access to cultural tourism portals," in Proc. of DEXA '04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 67–72.
- [21] M. Dionisio, V. Nisi, and J. P. Van Leeuwen, "The island of madeira location aware multimedia stories," in Proc. of ICIDS '10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 147–152.
- [22] A. Celentano and F. Pittarello, "Observing and adapting user behavior in navigational 3d interfaces," in Proc. of AVI '04. New York, NY, USA: ACM, 2004, pp. 275–282.
- [23] R. W. Picard, "Affective computing," M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 321, Tech. Rep., 1995.
- [24] K. Ryokai, H. Raffle, and R. Kowalski, "Storyfaces: Pretend-play with ebooks to support social-emotional storytelling," in Proc. of IDC '12. New York, NY, USA: ACM, 2012, pp. 125–133.
- [25] F. Pittarello and L. Bertani, "Castor: Learning to create context-sensitive and emotionally engaging narrations in-situ," in Proc. of IDC '12. New York, NY, USA: ACM, 2012, pp. 1–10.
- [26] Phonegap, "Open Source Mobile Framework," <http://phonegap.com>.
- [27] H. L. O'Brien and E. G. Toms, "The development and evaluation of a survey to measure user engagement," J. Am. Soc. Inf. Sci. Technol., vol. 61, no. 1, pp. 50–69, Jan. 2010.
- [28] E. Hall, The hidden dimension, ser. A Doubleday anchor book. Doubleday, 1966.

# Demo/Poster



# Distance Learning Immersive Environments: Sense of Presence Exploration

Max M. North

Visualization & Simulation Research Center  
Southern Polytechnic State University, Georgia, USA  
max@spsu.edu

**Abstract**—As distance education rapidly appears on the horizon, distance learning technologies are becoming more important in delivering contents and allowing collaboration between students and teachers in globally disparate environments. One of the recent innovative technologies to be explored in distance learning field is immersive environments and sense of presence that offers. The primary purpose of this research is to study the sense of presence—the sense of “being there”—using an immersive environment. While this is a work-in-progress, the authors provide preliminary results and conclusions.

**Keywords**—Distance Learning-VR- Sense of Presence

## I. INTRODUCTION

There are different aspects that make immersive environments (also known as virtual reality) feel as “real” as possible. One well-known aspect of an immersive environment is “the sense of presence” in the environment. It is often thought of as the sense of “being there.” In the 1990’s, a few theoretical research articles were published in the journal of *Presence, Teleoperators and Virtual Environments*, published by the Massachusetts Institute of Technology (MIT). In the 2000’s, researchers increasingly have been exploring the topic of sense of presence and using the knowledge gained from various immersive environments studies to further their applications.

## II. SENSE OF PRESENCE

How is the presence defined? In this research paper, the main definition of presence that will be used is the awareness or state of the mind of being in an environment, either real or virtual [1, 2]. Experiencing a sense of presence can happen whenever a person mentally feels that he or she is present in a situation. For example, when reading a book, some people may feel what the characters are feeling and become wrapped up in the book—they may feel as though they are watching a movie, or have an even deeper connection with the text.

## III. IMMERSIVE ENVIRONMENT EXPERIMENT

**Participants:** Thirty-five volunteers participated in this part of the study.

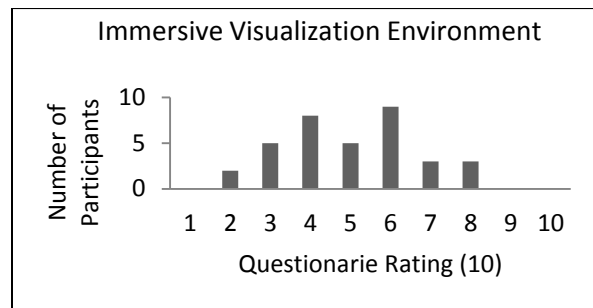
**Apparatus:** The device used for this part of the experiment was the Immersive Visualization Environment, also known as a dome-shaped system (see figure 1). This state-of-the-art equipment enhances the 3-D virtual environment imaging with four digital projectors and an 8’x10’ cylindrical screen.

**Procedure:** Each participant was seated in a chair placed a specific distance away from the screen, providing the individual with a fully-immersed environment, and cancelling out any outer disturbances. Though the environment was controlled, the participant had the ability to freely move his/her body without the restrictions of a helmet or hand-held device. A survey was administered to each participant which attempted to measure their sense of presence (the sense of

being there in the specified environment), with 10 indicating the highest presence and 0 indicating the lowest. One of the major questions asked was, “Rate your sense of presence in the virtual world during the experiment.” A graph of the results for this question is shown in Figure 2.



**Figure 1.** This figure depicts several subjects are engaged in Immersive Visualization Environment (Distance Learning)



**Figure 2.** Rating of the sense of presence in the virtual world during the experiment in the immersive environment.

## IV. PRELIMINARY RESULTS AND CONCLUSIONS

A strong sense of presence of the virtual world in the immersive environment was experienced by a majority of participants. Specifically, 86% of the participants—30 participants—responded with a rating between 7 and 9, indicating that the participants felt a high sense presence in the virtual environment. This single finding raises several possibilities for future research.

## ACKNOWLEDGMENT

This effort was supported by an equipment grant from the Army Research Office (ARO).

## REFERENCES

- [1] Usoh, M., Arthur, K., Whitton, M., Bastos, R., Steed, A., Slater, M., Brooks, F. (1999) Walking > Walking-in-Place > Flying, in Virtual Environments. Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. 359-364.
- [2] Nunez, D., & Blake, E. (2001) Cognitive Presence As A Unified Concept Of Virtual Reality Effectiveness. Proceedings of the 1st international conference on Computer graphics, virtual reality and visualization. South Africa 115-118.

# The Recovery System for Hadoop Cluster

Prof. Priya Deshpande  
Dept. of Information Technology  
MIT College of engineering  
Pune, India  
priyardeshpande@gmail.com

Darshan Bora  
Dept. of Information Technology  
MIT College of engineering  
Pune, India  
darshanbora@hotmail.com

**Abstract**—Due to brisk growth of data volume in many organizations, large-scale data processing became a demanding topic for industry as well as for academic fields. Hadoop is widely adopted in Cloud Computing environment for unstructured data. Hadoop is an open source, a java based distributed computing framework, and supports large-scale distributed data processing. In the recent years, Hadoop Distributed File System (HDFS) is popular for huge data sets and streams of operation on it. Availability of Hadoop is the important factor in Cloud Computing. But, in HDFS, Namenode failure affects the performance of the Hadoop cluster. It can be a single point failure. In this paper, we analysed the behaviour of Namenode, what are effects of Namenode failure. This paper presents a scenario to overcome this failure. Our scenario replicates the Namenode on the other Datanode so that the availability of the metadata is increases which will reduce the loss of data as well as delay.

**Keywords**— Hadoop; Cloud Computing; HDFS; Namenode; availability; failure.

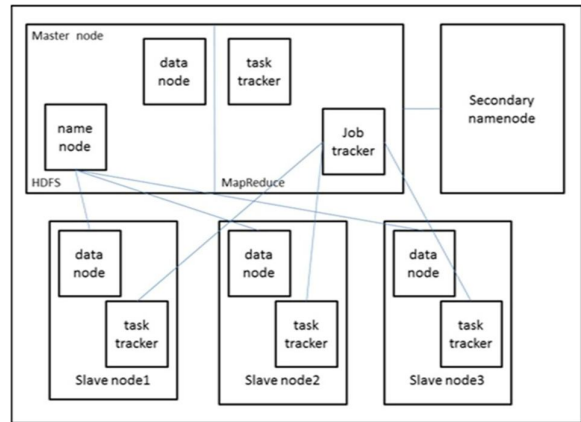
## I. INTRODUCTION

Cloud Computing is now mainstream commodity in the IT sector [1]. Accordingly, the impact of hardware failure as well as software failure decreases the performance of the cloud infrastructure. Failure occurrence may have major impact on efficiency of an application or it may cause an application being temporarily out of service. Cloud infrastructure should overcome from these kinds of failures. Now, Hadoop is a cloud workhorse [2]. Many internet companies are dependent on Hadoop for their large datasets. Every day, they are generating large amount of data which is in Terabytes, for example Facebook is generating everyday up to 5 Terabytes data. As a platform of computing and storage, Hadoop on a wider range deals with these kinds of data. Hadoop is a framework of open-source implementation of MapReduce for the analysis of large datasets. To manage storage resources across the cluster, Hadoop uses a distributed user-level file system called as Hadoop Distributed File System (HDFS) [3]. The HDFS is robust and highly scalable. The architectural representation of Hadoop is as shown in the figure 1.

Hadoop architecture is based on master-slave architecture. MapReduce Engine and HDFS are the important component of this architecture. JobTracker and TaskTracker are key parts of MapReduce engine while Namenode and Datanode are of HDFS. MapReduce deals with the computation while HDFS

is used for storage purpose. In this paper, we focused on HDFS and its failure.

As states earlier working of HDFS is based on Namenode and Datanode while the design of HDFS is based on the design of *GFS*, the Google File System [15][16].



**Figure 1:** Hadoop Architecture in Multi-node Cluster [3][14]

HDFS is nothing but a block-structured file system: individual files are broken into blocks of a fixed size called as **Chunk**; size may be in the order of 4 KB or 8 KB, currently sizes are in MBs. These chunks are stored across a cluster of one or more machines with data storage capacity. Individual machines in the cluster are referred to as **Datanodes**. A file can be made up of several chunks, and they are stored on the different Datanodes. If several Datanodes must be involved in the serving of a file, then a file could be caused unavailable by the failure of any one of those Datanodes. HDFS combats this problem by replicating each chunk across a number of Datanodes and by default 3 Datanodes are selected for replication.

In this scenario, it is important for this file system to store its metadata reliably, a node holding metadata of the stored data on different Datanodes is referred as **Namenode** [6][12]. Metadata holds the data, which contains information about data, which are stored on different Datanodes. In HDFS after every transaction or read-write operation metadata is going to be updated.

Namenode is the pillar of the HDFS architecture; therefore, reliability of Namenode is having significant value



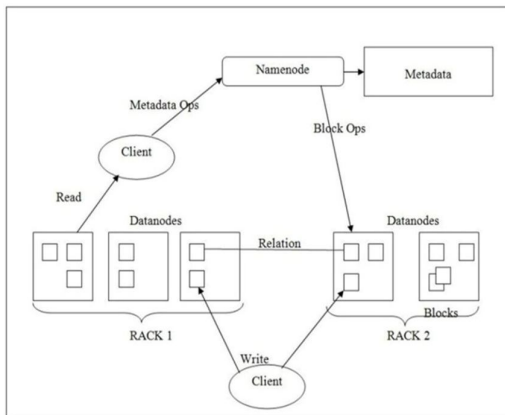
in HDFS. Whenever Namenode gets down the working of HDFS gets affected.

The rest of paper is as organized as follows. Section II provides brief information about HDFS architecture as well as it analyses the behaviour of HDFS under failures. The proposed scenario of the system is explained in the section III. The architecture of proposed scenario is described in Section IV while Section V is the conclusion of our proposed scenario.

## II. BACKGROUND

### A. Hadoop Architecture

Hadoop comes with a distributed file system called HDFS, which stands for Hadoop Distributed File system [5]. HDFS is a file system designed for storing very large files with streaming data access patterns, running on clusters on commodity hardware. HDFS is a distributed file system.



**Figure 2:** Proposed HDFS Architecture in Multi-node Cluster [3]

HDFS is a framework for analysis and transformation of very large data sets using MapReduce paradigm. HDFS stores file system metadata and application data separately. HDFS is built around the idea that the most efficient data processing pattern is a write-once, read-many-times pattern. A dataset is typically generated or copied from source, and then various analyses are performed on that dataset over time. HDFS is optimized for delivering a high throughput of data, and this may be at the expense of latency. Files in HDFS may be written to by a single writer. Writes are always made at the end of the file. There is no support for multiple writers, or for modifications at arbitrary offsets in the file.

As in other distributed file systems, like PVFS, Lustre and GFS, HDFS stores metadata on a dedicated server, called the Name Node. Application data are stored on other servers called Data Nodes. All servers are fully connected and communicate with each other using TCP-based protocols. Unlike Lustre and PVFS, the Data Nodes in HDFS do not rely on data protection mechanisms such as RAID to make the data durable. Instead, like GFS, the file content is replicated on multiple Data Nodes for reliability. While ensuring data

durability, this strategy has the added advantage that data transfer bandwidth is multiplied, and there are more opportunities for locating computation near the needed data.

The Hadoop Distributed File System (HDFS) is designed to store very large data sets reliably, and to stream those data sets at high bandwidth to user applications. In a large cluster, thousands of servers both host directly attached storage and execute user application tasks. By distributing storage and computation across many servers, the resource can grow with demand while remaining economical at every size. The architecture of HDFS and report on experience using HDFS to manage 40 petabytes of enterprise data is described at Yahoo!

### B. Single Point Failure in HDFS

HDFS architecture is mainly based on Namenode and Datanode, where Namenode act as a master while Datanode act as a slave. If Datanode gets failed in this scenario, only one machine will down, in that case Namenode will divert the work of failed Datanode to other available Datanode [6][9]. But suppose, Namenode goes down, then there will be a single point failure.

To avoid this, HDFS architecture selects a Secondary Namenode which will work after Primary Namenode fails. Fig. 1 shows the architecture of Hadoop with Secondary Namenode. Secondary Namenode is not a Namenode in the sense that Datanodes cannot connect to the secondary Namenode and in no event it can replace Namenode in case of its failure. If Hadoop is not able to use Namenode anymore it will need to copy latest image and logs somewhere else and need to restart whole cluster. It is a time consuming process, and also affects the performance of HDFS.

To overcome above problem we are going to propose a system which deals with single point failure which means without restarting cluster.

## III. PROPOSED WORK

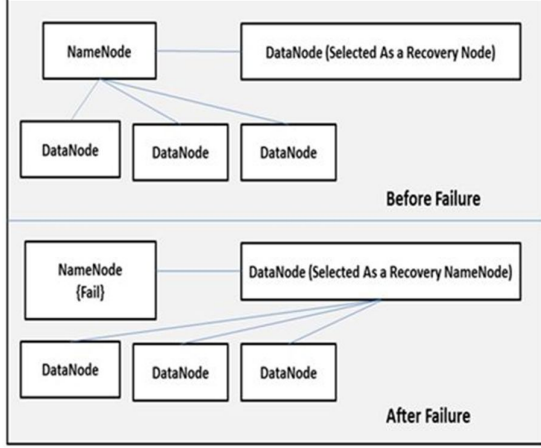
In this HDFS architecture, if Namenode gets fails; it means a single point failure. Because as we have stated earlier Namenode is a pillar of HDFS architecture and which contains metadata of the data which are stored on different Datanodes. If Namenode goes down, it will affect the entire cluster. To overcome this single point failure, our suggestion is, replicate the entire Namenode on the other Datanode called as "Recovery Namenode". Recovery Namenode will update all the information of Namenode simultaneously. After failure, Recovery Namenode will act as Namenode.

Recovery Namenode will keep track Namenode. After a periodic time interval Recovery Namenode is updated. Initially all the Datanodes sending to heartbeats Namenode, when Namenode gets down Recovery Namenode will broadcast a message to all Datanodes of new Namenode. After that every Datanode will send heartbeats to Recovery Namenode.

The detailed description of the proposed system is given in the architecture section. It will select new Namenode from available Datanodes.

#### IV. ARCHITECTURE

The behaviour of the HDFS architecture will be shown as:



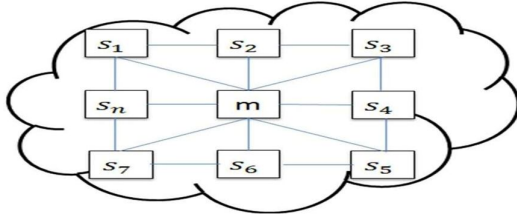
**Figure 3:** Proposed HDFS Architecture

There are two cases; first one gives the architecture of the new HDFS before failure while second one gives the architecture of new HDFS after failure with new Namenode. Let us see how this system will work:

##### A. Selection of Recovery Namenode

As we said earlier Namenode is a pillar of HDFS architecture, considering this part the next node means “Recovery Namenode” is having significant responsibilities.

If there are ‘n’ number of nodes in the cluster as shown in the figure below:



**Figure 4:** Hadoop Cluster

We have to set some appropriate methods for selection of recovery node so that selection will be fast and efficient. We have to also consider an availability of nodes so that which will not affect performance of the cluster. Here, ‘m’ is the Namenode while ‘s<sub>1</sub>, s<sub>2</sub>, ..., s<sub>n</sub>’ are the Datanodes. From these data node our proposed scenario will select one node as a ‘Recovery Namenode’.

As we know, after every 3s Datanode sends a heartbeat to Namenode to show his availability. In our scenario, each

Datanode sends a heartbeat along with its time of generation i.e.  $t_{dt}$ . Every Datanode has variant time of heartbeat generation such as  $t_{dt1}, t_{dt2}, \dots, t_{dtn}$  for ‘n’ nodes. After that, at Namenode this time will stores in a log referred as “Heartbeat Log” along with arrival time  $t_{at}$  of each Datanode. Now, calculate the time, to reach Namenode from Datanode for each and every Datanode as shown:

$$tt_{dt} = t_{at} - t_{dt}$$

Where,  $tt_{dt}$ =actual time taken by a heartbeat from Datanode to Namenode.

For every Datanode we are considering first ‘x’ readings. Then, calculate a mean time for every Datanode as follows:

$$mt_{dt_n} = \frac{1}{x} \sum_{1}^x tt_{dt}$$

Where,  $mt_{dt_n}$ = mean time of heartbeat to reach namenode to Datanode of Datanode ‘n’.

Now, we are having a log of all Datanodes with their respective mean time ‘ $mt_{dt}$ ’. By applying Quick sort algorithm heartbeat log is updated called as “Recovery Namenode List”.

According to this *Recovery Namenode List*, the first node from the list is selected as “Recovery Namenode”. At the time of cluster formation as per the node registration log is maintained. According to availability of Datanodes, *Recovery Namenode List* is updated, after every 600s.

##### B. Create a Recovery Namenode List

For the selection of Recovery Namenode we are creating a Recovery Namenode List from that, according to mean time, list is going to be sorted. In that, algorithm will first check heartbeat response of node. If node is up, add it to recovery node list else ignore that node, while adding calculate travelling mean time of that node so that we will get mean response time.

Calculate mean response time for each and every node which is up. Then sort this list according to mean response time. The first node will be new “Recovery Namenode” for the cluster.

##### C. Communication between Namenode and Recovery Namenode

When Namenode will select a Recovery Namenode, the communication between Namenode and Recovery Namenode is an important factor. There will be an instant messaging from Namenode to Recovery Namenode. After a certain time period Namenode will generate an instant message and send it to Recovery Namenode, so that Recovery Namenode will come to know that Namenode is alive. If for 600s Namenode is failed to send an instant message then Recovery Namenode will be declared as new Namenode. Recovery Namenode will broadcast that message to all Datanodes

```

create_recovery_namenode_list()
{
    get_all_node_list();
    while (list is not empty)
    {
        if(node.heartbeat_response == TRUE)
        {
            node_mean_time=
calculate_heartbeat_mean_time(node);
            add_node_to_recovery_namenode_list(node,
node_mean_time);
            goto_next_node;
        }
        else
        {
            goto_next_node;
        }
    }

calculate_heartbeat_mean_time(node)
{
    total_travelling_time=0;

    for(from starting time up to a certain time)
    {
        hb_start_time[]= get_start_time_time();
        hb_recieved_time[]=get_received_time();
        hb_traveeling_time = hb_recieved_time[] -
hb_start_time[];
        total_travelling_time= total_travelling_time+
hb_traveeling_time;
    }

    node_mean_time= total_travelling_time / (certain_time -
starting_time);
    return node_mean_time;
}

quick_sort (recovery_namenode_list with respect to
node_mean_time);

```

**Figure 5:** Algorithm for Selection of Recovery Namenode

#### D. Set a Checkpoint

Checkpoint method is widely used in different recovery model [10]. It allows system to recover from unpredictable fault. The idea behind this system is the saving and restoration of the system state.

Here, checkpoints are nothing but a time interval which is periodic. To replicate Namenode on Recovery Namenode, a time interval is considered. On a certain time interval checkpoints are created. After every 600s Namenode is replicated on Recovery Namenode. It means checkpoints are created after 600s.

Here, checkpoints are sets only for Namenode. Creating periodic checkpoints is the way to protect the metadata of file system.

#### E. Availability of Namenode

As per the HDFS architecture, Namenode does share any information about his failure, to overcome this problem in our scenario; Namenode will generate an instant message, sends to Recovery Namenode after 3s to ensure his availability. If Namenode failed to send instant message up to 600s then namenode will be declared as dead node. After that Recovery Namenode will send a message to all Datanode, that he is the new Namenode and send heartbeats to him. Recovery Namenode will start his work from a last checkpoint, before failure of Namenode.

#### F. Failure of Recovery Namenode

Whenever Namenode gets fail, Recovery Namenode will occupy his place. When Recovery Namenode will be a new Namenode, again new Recovery Namenode is selected by using same parameters. According to available Datanodes new Recover Namenode list is generated in a similar way, again checkpoints for a new Namenode are created. Though it increases overheads on Namenode as well as Datanode but provides high availability.

### V. CONCLUSION

In cloud computing, unstructured data storage is popular issue. Hadoop deals with unstructured data storage. In this paper, we have studied and analyzed the architecture of the Hadoop Distributed File System under Namenode failure. To overcome single point Namenode failure in HDFS we have proposed an architecture which increases reliability as well as availability of Hadoop. We also focused on selection of Recovery Namenode after failure of Namenode. This proposed architecture is massively helpful for an unrecoverable Namenode failure.

### REFERENCES

- [1] Florin Dinu, T. S. Eugene Ng, Understanding the Effects and implications of compute Node Related Failure in Hadoop, HPDC'12, Delft, The Netherlands, June 18-22, 2012.
- [2] Jeffrey Shafer, Scott Rixner, and Alan L. Cox, The Hadoop Distributed Filesystem: Balancing Portability and Performance, Presentation, ISPASS 2010, March 30<sup>th</sup> 2010.
- [3] Dhruba Borthakur, The Hadoop Distributed File System: Architecture and Design, The Apache Software Foundation.
- [4] Ronald Taylor, Pacific Northwest National Laboratory, Richland, WA, An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics, Bioinformatics Open Source Conference 2010, doi:10.1186/1471-2105-11-S12-S1.
- [5] Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler "The Hadoop Distributed File System", IEEE 2010.

- [6] MohommadAsif Khan, Zulfiqar A. Menon, Sajid Khan, Highly Available Hadoop Namenode Architecture, International Conference on Advanced Computer Science Applications and Technologies2012, Confrence Publishing Services, DOI 10.1109/ACSAT.2012.52.
- [7] AsafCidon, Stephen Rumble, Ryan Stutsman, SachinKatti, John Ousterhout and MendalRosenblum, Copysets: Reducing the Frequency of Data Loss in Cloud Storage.
- [8] FarazFaghri, SobirBazabayev, Mark Overholt, Reza Farivar, Roy H. Campbell and William H. Sanders, Failure Scenario as a Service (FSaaS) for Hadoop Cluster, SDMM'12, Montreal, Quebec, Canada, December 3-4,2012.
- [9] Florin Dinu, T. S. Eugene Ng, Analysis of Hadoop's Performance under Failures.
- [10] Jorge-Arnulfo Quijano-Ruiz, Christoph Pinkel, Jorg Schad, Jens Dittrich, RAFT at Work: Speeding-Up MapReduce Applications under Task and Node Failures, SIGMOD'11, Athens, Greece, June 12-16, 2011
- [11] Big Data – Hadoop HDFS and MapReduce  
<http://codemphysis.wordpress.com/2012/09/27/big-data-hadoop-hdfs-and-mapreduce/>
- [12] Hadoop Getting Started  
[http://docs.hortonworks.com/HDPDocuments/HDP1/HD P-Win-1.1/bk\\_getting-started-guide/content/ch\\_hdp1\\_getting\\_started \\_ chp3.html](http://docs.hortonworks.com/HDPDocuments/HDP1/HD P-Win-1.1/bk_getting-started-guide/content/ch_hdp1_getting_started _ chp3.html)
- [13] Hadoop Tutorial  
<http://developer.yahoo.com/hadoop/tutorial/module2.html>
- [14] Understanding Hadoop Clusters and the network  
<http://bradhedlund.com/2011/09/10/understanding-hadoop-clusters-and-the-network/>
- [15] Hadoop in Practice  
<http://techannotation.wordpress.com/2012/09/10/hadoop-in-practice/>
- [16] The Building Blocks of Hadoop  
<http://pramodgampa.blogspot.sg/2013/06/the-building-blocks-of-hadoop.html>

**Author Information:**

**Prof. Priya Deshpande**

Assistant Professor -MITCOE Pune  
priyadeshpande@gmail.com

**Mr. Darshan Bora**

ME-Student ,MITCOE ,Pune  
darshanbora@hotmail.com

# Improving App Inventor Usability via Conversion between Blocks and Text

Karishma Chadha

MIT Lincoln Lab, Lexington, MA 02421, USA

Email: karishma.chadha@ll.mit.edu

Franklyn Turbak

Wellesley College, Wellesley, MA 02481, USA

Email: fturbak@wellesley.edu

## Abstract

We have developed TAIL, a textual programming language isomorphic to the blocks language of MIT App Inventor 2 (AI2), and have extended AI2 with code blocks, a novel mechanism that enables bidirectional isomorphic conversions between blocks and text program fragments. TAIL improves AI2's usability by facilitating the reading, writing, and sharing of programs, and may also ease the transition from blocks programming to text programming

## 1. Project Overview

In blocks programming languages, programs are constructed by connecting visual fragments (blocks) shaped like jigsaw puzzle pieces. MIT App Inventor 2 (AI2) [7], a popular online environment for Android app development, democratizes mobile programming through its easy-to-use blocks language. For example, the AI2 blocks shown in Figure 1 are a *No Texting While Driving* app that automatically replies to an incoming text message with a canned response and then speaks the phone number and content of the message. The shapes of plugs and sockets on the blocks distinguish expressions and statements and suggest how the blocks fit together, reducing syntactic frustrations experienced by novices when learning textual programming. Sockets have labels documenting their purpose, helping programmers remember the number and order of operands. Selecting a block from a menu of related blocks (e.g. math blocks, control blocks) can be easier than remembering its name. Other blocks languages include Scratch [6, 8], Blockly [4], StarLogo Nova [12], and PicoBlocks [11].

While simple AI2 blocks programs are easy to read and write, more complex ones can become overwhelming. Creating and navigating nontrivial blocks programs is tedious,

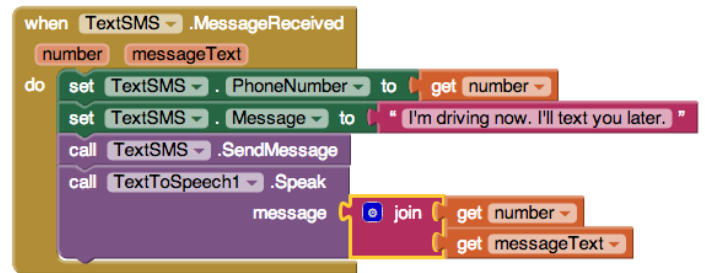
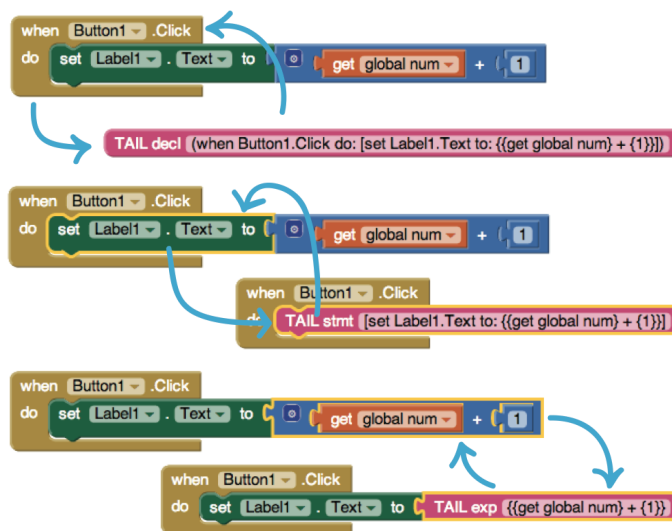


Figure 1. AI2 blocks for a simple *No Texting While Driving* app.

and AI2's inability to copy blocks between projects inhibits reusing code between projects and between programmers.

To address these issues, we have created a new textual language, TAIL (the Textual App Inventor Language), that is isomorphic to AI2's blocks language and have provided a means for converting between them. TAIL syntax is designed to provide users with a systematic way to translate the visual information on the blocks into text. We have extended AI2 with *code blocks* in which users can type TAIL code representing AI2 expression, statement, and top-level declaration blocks. These code blocks have the same meaning as the larger block assemblies they represent. Programmers can also convert back and forth between these code blocks and the original AI2 blocks (Fig. 2). Language isomorphism guarantees that round-trip conversions (from text to blocks and back, or blocks to text and back) begin and end with the identical program.

Ours is the first blocks language to support isomorphic conversions between blocks and text languages in both directions. This distinguishes it from systems that can convert blocks to text programs, but not vice versa (e.g. [1, 3, 5, 9] and from PicoBlocks, which can define blocks in a text language more expressive than its blocks language but cannot convert blocks to text [11]. Bidirectional conversions



**Figure 2. Sample TAIL ↔ blocks conversions.**

(1) increase AI2’s usability by providing an efficient means for reading, writing, and reusing/sharing programs, and (2) may ease users’ transitions from blocks programming to text programming.

The TAIL language is implemented using an ANTLR parser generator [10] to generate a JavaScript lexer and parser for TAIL. Actions in the grammar are used to translate the TAIL parse tree into AI2’s XML tree representation for blocks. These and other details are explained in [2].

This is work in progress. Currently, TAIL handles only a subset of AI2 blocks. We are extending TAIL to handle the full AI2 language and to express whole programs in addition to program fragments. We also need to test the usability of TAIL and code blocks in user studies so that we can refine them before including them in a release of AI2.

## 2. Why is this Research of Interest?

This research is of interest to the VLC/DMS community because App Inventor is becoming an increasingly popular visual language for creating mobile apps and teaching programming. It is a step towards addressing common criticisms of App Inventor: that large programs are difficult to read and write; that program fragments cannot be reused between programs; and that there is no clear transition between blocks programming and text-based programming.

## 3. Nature of the Demo & Poster

Our presentation will have two parts: (1) a poster that summarizes the TAIL language and the mechanisms for converting between TAIL textual program fragments and

App Inventor 2 blocks; and (2) an interactive demo of the conversion between TAIL and blocks in a version of App Inventor 2 running on a laptop computer.

## Acknowledgments

This work was supported by the National Science Foundation under grant DUE-1226216, by sabbatical funding from Wellesley College, and by the Wellesley College Science Center Summer Research program.

## References

- [1] K. A. Behnke. Slash: Scratch-based visual programming in Second Life for introductory computer science education. Poster presented at SIGCSE’13. University of Colorado Boulder, ATLAS Institute, 2013.
- [2] K. Chadha. Improving the usability of App Inventor through conversion between blocks and text. Undergraduate thesis, Wellesley College, May, 2014.
- [3] N. Fraser. Blockly code demo. <https://blockly-demo.appspot.com/static/apps/code/index.html>, accessed Jul. 5, 2014.
- [4] N. Fraser. Blockly website. <https://code.google.com/p/blockly>, accessed Jul. 5, 2014.
- [5] P. Guo. Proposal to render Android App Inventor visual code blocks as pseudo-Python code. [http://people.csail.mit.edu/pgbovine/android\\_to\\_python](http://people.csail.mit.edu/pgbovine/android_to_python), accessed Jul. 5, 2014.
- [6] J. Maloney, M. Resnick, N. Rusk, B. Silverman, and E. Eastmond. The Scratch programming language and environment. *ACM Transactions on Computing Education*, 10(4), Nov. 2010.
- [7] MIT App Inventor 2 home page. <http://appinventor.mit.edu>, accessed Jul. 5, 2014.
- [8] MIT Lifelong Kindergarten Group. Scratch website. <http://scratch.mit.edu>, accessed Jul. 5, 2014.
- [9] M. Myburgh. Printing the scripts for a Scratch program. <http://itisgr8.blogspot.com/2012/01/printing-scripts-for-scratch-program.html>, accessed Jun 17, 2014.
- [10] T. Parr. ANTLR 3 website. <http://www.antlr3.org/>, accessed Jul. 5, 2014.
- [11] Playful Invention Company. PicoCricket Reference Guide, version 1.2a. [http://www.picocricket.com/pdfs/Reference\\_Guide\\_V1\\_2a.pdf](http://www.picocricket.com/pdfs/Reference_Guide_V1_2a.pdf), accessed Jul. 5, 2014.
- [12] Scheller Teacher Education Program. StarLogo Nova website. <http://www.slnova.org>, accessed Jul. 5, 2014.



# A Decision Support System for Evidence Based Medicine

Giuseppe Polese

Department of Management & Information Technology  
University of Salerno, Italy  
gpolese@unisa.it

## Abstract

*We present a decision support system to let medical doctors analyze important clinical data, like patients medical history, diagnosis, or therapy, in order to detect patterns of knowledge useful in the diagnosis process.*

## 1. Introduction

We present a decision support system for Evidence Based Medicine (EBM), which combines data warehouse and data mining techniques [6, 7], together with additional inference mechanisms based on the Schanks theory [5], a non-logical approach widely used in the natural language processing area, with extensions used in the context of visual language understanding [2].

The paper is organized as follows. Section 2 discusses the approach underlying the proposed system, described in section 3. Finally, conclusion are given in section 4.

## 2. The Approach

We have adopted a case-based reasoning (CBR) approach, which is useful to extract knowledge from previously experienced cases. In particular, we have selected data from a large database of an infectious diseases department of a big public hospital, which has led to the selection of the features described in table 1.

We had to first identify and solve problems like data duplication, inconsistencies between logically associated values, missing data, unexpected use of one or more fields, and inconsistent values possibly due to different conventions, abbreviations, or to data entry errors. A particular effort was required for the presence of free text fields containing significant information, which was handled by using natural language processing and data classification.

Finally, the data mining strategy we have used is based on a sequence clustering algorithm. It is a sequence anal-

ysis algorithm, which allows to find the most common sequences by grouping those that are similar or identical.

## 3. The Proposed System

The system is structured according to a classical three tier architecture, shown in Fig. 1.

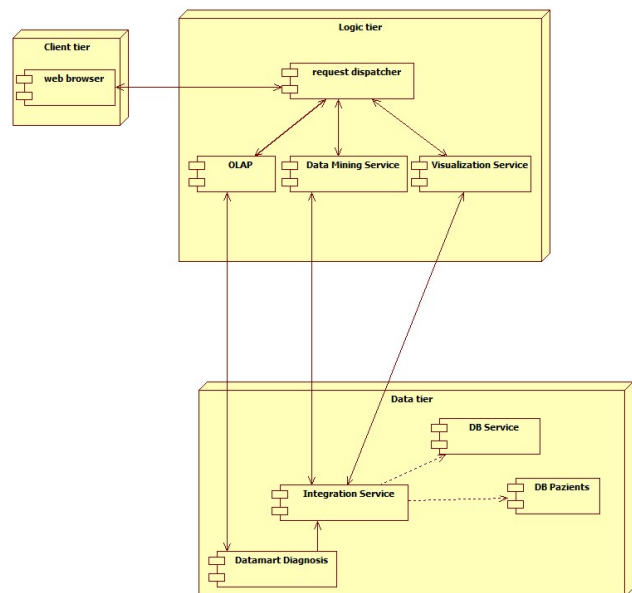


Figure 1. System architecture.

The client tier is composed of a web application representing the user interface.

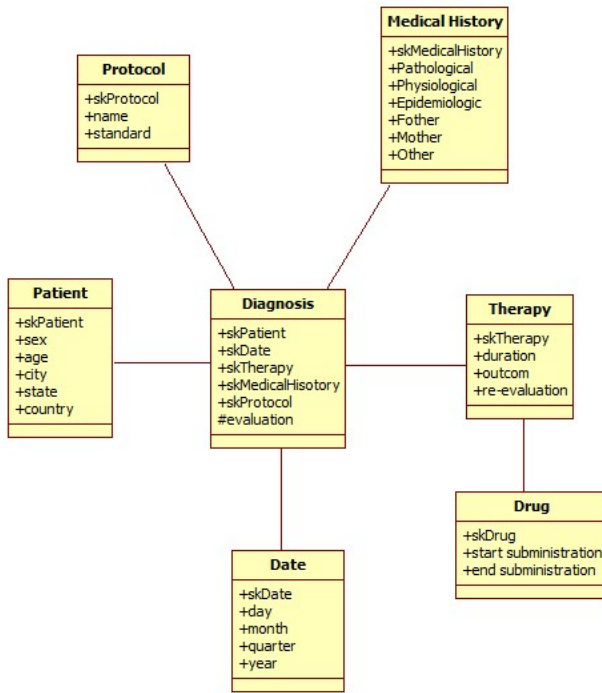
The *Logic Tier* provides two main services: Data Mining and Visualization, the former of which is responsible for the analysis of clinical records in order to find evidences in diagnostic-therapeutic processes of patients.

The *Data Tier* implements the data model of the application, and contains the *Healthcare Data Warehouse* providing the *OLAP Query and Visualization Service*, and whose

**Table 1. Selected Features**

Patient id	Patient sex	Patient nationality	...
Pathological medical history	Physiological medical history	Epidemiological medical history	...
Family medical history	Risk factors	Diagnosis	...
Clinical problems	Medical activities	Pharmacological Therapy	

star schema is shown in Fig. 2. Main efforts have been devoted to the analysis and reconciliation of the operational data sources [1].

**Figure 2. Logical Schema of the Data Mart.**

A usability test was carried out to validate the implemented system, through a one-to-one session using the think aloud technique. A group of students of our medical school were asked to use the system to examine medical records of several patients. Then, they were asked to fill in a questionnaire to report on the perceived usability. The questionnaire was subdivided into five categories (see Table 2), and answers were provided through a Likert scale [4]. The general evaluation was fairly good, and nearly the totality of them had a good reaction concerning system usage.

## 4. Conclusion

We have presented an health-care decision support system for medical guide lines identification, and have per-

formed a preliminary evaluation focusing on system usability. We are currently designing usage scenarios with medical doctors, in order to evaluate the quality of the inferred information, and to design additional inference mechanisms. Finally, since data privacy aspects are extremely critical in this context, we plan to extend the current system prototype with modules enabling the visual specification of role based access control mechanisms [3].

**Table 2. The usability questionnaire**

Category	Id	Question
General Evaluation	Q.2.1	The tool provides a nice user interface
	Q.2.2	Using the tool is simple
	Q.2.3	The aroused feeling by the tool use is satisfactory
Special Judgment	Q.3.1	The user interface is pleasant
	Q.3.2	The tool is simple to use
	Q.3.3	The tool proposes specific error messages
Tool Learning	Q.4.1	Learning to use the tool is simple
	Q.4.2	The required time to use the tool is appropriate
	Q.4.3	Remembering the commands and their use is simple
	Q.4.4	The number of steps to carry out a task is appropriate
	Q.4.5	The time to examine medical records and therapeutic protocols is appropriate
	Q.4.6	The number of steps to compare protocol applications is appropriate
	Q.4.7	Suggested Therapeutic Activities are correct
Information Grant	Q.5.1	Icon names and objects have a clear meaning
	Q.5.2	Each set of operations produces a predictable result

## References

- [1] L. Caruccio, V. Deufemia, and G. Polese. Visual data integration based on description logic reasoning. In *Proceedings of 18th International Database Engineering & Applications Symposium (IDEAS), Porto, Portugal*. ACM, 2014.
- [2] S. Chang, S. Orefice, G. Polese, and M. Tucci. A methodology and interactive environment for iconic language design. *International Journal of Human-Computer Studies*, 41(5):683–716, 1994.
- [3] M. Giordano and G. Polese. Visual computer-managed security: A framework to support the development of access control in enterprise applications. *IEEE Software*, 30(5):62–69, 2013.
- [4] A. N. Oppenheim. *Questionnaire design, interviewing, and attitude measurement*. Martin's Press, London, new edition, 1992.
- [5] R. C. Schank. *Conceptual Information Processing*. North-Holland Publishing Co, 1975.
- [6] N. Stolba and A. M. Tjoa. The relevance of data warehousing and data mining in the field of evidence-based medicine to support healthcare decision making. In K. Ardil, editor, *Computer Science*, pages 12–17, Volume 11, 2006. Enformatika. Vortrag: ICCS 2006, Prag, Czech Republic.
- [7] N. Ye. *Introduction In: The Handbook of Data Mining*. Lawrence Erlbaum Associates Publishers, 2003.

## Authors' Index

Ardito, Carmelo	104
Avola, Danilo	79, 260
Bellini, Pierfrancesco	8, 94, 211
Benigni, Monica	211
Berjawi, Bilal	365
Bertani, Luca	405
Billero, Riccardo	211
Bora, Darshan	416
Bottoni, Paolo	79, 104, 121
Brylow, Dennis	329
Buono, Paolo	276
Cabitza, Federico	47
Casaburi, Luca	64
Casagrande, Milena	399
Ceriani, Miguel	121
Chadha, Karishma	421
Chang, Shi-Kuo	86, 204
Chapman, Peter	310
Chen, Huan	128, 168
Chen, Wenzhi	39
Cheng, Bo-Chao	128, 168
Chesneau, Elisabeth	365
Chien, Min-Sheng	128
Chu, Yuan-Sun	168
Cinque, Luigi	260
Coccoli, Mauro	375, 383
Colace, Francesco	64, 196
Colazzo, Luigi	399
Cooper, Kendra	329
Costabile, Maria Francesca	104, 276
Costagliola, Gennaro	231, 321
Cui, Jing	3
Cunty, Claire	365
Dahleen, Jacob	329
de Lumley, Henry	345
De Rosa, Mattia	321
De Santo, Massimo	64, 196
Delaney, Aidan	300
Deshpande, Priya	416
Desolda, Giuseppe	104
Deufemia, Vincenzo	345

Di Giovanni, Pasquale	354
Di Martino, Sergio	357
Dib, Fadi	283
Ding, Yaoming	22
Duchateau, Fabien	365
Díaz, Paloma	335
Eicholtz, Matthew	267, 391
Eloe, Nathan	151
Favetta, Franck	365
Feng, Nan	30
Fogli, Daniela	47
Fuccella, Vittorio	231, 321
Fujinami, Kaori	250
Gatto, Ivano	189
Ginige, Athula	354
Greco, Luca	64, 196
Guercio, Angela	375
Göbel, Richard	182
Hallberg, Niklas	17
Hawash, Amjad	79
Herranz, Sergio	335
Hsu, Ya-Chi	168
Hu, Zhigang	175
Hunsberger, Alex	329
Jin, Jesse S.	72
Jungert, Erland	17
Kara, Levent B.	391
Kara, Levent Burak	267
Kong, Jun	113
Kropf, Carsten	182
Kuo, Ting-Chun	128
Kwoczek, Simon	357
Lanzilotti, Rosa	276
Laurini, Robert	365
Leopold, Jennifer	151, 241
Levialdi, Stefano	260
Li, Dong-Jin	57
Li, Jiping	22
Li, Mao-Lin	204
Li, Yi-Na	57
Li, Zhe	143

Liao, Guo-Tan	128
Liu, Weibin	3, 160
Longstreet, Shaun	329
Lu, Wei	135
Lu, Xuequan	39
Ma, Hua	175
Maresca, Paolo	375
Mascardi, Viviana	345
Matera, Maristella	104
Micallef, Luana	292
Miquel, Maryvonne	365
Molinari, Andrea	399
Moscato, Vincenzo	196
Müller, Sven	182
Nejdl, Wolfgang	357
Nesi, Paolo	8, 94, 211
Nie, Yifan	221
North, Max	415
Ouyang, Yuanxin	221
Paolino, Luca	345
Pascuccio, Fernando Antonio	231
Peng, Baolin	221
Petracca, Andrea	260
Picariello, Antonio	196
Piccinno, Antonio	47
Picozzi, Matteo	104
Pittarello, Fabio	189, 405
Placidi, Giuseppe	260
Polese, Giuseppe	345, 423
Rauch, Nadia	211
Ren, Cheng	160
Ren, Gaojun	72
Ren, Jinchang	72
Rodgers, Peter	283, 292, 310
Romero, Rosa	335
Rong, Wenge	221
Roudaki, Ali	113
Sabharwal, Chaman	151, 241
Sato, Makoto	250
Sebillo, Monica	354
Song, Song	160
Spezialetti, Matteo	260

Stanganelli, Lidia	375
Stapleton, Gem	300, 310
Steif, Paul S.	391
Steurer, Joseph	151
Su, Kui	39
Sun, Meijun	72
Taylor, John	300
Thompson, Simon	300
Torre, Ilaria	383
Tortora, Genny	354
Tucci, Maurizio	354
Turbak, Franklyn	421
Venturi, Alessandro	94
Vitiello, Giuliana	354
Wadströmer, Niclas	17
Walia, Gurisimran	113
Wang, Guangwei	22
Wang, Zan	30
Wang, Zheng	72
Wang, Zhenhua	30
Wang, Zonghui	39
Ward, Katrina	241
Weber, Ryan	329
Wei, Ruxiang	160
Xing, Weiwei	3, 135, 160
Xiong, Zenggang	22
Xiong, Zhang	221
Yao, Chi	57
Ye, Conghuan	22
Yu, Xue	30
Yung, Duncan	1, 86, 239, 372, 413
Zhang, Kaibing	22
Zhang, Kang	57
Zheng, Jiangbin	143
Zhu, Tingge	143
Zong, Wei	135



## Program Committee's Index

Tim Arndt  
Danilo Avola  
Arvind Bansal  
Andrew Blake  
Paolo Bottoni  
Paolo Buono  
Alfonso Cardenas  
Augusto Celentano  
Maiga Chang  
Shi Kuo Chang  
Ing-Ray Chen  
Shu-Ching Chen  
William Cheng-Chung Chu  
Yuan-Sun Chu  
Mauro Coccoli  
Francesco Colace  
Luigi Colazzo  
Kendra Cooper  
Gennaro Costagliola  
Philip Cox  
Alfredo Cuzzocrea  
Sergiu Dascalu  
Andrea De Lucia  
Massimo De Santo  
Aidan Delaney  
Massimo Desanto  
Vincenzo Deufemia  
Tiansi Dong  
David Du  
Larbi Esmahi  
Filomena Ferrucci  
Andrew Fish  
Paul Fishwick  
Daniela Fogli  
Manuel Fonseca  
Rita Francese  
Vittorio Fucella  
Kaori Fujinami  
David Fuschi  
Ombretta Gaggi  
Angelo Gargantini  
Nikolaos Gkalelis  
Luca Greco  
Angela Guercio  
Niklas Hallberg  
Carlos Iglesias  
Pedro Isaías

Joaquim Jorge  
Erland Jungert  
Levent Burak Kara  
Jun Kong  
Yau-Hwang Kuo  
Robert Laurini  
Jennifer Leopold  
Lian Li  
Fuhua Lin  
Hong Lin  
Alan Liu  
Jonathan Liu  
Wei Lu  
Paolo Maresca  
Sean McDirmid  
Andrea Molinari  
Vincenzo Moscato  
Paolo Nesi  
Kia Ng  
Sethuraman Panchanathan  
Ignazio Passero  
Joseph Pfeiffer  
Antonio Picariello  
Antonio Piccinno  
Giuseppe Polese  
Elvinia Riccobene  
Michele Risi  
Peter Rodgers  
Teresa Roselli  
Veronica Rossano  
Giuseppe Scanniello  
Monica Sebillio  
Buntarou Shizuki  
Peter Stanchev  
Lidia Stanganelli  
Gem Stapleton  
Mahbubur Syed  
Christian Timmerer  
Genny Tortora  
Jiri Trnka  
Franklyn Turbak  
Giuliana Vitiello  
Weiwei Xing  
Kazuo Yana  
Atsuo Yoshitaka  
Ing Tomas Zeman  
Kang Zhang  
Roberto Zicari

## External Reviewers' Index

Cenni, Daniele  
Chapman, Peter  
De Rosa, Mattia  
De Santo, Massimo  
Di Giovanni, Pasquale  
Eicholtz, Matthew  
Fasano, Fausto  
Fuccella, Vittorio  
Greco, Luca  
Indelli Pisano, Valentina  
Nappi, Michele  
Palomba, Fabio  
Pantaleo, Gianni  
Pesare, Enrica  
Sabharwal, Chaman  
Sun, Meijun  
Tian, Haiman  
Tucci, Maurizio  
Wang, Zheng  
Yumer, Mehmet Ersin



# DMS

Pittsburgh  
August 27-29, 2014

# 2014

**Proceedings of the Twentieth  
International Conference on  
Distributed Multimedia Systems**

Copyright © 2014  
Printed by  
Knowledge Systems Institute  
Graduate School  
3420 Main Street  
Skokie, Illinois 60076  
(847) 679-3135  
office@ksi.edu  
www.ksi.edu  
Printed in USA, 2014  
ISBN 1-891706-36-5 (paper)  
ISSN 2326-3261 (print)

