

# DMS 2015

An aerial photograph of the Vancouver skyline at sunset. The city is densely packed with skyscrapers, many of which have their lights on. The sky is a warm orange and yellow, and the mountains in the background are silhouetted against the light. The water of the harbor is visible on the right side of the image.

**Proceedings of the Twenty-First  
International Conference on  
Distributed Multimedia Systems**

**and**

**Journal of Visual Languages and  
Sentient Systems, Volume 1, 2015**

Vancouver, Canada  
August 31 - September 2, 2015



**PROCEEDINGS**

**DMS 2015**

**The 21<sup>st</sup> International Conference on  
Distributed Multimedia Systems**

**Sponsored by**

KSI Research Inc. and Knowledge Systems Institute Graduate School, USA

**Technical Program**

**August 31 – September 2, 2015**

**Hyatt Regency, Vancouver, Canada**

**Organized by**

KSI Research Inc. and Knowledge Systems Institute Graduate School, USA



Copyright © 2015 by KSI Research Inc. and 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: 1-891706-38-1

ISSN: 2326-3261 (print)

2326-3318 (online)

DOI: 10.18293/DMS2015

Additional copies can be ordered 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 KSI Research Inc. and Knowledge Systems Institute Graduate School, USA.

Printed by KSI Research Inc. and Knowledge Systems Institute Graduate School, USA.



# FOREWORD

Welcome to the 21<sup>st</sup> International Conference on Distributed Multimedia Systems (DMS 2015) that takes place this year in Vancouver, in beautiful British Columbia, Canada. This year conference follows a well-established sequence of yearly appointments where researchers from academia, industry, and government agencies from around the world meet to discuss issues, ideas, and innovations within the areas of multimedia and sentient systems.

With the continuous evolution of media, social media, and semantic computing, many challenges continue to arise and distributed systems capable of interacting with the environment by gathering, processing, interpreting, storing and retrieving multimedia information originating from sensors, robots, actuators, websites, and other information sources have become one the main areas of research. Those distributed systems, which go by the name of sentient systems, are the primary focus of this year's conference. However the DMS conference is not limited simply to research on multimedia sentient systems, and with the support of two additional workshops, one on Distance Education Technology (DET 2015) and one on Visual Languages and Computing (VLC 2015), the DMS conference continues to provide an international forum for discussion that expands also into the areas of education and visual languages.

The conference is organized in sessions which focus on many specialized topics. Exchange of ideas, discussions, research results, and experiences in the longstanding history of the conference have had a positive influence on the research in the past, and we believe that they will have a positive influence this year as well, thanks to the quality of the meeting and the research contributions from researchers from many countries. With the high quality of this year's technical program, the DMS community will continue to be an important venue and a source of new ideas and innovations.

We have received 50 submissions and the paper selection was based upon a rigorous review process, with an acceptance rate for full papers of 46%. We are expecting authors and guests from 10 countries: Canada, Chile, China, France, Italy, Japan, The Netherlands, Sweden, UK, and USA. This year's conference program contains contributions of high quality research papers, short papers, posters and demo to discuss ongoing research activities and applications.

DMS 2015 is pleased to welcome Prof. Franklyn Turbak as keynote speaker. Prof. Franklyn Turbak is Professor at Wellesley College. He is internationally renowned in the domain of visual languages and visual programming.

The DMS conference also has the pleasure to welcome the inauguration of a new important research journal in the area of sentient systems, the Journal of Visual Languages and Sentient Systems (JVLSS). Consequently, up to 6 papers and 2 research notes will be selected for inclusion in the inaugural issue to be published as part of the DMS2015 Proceedings. We believe that this journal will be an important venue for research in the area of multimedia sentient systems and we wish this new journal great success. In addition to this, up to 8 papers will be invited and further reviewed for possible inclusion in the special issues on best papers from DMS2015, to be published in December 2015 in Journal of Visual Languages and Computing (JVLC). Only papers presented at the DMS2015 will be considered for this special issue. Invitation will be made after the DMS2015 conference.

As Program Co-Chairs, we would like to express all our gratitude and appreciation to the Steering Committee Chair Dr. S.K. Chang for his support, dedication to the conference, and his invaluable



experience. However the high quality of DMS 2015 technical program would not have been possible without the tireless efforts of many individuals. First of all, we would like to thank the entire Steering Committee for their continuous support and guidance; the entire Program Committee whose invaluable, attentive, and timely work has made possible the creation of a high quality technical program. Then, we would like to extend our sincere appreciation to all the authors who have submitted their papers to the conference, thus contributing with their work and ideas to the success of this venue. Last but not least, we like to acknowledge the important contribution of the KSI staff whose assistance and support has been truly remarkable throughout the entire organization process.

On behalf of the Program Committee, Co-Chairs and the entire Program Committee, we are delighted to extend to you our really warm welcome to the 21th International Conference on Distributed Multimedia Systems (DMS 2015). We hope that you will find this year conference an exciting place for ideas exchanges, for fostering new projects, and a rewarding place for your research challenges. We wish you a nice staying in Vancouver and we hope that you will find some time to enjoy, among the other things, the beauty that the city offers.

Angela Guercio and Mahbubur Syed  
DMS 2015 Program Co-Chairs



# **DMS 2015**

## **The 21<sup>st</sup> International Conference on Distributed Multimedia Systems**

**August 31 – September 2, 2015**

**Hyatt Regency, Vancouver, Canada**

### **Conference Organization**

#### **DMS'15 Conference Chair**

Erland Jungert, Linkoping University, Sweden; conference chair

#### **DMS'15 Steering Committee Chair**

Shi-Kuo Chang, University of Pittsburgh, USA; Steering Committee Chair

#### **DMS'15 Steering Committee**

Paolo Nesi, University of Florence, Italy; Steering Committee Member

Kia Ng, University of Leeds, UK; Steering Committee Member

#### **DMS'15 Program Co-Chairs**

Angela Guercio, Kent State University, USA; Program Chair

Mahbubur Syed, Minnesota State University, USA; Program Co-Chair

#### **DMS'15 Program Committee**

Arvind K. Bansal, Kent State University, USA

Andrew Blake, University of Brighton, UK

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  
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  
Max North, Southern Polytechnic State University, 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  
Genny Tortora, University of Salerno, Italy  
Atsuo Yoshitaka, JAIST, Japan  
Ing Tomas Zeman, Czech Technical University, Czech Republic  
Kang Zhang, University of Texas at Dallas, USA

### **DET'15 Workshop Chair**

Paolo Maresca, University Federico II, Napoli, Italy; DET'15 workshop chair

### **DET'15 Program Co-Chairs**

Maiga Chang, Athabasca University, Canada; DET'15 Program Co-Chair  
Rita Francese, University of Salerno, Italy; DET'15 Program Co-Chair

### **DET'15 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  
Rita Francese, University of Salerno, Italy  
Angelo Gargantini, University of Bergamo, Italy  
Angela Guercio, Kent State University, USA  
Robert Heller, Athabasca University, Canada  
Pedro Isaias, Open University, Portugal  
Paolo Maresca, University Federico II, Napoli, Italy  
Andrea Molinari, University of Trento, Trento, Italy  
Mario Arrigoni Neri, University of Bergamo, Italy  
Ignazio Passero, University of Salerno, Italy  
Michele Risi, University of Salerno, Italy  
Teresa Roselli, University of Bari, Italy  
Veronica Rossano, University of Bari, Italy  
Giuseppe Scanniello, University of Basilicata, Italy  
Lidia Stanganelli, Universita Telematica eCampus, Novedrate (CO), Italy

### **VLC'15 Workshop Chair**

Franklyn A Turbak, Wellesley College, USA; VLC'15 Workshop chair

### **VLC'15 Program Chair**

Gem Stapleton, University of Brighton, UK; VLC'15 Program chair

### **VLC'15 Program Committee**

Bilal Alsallakh, Vienna University of Technology, Austria  
Danilo Avola, University of Rome, Italy  
Paolo Bottoni, Universita Sapienza, Italy  
Paolo Buono, University of Bari, Italy  
Alfonso F. Cardenas, University of California, USA  
Peter Chapman, University of Brighton, UK  
Gennaro Costagliola, University of Salerno, Italy  
Aidan Delaney, University of Brighton, UK  
Vincenzo Deufemia, University of Salerno, Italy  
Filomena Ferrucci, University of Salerno, Italy  
Andrew Fish, University of Brighton, UK  
Vittorio Fuccella, University of Salerno, Italy  
Levent Burak Kara, Carnegie Mellon University, USA  
Robert Laurini, University of Lyon, France  
Jennifer Leopold, Missouri University of Science & Technology, USA  
Luana Micallef, Helsinki Institute for Information Technology, Finland  
Nikolay Mirenkov, University of Aizu, Japan

Joseph J. Pfeiffer, Jr., New Mexico State University, USA  
Peter Rodgers, University of Kent, UK  
Monica Sebillio, University of Salerno, Italy  
Gem Stapleton, University of Brighton, UK  
Franklyn Turbak, Wellesley College, USA  
Giuliana Vitiello, University of Salerno, Italy

### **DMS/DET/VLC'15 Publicity Co-Chairs**

Kao-Shing Hwang, National Chung Cheng University, Taiwan; Publicity Co-Chair  
Lidia Stanganelli, University of Napoli Federico II, Italy; Publicity Co-Chair

# Keynote

## Democratizing Programming with Blocks Languages

Franklyn Turbak

Wellesley College

**Abstract:** In blocks programming languages (such as Scratch, Blockly, App Inventor, Snap!, Pencil Code, Alice/Looking Glass, AgentSheets/AgentCubes), programs are constructed by connecting visual blocks shaped like puzzle pieces. Through activities like Code.org's Hour of Code and both online and traditional courses, these languages have become extremely popular ways to introduce programming and computational thinking to tens of millions of people of all ages and backgrounds. By lowering barriers to programming in key programming language dimensions (syntax, static semantics, and dynamic semantics), blocks languages are helping to democratize programming by putting the power of programming in the hands of nonexperts. In my talk, I will focus on blocks language work done in the context of MIT App Inventor and the Wellesley College TinkerBlocks research project. Despite recent advances in blocks languages, there are still many challenges to address, including enhancing their usability and expressiveness, developing paths for transitioning to more traditional programming, and dealing with the perception that they are just toy languages for kids. I encourage members of the DMS community to join me in investigating these challenges.

**About the Speaker:** Franklyn Turbak is an associate professor of Computer Science at Wellesley College. His interests include the design, analysis, and implementation of expressive programming languages and visual representations of programs and computational processes. He is co-author of the textbook *Design Concepts in Programming Languages*. As head of the Wellesley TinkerBlocks research group, member of the MIT App Inventor development team, and lead PI on the NSF-funded Computational Thinking Through Mobile Computing project, his current goal is to improve the expressiveness and pedagogy of blocks programming languages.



# Panel Discussion

## Future of VL Research and Sentient Systems

### Session chair and moderator

Shi-Kuo Chang, University of Pittsburgh, USA ([chang@cs.pitt.edu](mailto:chang@cs.pitt.edu))

### Panelists

Gennaro Costagliola, University of Salerno, Italy ([gencos@unisa.it](mailto:gencos@unisa.it))

Gem Stapleton, University of Brighton, UK ([g.e.stapleton@brighton.ac.uk](mailto:g.e.stapleton@brighton.ac.uk))

Franklyn A Turbak, Wellesley College, USA ([fturbak@wellesley.edu](mailto:fturbak@wellesley.edu))

Paolo Nesi, University of Florence, Italy ([paolo.nesi@unifi.it](mailto:paolo.nesi@unifi.it))

**Panel Description:** The success of visual languages especially iconic languages is evident to everyone because most smart phones these days use iconic languages to communicate with the end user. Ironically the success of visual languages in practice has led to doubt and uncertainty about the future of visual languages research. The advances of sentient systems can motivate more research on visual languages. Therefore panelists are invited to explore the future of VL research and sentient systems. Panelists can discuss the theoretical implications as well as practical implementations of next generation visual languages, investigate the relations between visual languages and visualization, the impact of big data research and other related topics. Description of example research projects and the introduction of new research paradigms are especially welcome. The panelists will present their views. Comments from the audience are also welcome.

### Position Statements from the Panelists

**Gennaro Costagliola:** *A graphical review of visual language research* (co-authored with Vittorio Fuccella and Stefano Perna)

We summarize two decades (1995-2014) of research on visual languages through a timeline of terms, similar to the work done by Panisson and Quaggiotto in [1]. To this end, we extracted terms from the titles of the papers published in the considered twenty years on the journal *Journal of Visual Languages and Computing* (JVLC) and on the proceedings of the *IEEE Symposium on Visual Languages and Human-Centric Computing* (VL/HCC).

Basically, the methods used for creating the visualization can be summarized as follows:

- **Data Gathering and Processing.** We downloaded titles from DBLP digital library and extracted terms from them using Apache Lucene [2].

- **Graph Creation.** A graph is created with two types of nodes: *year* and *stemmed term*. The former type represents the year of publication (for a total of 20 nodes), the latter type represent the terms resulting from the stemming process. Each

*term* node is connected to one or more *year* nodes through an edge whose weight is equal to the frequency of the term in that year.

- **Graph Visualization.** The graph layout was built through the Gephi visualization tool [3]. The *year* nodes were positioned linearly in fixed, equally spaced locations. The position of the *term* nodes, was established through a force-directed layout exploiting the frequency information.

The final visualization is shown in Figure 1. The nodes of the timeline are colored in red, while the color of the nodes corresponding to terms expresses, together with the size, the term frequency. In order to avoid cluttering, the edges are omitted. Due to the particular adopted visualization technique discussed above, the position of terms along the timeline is close to the weighted average year of occurrence. As an example, the appearance of the term *grammar* in a medium/small sized font and close to year 1997 means that grammars were moderately considered in visual language research through 90's and at the beginning of year 2000.

As we can see in the graph, the research of the late 90's was characterized by topics as search queries and video databases. In the first decade of the 2000s the researchers mainly worked on topics such as virtual reality, modeling of diagrams through UML and semantic Web. In recent years emerged, among others, topics such as sketch recognition and visual representation of source code.

### References

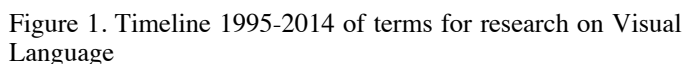
- [1] A. Panisson and M. Quaggiotto. From cortisone to graphene: 60 years of breakthroughs in pubmed publications. *WebSci 2014 Data Visualization Challenge*, 2014.
- [2] The Apache Software Foundation. Apache lucene - ultra-fast search library and server. <https://lucene.apache.org/>, 2015.
- [3] The Gephi Consortium. Gephi - the open graph viz platform. <http://gephi.github.io/>, 2015.

The development of visual languages has a long history and this field has evolved considerably over the last few decades. The research community has reached a stage where we can identify advantages and pitfalls of existing visual language designs and we have insight into the infrastructure needed for their use. Such infrastructure can include automated theorem proving (reasoning) support, automated diagram layout, and sketch recognition interfaces.

There have been many successful visual languages developed in the past, with UML being a prominent example, but also cases where such languages may not have delivered their full potential. To effectively translate research on visual languages into practice, target end-users need to see, and be convinced of, the benefit of adopting a new approach.

It is suggested that the community could benefit from a methodology that can be used for designing, developing and evaluating new visual languages. Such a methodology should support the process from conception of the visual language to end-user take-up. The development of visual languages should pay due regard to the types of support needed to make them practically useful. There could be clear benefits of such a methodology: the development process would be more robust and less likely to overlook key elements that are necessary for the resulting language to be truly fit-for-purpose. Of course, it is perhaps unlikely that one methodology could be sufficiently comprehensive and flexible enough to support all types of visual languages and application areas, but step in this direction could reap dividends.

Thus, it is proposed that the time is right for providing an overarching framework that guides visual language design, development and evaluation.



## **Franklyn Turbak: *Visual Languages and the Future of Programming***

The convergence of desktop computers, mobile phones, cameras, GPS navigators, web browsers, high-resolution touch-sensitive displays, and other personal electronic devices has resulted in smartphones and tablets — powerful personal computers that we carry with us everywhere and use constantly. At the same time, the entities we want to control in our environment have exploded — including these devices themselves, the social/communication networks and Internet-based information sources to which they're connected, and a host of objects in the emerging Internet of Things (e.g., home appliances, robots, wearable computers, sensor networks, large displays, 3D printers, drones).

In this new situated and ubiquitous computing landscape, there is an increasing need for everyone to specify computational behaviors involving all of these kinds of entities. Smartphone and tablet owners want to at least customize and compose existing apps for their devices and in some cases create new apps for themselves and their communities. Artists, scientists, business people, and civil servants need domain-specific ways to create artifacts and to gather, generate, simulate, analyze, and visualize information specific to their interests.

Spreadsheets, visual dataflow languages, and blocks-based syntactic representations of control-flow languages have emerged as popular ways for those without a programming background to specify computational behaviors.

But these only scratch the surface of what is possible. How can we leverage other advances from the visual languages community (including sketch and gesture recognition, diagram understanding, tangible user interfaces, informative animations, and programming by example) to make programming systems that device users find understandable, easy-to-use, and sufficiently powerful? Some key challenges include the relatively small displays of these devices relative to computer screens, effectively utilizing the sensors and actuators of these devices for programming purposes, providing high-level composable computational abstractions, and developing new visual models of the dynamic aspects of computational processes that enable end users to create and debug programs.

## **Paolo Nesi: *Sentient Multimedia Systems***

The era of artificial intelligence is probably overcome or revitalized by a number of new challenges and research lines. Among them, the autonomous vehicles, the personal assistant, the smart cities, the resilient solutions. I feel that these issues and topics are only a few of the new research areas that are going to appear, while are those in which I feel to be more confident and making some observations. In most of these activities, the human behavior is observed and modeled by using several different technologies and mathematical instruments. Some of them can be called cyber-physical and/or sociotechnical systems. Most of them are distributed systems

in which some capabilities of reasoning are enforced. The enforcement is performed by modeling, via machine learning, semantic computing and deductive system, mathematical and statistical and heuristic approaches, etc. In most cases these systems are acquiring large amount of data, to create evolving models on the basis of which, at the presentation on new or similar situation, they may produce outputs. The data are acquired continuously observing the users and as well as posing questions to the users as in the last generation of voice answering systems. To this end the models constructed for representing the knowledge are evolving as well. The recent technologies are capable of processing huge amount of data and thus disciplines as Big Data, data analytics, and statistic are getting a new momentum, together with data visualization and visual analysis tools. A new generation of models and applications of visual languages connected to this world would be needed to cope with manipulating and evolving data related to user profiles, conceptual models, questions and answers, data processing and analytics, recommended system, adaptive system, etc.

The new solutions are going to expose step by step more capabilities of smartness, adaptability and resilience to unexpected events and conditions. One could think that we are becoming more intelligent. In the common belief there is the idea to see as the next implied/expected step the construction of sentient systems, but only when the computer system will be much faster and capable to store more information, and thus not earlier than 10 years. To this end, there is a demand of Quantum Artificial Intelligence or in any way of quantum computing capabilities. It is not clear if this massive capability is needed to really create an intelligent system or just to study the phenomena and learning how to do it. Probably the second, since our computers are already capable to do things that we are not capable to do. Thus, the real computational needs and storage would be probably demonstrated much limited when the first really sentient system will be developed. A change of paradigm is needed to dominate the new challenge; we need new instruments and may be to advance the former instruments of knowledge modeling, expressive languages, decision making and executable models, heuristic and probabilistic engines, etc. To this end, large and multidisciplinary teams could be needed to conquer the new domain in which computational aspects and human factors are going to be fuse and enrich each other.

## **Shi-Kuo Chang: *Future Directions for Journal of Visual Languages and Sentient Systems***

The interesting ideas and proposed directions discussed in this panel will lead to special issues, focused topics and special sections in the journal of visual languages and sentient systems (VLSS). Anyone who has ideas and thoughts on future directions for VL is welcome to write a Viewpoints article for VLSS, and the length of such articles shall not exceed three double-column formatted pages.



## Table of Contents

<b>Foreword</b> .....	iii
<b>Conference Organization</b> .....	v
<b>KeyNote: Democratizing Programming with Blocks Languages</b>	
<b>Professor Franklyn Turbak</b> .....	ix
<b>Panel: The Future of VL Research and Sentient Systems</b> .....	x

---

### DMS-1

---

Effective Removal of Artifacts from Views Synthesized using Depth Image Based Rendering .....	65
<i>Danyang Zhao, Jiangbin Zheng and Jinchang Ren</i>	
Cross-Covariance-based Features for Speech Classification in Film Audio .....	72
<i>Matt Benatan and Kia Ng</i>	
A Symbolic Representation of Motion Capture Data for Behavioral Segmentation (S) .....	78
<i>Ruziang Wei, Weibin Liu and Weiwei Xing</i>	
Shadow Detection in Complex Environments via An Innovative Information Fusion Approach (S) .....	85
<i>Alfredo Cuzzocrea, Enzo Mumolo, Alessandro Moro, Kazunori Umeda and Gianni Vercelli</i>	
A Complete Classification of Occlusion Observer's Point of View for 3D Qualitative Spatial Reasoning (S) .....	94
<i>Chaman Sabharwal</i>	

---

### VLC-1

---

WiSPY: A Tool for Visual Specification and Verification of Spatial Integrity Constraints ..	39
<i>Vincenzo Del Fatto, Vincenzo Deufemia, Luca Paolino and Sara Tumati</i>	
BopoNoto: An Intelligent Sketch Education Application for Learning Zhuyin Phonetic Script (S) .....	101
<i>Paul Taele and Tracy Hammond</i>	
Fast prototyping of visual languages using local context-based specifications .....	14
<i>Gennaro Costagliola, Mattia De Rosa and Vittorio Fuccella</i>	
Visually Mapping Requirements Models to Cloud Services (S) .....	108
<i>Shaun Shei, Aidan Delaney, Stelios Kapetanakis and Haralambos Mouratidis</i>	

---

### DET-1

---

Pupils's collaboration around a large display .....	115
<i>Rosa Lanzilotti, Carmelo Ardito, Maria Francesca Costabile, Antonella De Angeli and Giuseppe Desolda</i>	

Experiencing a New Method in Teaching Databases using Blended eXtreme Apprenticeship. (S) .....	124
<i>Vincenzo Del Fatto, Gabriella Dodero and Roberta Lena</i>	
A Smart Material Interfaces Learning Experience .....	131
<i>Andrea Minuto, Fabio Pittarello and Anton Nijholt</i>	
Semantic video annotation for accessible resources in flipped classrooms (S) .....	141
<i>Ilaria Torre and Gianni Vercelli</i>	
Digitally Enhanced Assessment in Virtual Learning Environments .....	148
<i>Pierpaolo Di Bitonto, Enrica Pesare, Teresa Roselli and Veronica Rossano</i>	

---

**DMS-2**


---

A Distributed Framework for NLP-Based Keyword and Keyphrase Extraction From Web Pages and Documents .....	155
<i>Paolo Nesi, Gianni Pantaleo and Gianmarco Sanesi</i>	
Discovery and registration of components in multimodal systems distributed on the IoT ..	162
<i>Bertha Helena Rodriguez and Jean-Claude Moissinac</i>	
Vehicle Type Identification Based on Car Tail Text Information (S) .....	172
<i>Ruixue Yin, Weibin Liu and Weiwei Xing</i>	

---

**VLC-2**


---

Incremental indexing of objects in pictorial databases .....	23
<i>Castellano Giovanna, Fanelli Anna Maria and Torsello Maria Alessandra</i>	
RankFrag: A Machine Learning-Based Technique for Finding Corners in Hand-Drawn Digital Curves .....	29
<i>Gennaro Costagliola, Mattia De Rosa and Vittorio Fuccella</i>	
Generative Interface Structure Design for Supporting Existing Objects .....	5
<i>Nurcan Gecer Ulu and Levent Kara</i>	

---

**DET-2**


---

Design of an Educational Adventure Game to teach computer security in the working environment .....	179
<i>Ciro D'Apice, Claudia Grieco, Luca Liscio and Rossella Piscopo</i>	
Reward Points Calculation based on Sequential Pattern Analysis in Educational Mobile App (S) .....	186
<i>Bo-Shi Li, Rita Kuo, Maiga Chang and Kristin Garn</i>	

---

**DET-3**


---

Multimedia data integration and processing for E-Government .....	191
<i>Luca Greco, Francesco Colace, Flora Amato, Vincenzo Moscato and Antonio Picariello</i>	
Learning a Semantic Space by Deep Network for Cross-media Retrieval .....	199
<i>Zhao Li, Wei Lu, Egude Bao and Weiwei Xing</i>	

Intelligent Agent and Virtual Game to support education in e-health .....	204
<i>Enrica Pesare, Teresa Roselli and Veronica Rossano</i>	

---

**DMS-4**


---

A Surveillance System with SIS Controller for Incident Handling using a Situation-based Recommendations Handbook .....	212
<i>Erland Jungert and Shi-Kuo Chang</i>	
Graph Databases Lifecycle Methodology and Tool to Support Index/Store Versioning ....	221
<i>Paolo Nesi, Pierfrancesco Bellini and Ivan Bruno</i>	
GO-Bayes Method for System Modeling and Safety Analysis .....	49
<i>Guoqiang Cai</i>	

---

**DMS-5**


---

Robust Radial Distortion Estimation Using Good Circular Arcs .....	231
<i>Xiaohui Zhang, Weibin Liu and Weiwei Xing</i>	
Differential Evolutionary Algorithm Based on Multiple Vector Metrics for Semantic Similarity Assessment in Continuous Vector Space.....	241
<i>Yuanyuan Cai, Wei Lu, Xiaoping Che and Kailun Shi</i>	
PhysQSR: Improving Reasoning in Three Dimensions and Time With Image Processing and Physics (S) .....	250
<i>Nathan Eloie and Jennifer Leopold</i>	

---

**DMS-6**


---

TEco: an integration model to augment the Web with a trust area for inter-pares interactions. (S) .....	257
<i>Gennaro Costagliola, Vittorio Fuccella and Fernando Antonio Pascuccio</i>	
Joint Fingerprinting and Encryption for JPEG Images Sharing in Mobile Social Network (S).....	264
<i>Conghuan Ye, Zenggang Xiong, Yaoming Ding, Guangwei Wang, Xuemin Zhang and Fang Xu</i>	
An Interaction Mining Approach for Classifying User Intent on the Web.....	274
<i>Loredana Caruccio, Vincenzo Deufemia and Giuseppe Polese</i>	
MOSAIC+: tools to assist virtual restoration.....	284
<i>Daniel Riccio, Sonia Caggiano, Maria De Marsico, Riccardo Distasi and Michele Nappi</i>	

---

**DET-3**


---

On the Experience of Using Git-Hub in the Context of an Academic Course for the Development of Apps for Smart Devices.....	292
<i>Rita Francese, Carmine Gravino, Michele Risi, Giuseppe Scanniello and Genoveffa Tortora</i>	



Teaching Computer Programming in a Platform as a Service Environment (S) .....	300
<i>Mauro Coccoli, Paolo Maresca, Lidia Stanganelli and Angela Guercio</i>	

---

**Poster**


---

Testing a Storytelling Tool for Digital Humanities (P) .....	59
<i>Fabio Pittarello</i>	
A Quick Survey on Sentiment Analysis Techniques: a lexical based perspective (P) .....	62
<i>Luca Greco, Francesco Colace, Vincenzo Moscato, Flora Amato and Antonio Picariello</i>	

<b>Author's Index</b> .....	A-1
<b>Program Committee's Index</b> .....	A-4
<b>External Reviews' Index</b> .....	A-6

**Note:**

(S) indicates a short paper.

(P) indicates a poster.

**Journal of  
Visual Languages and Sentient  
Systems**





# **Journal of Visual Languages and Sentient Systems**

## **Editor-in-Chief**

**Shi-Kuo Chang, University of Pittsburgh, USA**

## **Co-Editors-in-Chief**

**Gennaro Costagliola, University of Salerno, Italy**

**Paolo Nesi, University of Florence, Italy**

**Gem Stapleton, University of Brighton, UK**

**Franklyn Turbak, Wellesley College, USA**

**An Open Access Journal published by**

**KSI Research Inc.**

**and**

**Knowledge Systems Institute Graduate School**

# **VLSS Editorial Board**

Tim Arndt, Cleveland State University, USA

Alan F. Blackwell, University of Cambridge, United Kingdom

Paolo Bottoni, University of Rome, Italy

Francesco Colace, University of Salerno, Italy

Maria Francesca Costabile, University of Bari, Italy

Philip T. Cox, Dalhousie University, Canada

Martin Erwig, Oregon State University, USA

Vittorio Fuccella, University of Salerno, Italy

Angela Guercio, Kent State University, USA

Erland Jungert, Swedish Defence Research Establishment, Sweden

Kamen Kanev, Shizuoka University, Japan

Robert Laurini, University of Lyon, France

Mark Minas, University of Munich, Germany

Brad A. Myers, Carnegie Mellon University, USA

Joseph J. Pfeiffer, Jr., New Mexico State University, USA

Genny Tortora, University of Salerno, Italy

Kang Zhang, University of Texas at Dallas, USA

Copyright © 2015 by KSI Research Inc. and 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: 1-891706-38-1

ISSN: 2326-3261 (print)

2326-3318 (online)

DOI: 10.18293/VLSS2015

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

Printed by KSI Research Inc. and Knowledge Systems Institute Graduate School, USA.

# Journal of Visual Languages and Sentient Systems

Volume 1, 2015

## Table of Content

<b>Preface</b>	4
<b>Regular Papers</b>	
Nurcan Gecer Ulu and Levent Kara, “Generative Interface Structure Design for Supporting Existing Objects”	5
Gennaro Costagliola, Mattia De Rosa and Vittorio Fuccella, “Fast prototyping of visual languages using local context-based specifications”	14
Castellano Giovanna, Fanelli Anna Maria and Torsello Maria Alessandra, “Incremental indexing of objects in pictorial databases”	23
Gennaro Costagliola, Mattia De Rosa and Vittorio Fuccella. RankFrag: A Machine Learning-Based Technique for Finding Corners in Hand-Drawn Digital Curves”	29
Vincenzo Del Fatto, Vincenzo Deufemia, Luca Paolino and Sara Tumiatì, “WiSPY: A Tool for Visual Specification and Verification of Spatial Integrity Constraints”	39
Guoqiang Cai, “GO-Bayes Method for System Modeling and Safety Analysis”	49
<b>Research Notes</b>	
Fabio Pittarello, “Testing a Storytelling Tool for Digital Humanities”	59
Luca Greco, Francesco Colace, Vincenzo Moscato, Flora Amato and Antonio Picariello, “A Quick Survey on Sentiment Analysis Techniques: a lexical based perspective”	62

# PREFACE

The Journal of Visual Languages and Sentient Systems (VLSS) is intended to be a forum for researchers, practitioners and developers to exchange ideas and research results, for the advancement of visual languages and sentient multimedia systems. The success of visual languages especially iconic languages is evident to everyone because most smart phones these days use iconic languages to communicate with the end user. Ironically the success of visual languages in practice has led to doubt and uncertainty about the future of visual languages research. However the advances of sentient systems can motivate more research on visual languages, both at the practical level and at the theoretical level.

Sentient systems are distributed systems capable of actively interacting with the environment by gathering, processing, interpreting, storing and retrieving multimedia information originated from sensors, robots, actuators, websites and other information sources. In order for sentient systems to function efficiently and effectively, visual languages may play an important role. To stimulate research towards that goal, the Journal of Visual Languages and Sentient Systems is born.

VLSS publishes research papers, state-of-the-art surveys, review articles, in the areas of visual languages, sentient multimedia systems, distributed multimedia systems, sensor networks, multimedia interfaces, visual communication, multi-media communications, cognitive aspects of sensor-based systems, and parallel/distributed/neural computing & representations for multimedia information processing. Papers are also welcome to report on actual use, experience, transferred technologies in sentient multimedia systems and applications. Timely research notes, viewpoint articles, book reviews and tool reviews, not to exceed three pages, can also be submitted to VLSS.

Manuscripts shall be submitted electronically to VLSS. Original papers only will be considered. Manuscripts should follow the double-column format and be submitted in the form of a pdf file. Page 1 should contain the article title, author(s), and affiliation(s); the name and complete mailing address of the person to whom correspondence should be sent, and a short abstract (100-150 words). Any footnotes to the title (indicated by \*, +, etc.) should be placed at the bottom of page 1.

Manuscripts are accepted for review with the understanding that the same work has not been and will not be nor is presently submitted elsewhere, and that its submission for publication has been approved by all of the authors and by the institution where the work was carried out; further, that any person cited as a course of personal communications has approved such citation. Written authorization may be required at the Editor's discretion. Articles and any other material published in VLSS represent the opinions of the author(s) and should not be construed to reflect the opinions of the Editor(s) and the Publisher.

Shi-Kuo Chang  
Editor-in-Chief  
Journal of Visual Languages and Sentient Systems

# Generative Interface Structure Design for Supporting Existing Objects

Nurcan Gecer Ulu  
Carnegie Mellon University

Levent Burak Kara  
Carnegie Mellon University

**Abstract**—Increasing availability of high quality 3D printing devices and services now enable ordinary people to create, edit and repair products for their custom needs. However, an effective use of current 3D modeling and design software is still a challenge for most novice users. In this work, we introduce a new computational method to automatically generate an organic interface structure that allows existing objects to be statically supported within a prescribed physical environment. Taking the digital model of the environment and a set of points that the generated structure should touch as an input, our biologically inspired growth algorithm automatically produces a support structure that when physically fabricated helps keep the target object in the desired position and orientation. The proposed growth algorithm uses an attractor based form generation process based on the space colonization algorithm and introduces a novel target attractor concept. Moreover, obstacle avoidance, symmetrical growth, smoothing and sketch modification techniques have been developed to adapt the nature inspired growth algorithm into a design tool that is interactive with the design space. We present the details of our technique and illustrate its use on a collection of examples from different categories.

## I. INTRODUCTION

The customization and personalization of products started to compete with traditional mass production principles with the contribution of maker movement and DIY (Do-It-Yourself) culture. DIY commonly refers to any fabrication, modification or repair event that is outside of one's professional expertise [1]. With the rise of DIY culture, there is a growing interest for design and fabrication tools tailored towards non-expert users.

Recent advances in 3D design and manufacturing technologies now have made content creation accessible to novice users. Besides the basic consumer level 3D printers, online on-demand 3D printing services (e.g. Shapeways, i.Materialise) have enabled ordinary people to access high quality machines. 3D modeling software, such as Autodesk 123D and Tinkercad, allow consumers to create 3D shapes using simplified geometric interaction methods. However, current commercial design software do not take advantage of capabilities of 3D printing. While *almost anything* can be fabricated using 3D printing, these design software limit potential design outputs by mimicking features of traditional manufacturing and assembly methods. In this work, we extend the design possibilities by taking a generative design approach to create organic looking branching shapes that would be challenging to design and fabricate with traditional methods.

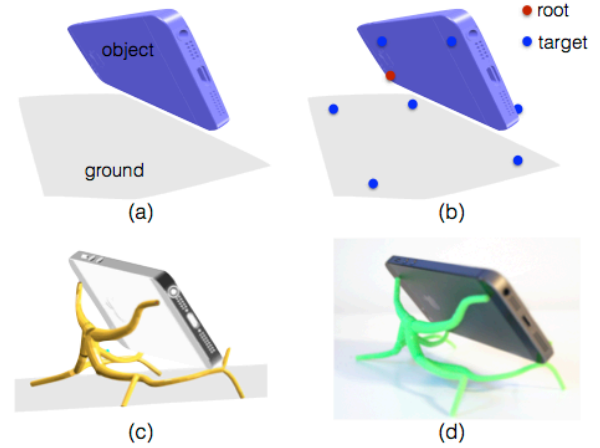


Fig. 1. Example problem to generate a phone stand (a) given object and environment configuration (b) user defined target and root points (c) generated interface structure (d) 3D printed result.

We propose a framework that automatically generates interface structures under prescribed constraints. The input to our algorithm is a surface mesh for the object to support and a mesh to represent the ground surface with target and root points to create a shape in between (Fig.1). Then, automatic interface structure generation is achieved by a nature inspired growth mechanism. Users can control the design by changing target-root combinations at the input phase as well as by using sketch modifications after the shape is created. Moreover, the stochastic nature of the growth algorithm lets users design one of a kind pieces by generating different outputs for the same problem on each run. The main contribution of this work is the novel application of a nature inspired growth algorithm for automatic product generation. This is accomplished by the introduction of target attractor and pruning concepts, embedding product design considerations and user interaction.

## II. RELATED WORK

**Design Tools for Non-Expert Users** have recently received significant attention. In [2], a chair design tool is proposed to create balanced chairs from extruded 2D profile sketches. To enable informed exploration, Umetani et al. [3] proposed a suggestive design tool for plank-based furniture. In that work, the user adds planks and edits their positions, orientations or size. A data-driven approach to interactive design of model

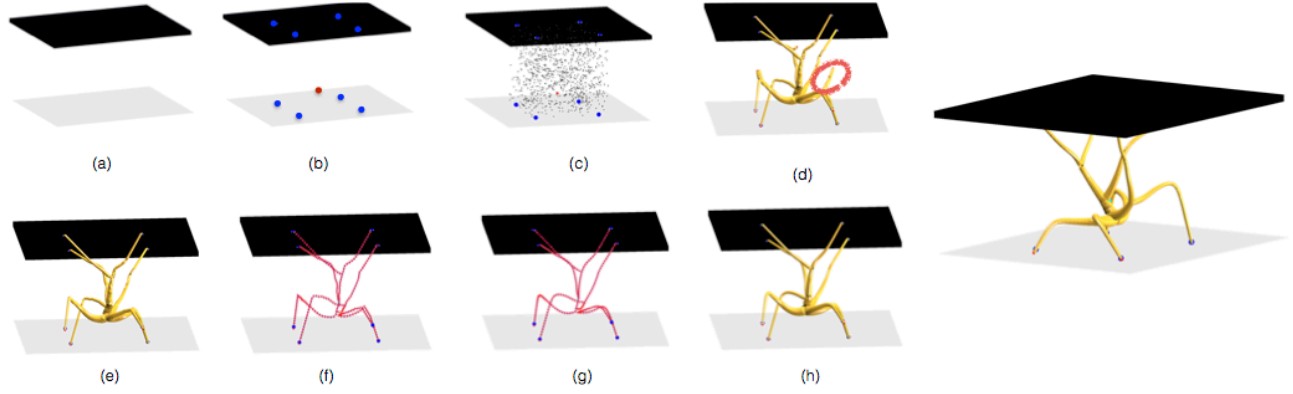


Fig. 2. Interface structure generation process. First, user defines the environment configuration (a) and selects the root and target points shown as red and blue, respectively (b). Attractors are generated randomly in the design space (c) and interface structure is grown automatically (d). Then, unnecessary branches are removed automatically (d-e) and the skeleton of the interface structure is smoothed as desired (f-h). Resulting shape is shown on the right.

airplanes is proposed in [4] where the user creates free-flight gliders with 2D sketches. In this work, we focus on creating a large range of products instead of one specific group such as chairs or gliders. While all three systems are notable interactive tools, components of the resulting designs are limited to 2D laser cut pieces. Our system generates organic 3D geometries that can take advantage of the opportunities in 3D printing.

Much of recent research on design for 3D printing addresses modifications of existing digital models by optimizing physical properties, such as balance and structural strength. For this purpose, inner carving with deformations [5, 6] and thickening of thin sections [7] have been used. Here, we focus on geometric shape *generation* whereas their focus is on shape modification.

**Generative Design** methods are recognized as significant technologies to rapidly generate different design alternatives. Fabrication of generatively produced designs have been examined illustrating how geometrically complex shapes can be physically created in [8, 9]. In computer graphics, generative design methods have been used to create architectural models such as buildings [10], virtual cities [11] and trees [12, 13]. In this paper, we are inspired by a specific generative design method developed to simulate the tree growth process for automatic shape generation.

**Tree Growth** methods have been widely used in computer graphics for urban modeling and computer animation. As such, tree growth has been vastly investigated in the literature. L-systems have been used to generate trees in [12]. Runions et al. [13] proposed a space colonization algorithm to mimic open and closed venation process in the leaf formation. Later, the space colonization algorithm has been extended to grow trees in [14] and [15]. Tree generation inside various geometries is investigated using spatial attractor distribution in [16]. In this paper, our automatic interface generation process has been inspired by the space colonization algorithm. We use its spatial attractor distribution feature to enable the interaction with the design space. Moreover, interaction of the tree model with

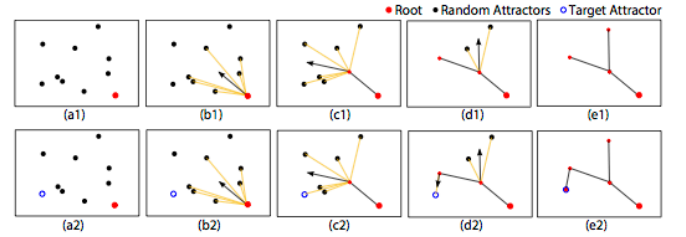


Fig. 3. Space colonization algorithm with (a1-e1) and without (a2-e2) target attractors.

the obstacles in the environment has been studied in [17]. In that, a fully grown tree model is placed around an object and the colliding branches are removed. For our purposes, this approach cannot be used since the grown structure has to be connected to the target points and a branch connected to a target can not be removed. Hence, we utilize obstacle avoidance during the growth process.

### III. AUTOMATIC SHAPE GENERATION

In this work, the aim is to automatically create an interface structure between given objects. This process is illustrated in Fig.2 starting with the user input and the steps of the automatic shape creation performed on the background. First, the user supplies the input geometries as 3D mesh models (a). Then, a set of target points are selected by the user to define where the interface structure should be in contact with the input models (b). Then, a root point or points are provided by the user to start the growth process (b). Attractors are randomly generated inside the design space (c). The structure is generated akin to a tree originating at its root and growing in 3D space to reach the targets (d). Branches that are not connected to the input objects through target points are removed from the structure (e). We also refer to this step as pruning or unnecessary branch removal. Finally, the skeleton of the structure (f) is smoothed (g-h). In the following sections, the details of these steps are described.



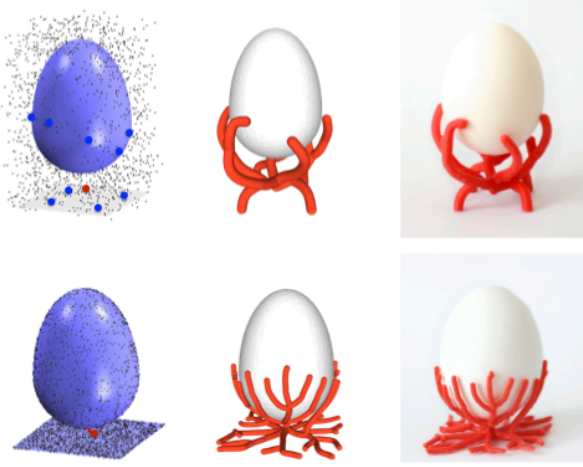


Fig. 4. Egg holder generated using volume (top) and surface (bottom) hormones. Left: problem setup, middle: digital model, right: 3D printed result.

#### A. Growth Algorithm

The proposed method uses an attractor based growth approach of space colonization algorithm given in [13]. Space colonization algorithm creates a branching tree structure in space as demonstrated in Fig.3.a1-e1. The tree structure grows without the guarantee that it would touch any specific point in the design environment. In this work, we need to create shapes between objects ensuring that the generated interface would be in contact with the target objects to support them. For this reason, we introduce a novel target attractor concept to create branching structures that grow to the required target positions (Fig.3.a2-e2).

The target based growth process starts with the definition of the design space, e.g. rectangle in (a2) and target-root point selections. Then, random attractors are sampled uniformly inside the design space. These random attractors have an *influence distance* that they can pull a branch to themselves as well as a *kill distance* that makes them inactive when they get too close to a branch in the growing structure. At every growth step, depending on the influence and kill distance, each attractor is associated with the tree node that is closest to it (yellow lines) if the node is within the influence distance. Then, normalized vectors from the node to the attractors are created and their average (black arrow) is calculated and used as the growth direction for the node (b2). The new node is added in the growth direction in the distance of branch length. All attractors are checked if they are in the kill distances of nodes. In other words, an attractor is killed if it is close enough to the tree (c2). This process iterates until all attractors are killed. While target attractors also pull the branches towards them, they are a special type of attractor with *zero* kill distance. If an attractor is a target attractor, it does not get *killed* until a tree node reaches it (notice difference in d1 and d2). The position of a new node is calculated as follows

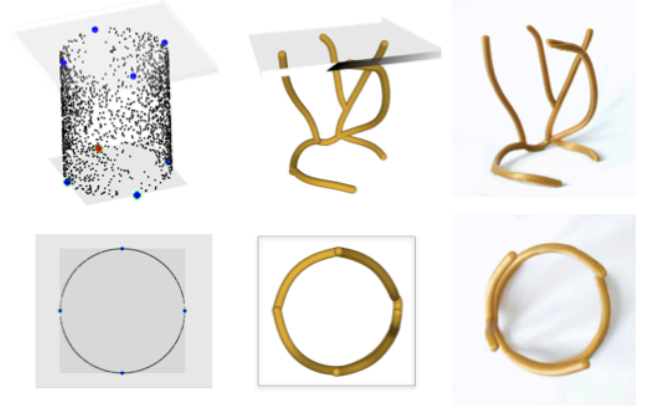


Fig. 5. Example projection growth. Generating the interface structure. Left: problem setup with root & target points illustrated, middle: digital model, right: 3D printed result. Orthogonal (top) and top (bottom) views of the same part are given.

$$\vec{v}' = \begin{cases} \vec{t}, & \text{if reaching target} \\ \vec{v} + L\hat{n}, & \text{otherwise} \end{cases} \quad (1)$$

$$\vec{n} = \frac{\vec{a} - \vec{v}}{\|\vec{a} - \vec{v}\|} \quad (2)$$

$$\hat{n} = \frac{\vec{n}}{\|\vec{n}\|} \quad (3)$$

$\vec{v}'$ ,  $\vec{v}$ ,  $\vec{t}$ ,  $L$ ,  $\hat{n}$ ,  $\vec{a}$  are the position vector of a new node, the position vector of node in the tree set, the position vector of the target attractor, branch length, unit growth direction vector and position vector of attractor in the set, respectively.

#### B. Random Attractor Placement

Placement of random attractors is a crucial step in our algorithm, especially to create variations for the same problem. Since the placement of the attractors defines the virtual design space in which our structure can grow, how attractors are placed in 3D drastically affects the resulting geometry. This effect is demonstrated for three distinct cases in Fig.4 and Fig.5. Figure 4 compares the use of volume and surface attractors to generate an interface structure between the same objects. In the first one, we use the bounding box volume of the two objects to generate the attractors inside of the volume. On the other hand, in the second one, attractors are sampled on the surface of these objects. From this figure, it can be seen that the resulting interface geometries with very distinct characteristics can be obtained by only changing the distribution of the attractors even for the same problem setting. Here, an important distinction between these two cases is that we do not require target attractors for the surface growth case simply because we are guaranteed to touch the surfaces of both objects in this case. Another distinct case for attractor distribution is illustrated in Fig. 5. Here, the aim is to generate an interface structure that would give a desired 2D profile



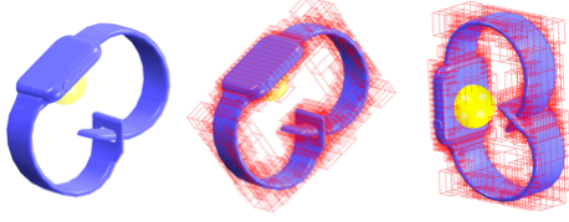


Fig. 6. Obstacle avoidance is used to restrict growth in specific parts in the space. The restricted regions may be the objects represented as octrees or user defined spheres.

when viewed from a certain direction. For this purpose, we sample the attractors on a surface that is created by extruding the desired 2D profile in the viewing direction. Hence, this specific attractor generation case can be classified as a subset of the surface growth explained previously. Moreover, any 3D swept/curved surface can be used to create 3D profiles. However, the main distinction here is that we require target attractors to be defined in this case, to ensure the resulting structure touches and supports the objects in the problem setup. This is mainly because the surface which the attractors generated on is a virtual one rather than the actual surface of the objects.

Apart from the aforementioned cases, we also enable symmetry in the resulting geometries. In product design, symmetry is considered to be a critical feature for everyday objects [18]. In our shape creation algorithm, we can ensure symmetry by simply placing the attractors in the design space symmetrically. Hence, the addition of a symmetry feature does not add any computational cost in our algorithm. However, the only case that may need special attention is the one where the root point is placed on the symmetry plane. In such cases, growth only happens on the symmetry plane because of the equal attraction from both sides. We solve this problem by moving the user defined root point slightly in both directions orthogonal to the symmetry plane by duplicating the root point.

### C. Obstacle Avoidance

When designing an object, its interaction with the environment is important. For this reason, structure growth may be restricted in some parts of the design space. First of all, the interface structure should not intersect with the objects that it is intended to support. For this purpose, we utilize mesh representations of the objects for collision detection. In addition, users may define geometric obstacles in the form of spheres to restrict the growth. As an example for the use of spheres for functional purposes, a sphere is placed under the smart watch to limit the growth of the interface structure on the magnetic touch charging area in Fig 6.

During growth, the intersection of the new branch and obstacles are tested at each step. If there is a collision between the obstacle and the branch, a random direction is chosen for growth until collision is eliminated or maximum number of trials is reached. Intersection tests are conducted using a

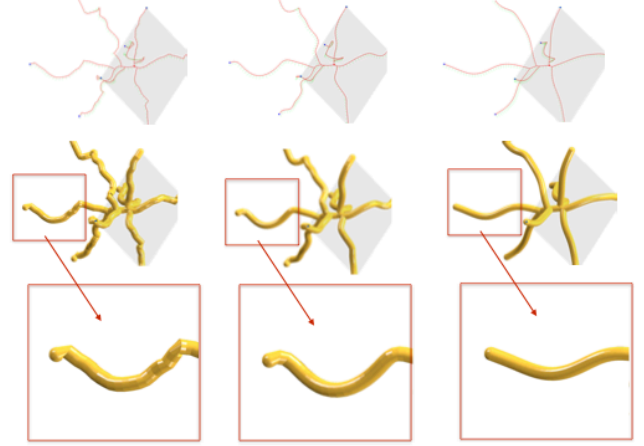


Fig. 7. Effect of Laplace and Biharmonic operators on smoothing is illustrated for skeleton (top) and skin (middle-bottom) of the resulting geometries. Left: the original, middle: after smoothing with Laplace & Biharmonic operators together, right: after smoothing with Laplace operator only. Note that the use of the combined Laplace & Biharmonic operators allows smoothing without significant shrinkage.

parametric representation of a line segment and an implicit representation of spherical and triangular objects. Details of the intersection test can be found in [19]. To increase the efficiency of collision detection for triangular meshes, we use octree representation [20].

### D. Smoothing

While jagged transitions between biological branches look realistic for trees, smooth transitions are usually more appealing in product design. For this purpose, we apply curve smoothing to the tree structure as shown in Fig.7. Here, the position of each node on the tree skeleton is updated based on the positions of the neighboring nodes using Laplacian and Biharmonic operators as follows [21, 22].

$$\vec{v}_i = \vec{v}_i + \lambda_1 \Delta \vec{v}_i + \lambda_2 \Delta(\Delta \vec{v}_i) \quad (4)$$

where Laplace and Biharmonic operators can be defined as:

$$\Delta \vec{v}_i = \nabla^2 \vec{v}_i = \frac{1}{2}(\vec{v}_{i+1} - \vec{v}_i) + \frac{1}{2}(\vec{v}_{i-1} - \vec{v}_i) \quad (5)$$

$$\Delta(\Delta \vec{v}_i) = \nabla^4 \vec{v}_i \quad (6)$$

$\vec{v}_{i-1}$  and  $\vec{v}_{i+1}$  denote two neighbors of the node,  $\vec{v}_i$ .

In order to achieve smooth curves, we linearly combined Laplace and Biharmonic operators. Although it is possible to accomplish smoothing with only the Laplacian term, Biharmonic term is included to suppress the shrinking behavior arising from the Laplace operator when used alone (Fig.7). In this paper, we select the coefficients  $\lambda_1$ ,  $\lambda_2$  as 0.2 and 0.1, respectively.

### E. Branch Pruning

As can be observed from Fig.2 and Fig.3, our approach creates many branches that may not serve a structural function on the interface (i.e., branches that do not touch a target point). Hence, branches that are not connected to the target object or ground are automatically detected and removed from the tree graph (Fig.2(d)-(e)).

### F. Modifications and Variation In The Design

Although we produce the interface structures automatically, we enable users to control many aspects of the geometry generation. The user control starts by importing 3D models of the objects and the selection of root-target attractor configurations. Then, another significant control comes from the placement of random attractors as explained previously. In addition to these inputs, there are four factors that affect the growth process (1) influence distance, (2) kill distance, (3) branch length and (4) number of random attractors. These factors are very important to generate variations in the space colonizations algorithm for tree generation such as trees with dense or sparse branch structures [14]. On the other hand, results of our proposed growth algorithm are not affected by the changes in those parameters primarily due to the target attractor and branch removal concepts. As long as the parameters are *suitable*, results do not change significantly. There is a wide range of suitable parameters for a given problem. Any suitable parameter set has the following properties:

- Influence distance is greater than kill distance.
- Branch length, which can be considered as step size, is small compared to the environment dimensions but it is large enough to facilitate efficient growth computation.
- The number of random attractors is high enough to create uniform distribution in the design space, we used 2000 attractors for the examples in the paper.
- Influence distance is high enough to enable attraction of a node for the created uniform distribution.

In this work, we choose default values using the given guidelines. For each problem setup, we use the default values by scaling them with the dimensions of the bounding box of the system.

Another set of important controls comes into play after the interface structure is generated automatically. At this point, users can control the radius variation in the branches of the interface structure as well as modify the skeleton of the structure by sketched strokes. Now that the skeleton of the structure is obtained, a 3D skin is created by covering each branch with a truncated cone and taking the union of all cones. The radius at each node is calculated based on its age as

$$r = r_{min} + (r_{max} - r_{min}) * e^{-k\alpha} \quad (7)$$

where  $r$ ,  $r_{min}$ ,  $r_{max}$ ,  $\alpha$ ,  $k$  is the radius, minimum radius, maximum radius, age and decay of radius, respectively. Here, age of each node is determined in such a way that every node starts with age of 0 and the age increases by 1 at each growth step. Decay of radius defines how fast the radius changes from

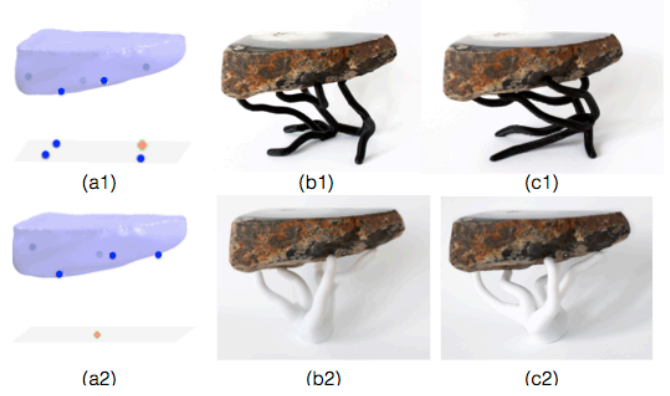


Fig. 8. Effect of target-root selection and stochastic growth demonstrated on two different target-root configurations (a1-c1, a2-c2). For the same problem setup (a), various stochastic growth results (b and c) can be obtained.

the root to the targets between maximum radius and minimum radius and is set by the user. Also,  $r_{min}$ ,  $r_{max}$  are set by the user.

Sketch modifications are performed through modifier sketches performed by the user to specify the new shape of the skeleton curve as it would occur from the current viewpoint. To do this, a surface is created by the rays emanating from the user's eyes, passing through the strokes and extends into the page. In theory, there are infinitely many candidate solutions on this surface. The best 3D configuration is thus found by computing the minimum distance projection of the original curve onto the surface. For the details of the sketch modifications please refer to [23].

## IV. RESULTS AND DISCUSSION

Our approach enables the automatic generation of interface structures for 3D printing. The user may control the geometry generation through target-root configurations, random attractor placement, skin radius selection and sketch modifications. We applied our algorithm to a variety of examples including gadget accessories, decoration and restoration of existing objects and furniture. In order to transform the existing objects into the digital design environment, we utilized 3D scanning using a Kinect device. We downloaded the 3D digital models through the stock 3D model websites like GrabCAD and Google 3D Warehouse for the common objects.

The latest trends in decorating and modern furniture design include hybrid design approaches where natural materials with imperfections are combined with machine-made parts to create innovative and original designs. In Figure 8, an example hybrid design created using our system is shown. Here, we take a natural rock piece and design a support structure that complements its organic geometrical features. One important control that our system provides is the target-root placement. We generated two different target-root configurations (a1, a2) for the same problem to demonstrate the significant variance in the resulting geometries (b1, b2). Moreover, we would like to draw attention to the stochastic nature of our algorithm





Fig. 9. Use of sketch modifications for a functional purpose. Top part of the planter holder is enlarged to be able to insert the planter. Left: Sketch modification steps, Right: 3D printed result with the inserted planter.

that comes from the random sampling of attractors inside the design space where different results are obtained for distinct set of random attractors. However, the effect of the stochastic nature on the resulting geometries (b1-c1 and b2-c2) are subtle compared to the effect of target-root selection.

Another direction for craft, arts and design is the restoration of broken objects through 3D printing to obtain new artistic expressions rather than restoring the original object [24]. In that, the motivation is not to restore the initial function of the object, but rather use it to function as a memorial. In Fig.10, we show that our method can be used for similar purposes. Here, a missing part of a broken vase is restored with the generated interface shape. For this, we first scanned the broken vase and placed desired target-root points. Then, the resulting piece to complete the broken part is grown using our algorithm. We also 3D printed and assembled the resulting part to the broken vase. Another alternative interface structure for this example can also be seen in Fig.12.a.

In addition to the aesthetic needs, the need for sketch modifications may arise from functional requirements. Use of sketch modifications for a functional purpose is illustrated in Fig.9. In this example, a hanger is designed to suspend the planter. However, the planter can not be inserted into this automatically generated structure. For this reason, sketch modifications are applied on the skeleton of the structure to enlarge the top part of the hanger so that the planter can be inserted.

In some configurations, users might need to control the growth process more strictly to achieve a geometry with particular desired properties. In such cases, our geometry creation process can be guided by progressively manipulating the problem setup. To do this, instead of defining all target points at once, we start with a subset of targets and progressively add the remaining ones as we grow the structure. Figure 11 demonstrates this on an example to attach the phone to a baseball cap for first person view camera shots. Here, the

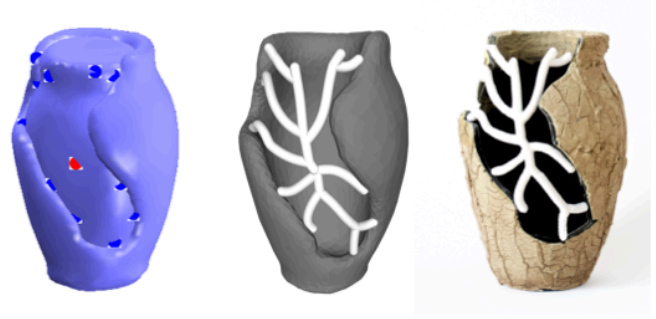


Fig. 10. Form completion: A broken vase is restored using a scanned model. Left: problem setup, middle: digital model, right: 3D printed result.

aim is to guide the growth on the side of the cap instead of any other possible outcome. First, only five of the targets are defined (a) and the growth process is completed (b). Then, a new target point is added (b) and another growth process is accomplished. This process is iterated (c) until the final desired shape is created (d). Since we are using a consumer level 3D printer with a limited build volume, we partitioned the resulting object into smaller pieces to be able to 3D print. For the assembly, we manually added dovetail structures on the assembly surfaces (f).

There are many communities that promote reuse of materials through community engagement, resource conservation and creativity e.g. Pittsburgh Center for Creative Reuse and Lancaster Creative Reuse. Since our design framework is developed to work with existing objects, users can easily utilize our algorithm for creative reuse purposes. A virtual example of material reuse is shown in Fig. 12.e. Here, the usage of a seat and back from a broken chair to design a new support and legs is demonstrated. In the example, while we have virtual models for the elements to be reused, as mentioned earlier any object can be scanned and used to create the interface structures. Although for the previous examples, we focus on creating attachment structures that hold the object in place without fixing or gluing, this example requires the interface structure to be fixed to the supported objects.

Since our framework is tailored towards non-expert users, we fabricated all our examples with a consumer level low-cost 3D printer, PrintrBot Simple 1405, to study the printability of our results. However, more advanced 3D printers can be used to fabricate resulting geometries with higher qualities using various material options.

We recorded the computation time for automatic shape generation for a number of examples. Since our method has a stochastic nature, computation time changes as the random attractor set changes. Thus, the results are reported for three different random attractor sets for each example in Table I. One reason computation time changes for each example is the change in the complexity of the objects that increases the time for collision checks. Another, important factor is how easy or difficult it is to reach the targets inside the design space.

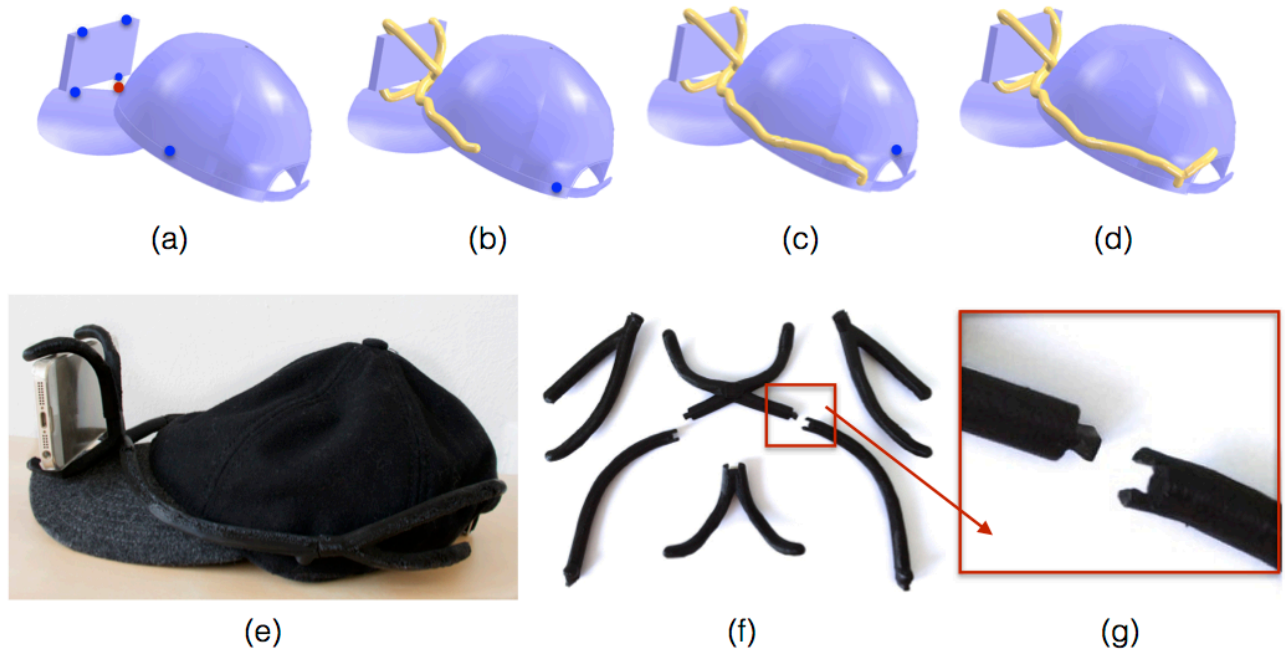


Fig. 11. Guided progressive growth (top) is shown on a baseball cap example to attach a phone for first person view camera shooting. Use of assembly structures to 3D print larger designs (bottom) have been demonstrated with the zoomed in dovetail joint detail (g).

TABLE I  
COMPUTATION PERFORMANCE OF OUR GENERATIVE DESIGN  
ALGORITHM

	Total Run Time [s]				
	Run 1	Run 2	Run 3	Run 4	Run 5
Fig.12.a	11	12	10	10	9
Fig.12.b	< 1	< 1	< 1	< 1	< 1
Fig.12.c	< 1	< 1	< 1	< 1	< 1
Fig.12.d	2	2	2	2	2
Fig.12.e	4	2	4	3	2

## V. LIMITATIONS AND FUTURE WORK

Our focus has been on generating tree-like skin structures on the skeletons we grow that makes the resulting designs resemble trees. We expect the proposed formulation to be readily applicable with different building blocks instead of our current truncated cones to achieve a richer variation in form. Moreover, since our obstacle avoidance is achieved through random search directions, our algorithm might not converge to a solution within predefined maximum number of trials. While we have observed this issue rarely, increasing maximum trial number for complex problem settings may be required. Finally, in this work, we do not consider structural performance of the resulting shapes. Yet, our algorithm can be extended to ensure structural soundness for a given problem configuration. This may require utilization of finite element analysis during the shape generation process.

## VI. CONCLUSION

We present a generative design framework to create interface structures to support existing objects. The proposed method enables novice users to automatically generate geometries and edit as well. Our approach introduces a novel application of a nature inspired growth algorithm with embedded product design considerations. Our current studies indicate that the approach works well for a variety of design problems with the presented actual 3D printed results alongside their digital models.

## ACKNOWLEDGMENT

Authors would like thank Ye Han for his help on the assembly structure creation.

## REFERENCES

- [1] C. Mota, "The rise of personal fabrication," in *Proceedings of the 8th ACM Conference on Creativity and Cognition*, ser. C&C '11. New York, NY, USA: ACM, 2011, pp. 279–288. [Online]. Available: <http://doi.acm.org/10.1145/2069618.2069665>
- [2] G. Saul, M. Lau, J. Mitani, and T. Igarashi, "Sketchchair: An all-in-one chair design system for end users," in *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '11. New York, NY, USA: ACM, 2011, pp. 73–80. [Online]. Available: <http://doi.acm.org/10.1145/1935701.1935717>
- [3] N. Umetani, T. Igarashi, and N. J. Mitra, "Guided exploration of physically valid shapes for furniture design," *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2012)*, vol. 31, no. 4, 2012.
- [4] N. Umetani, Y. Koyama, R. Schmidt, and T. Igarashi, "Pteromys: Interactive design and optimization of free-formed free-flight model airplanes," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 65:1–65:10, Jul. 2014. [Online]. Available: <http://doi.acm.org/10.1145/2601097.2601129>



- [5] R. Prévost, E. Whiting, S. Lefebvre, and O. Sorkine-Hornung, "Make It Stand: Balancing shapes for 3D fabrication," *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH)*, vol. 32, no. 4, pp. 81:1–81:10, 2013.
- [6] M. Bäcker, E. Whiting, B. Bickel, and O. Sorkine-Hornung, "Spin-It: Optimizing moment of inertia for spinnable objects," *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH)*, vol. 33, no. 4, 2014.
- [7] O. Stava, J. Vanek, B. Benes, N. Carr, and R. Měch, "Stress relief: Improving structural strength of 3d printable objects," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 48:1–48:11, Jul. 2012. [Online]. Available: <http://doi.acm.org/10.1145/2185520.2185544>
- [8] Y. Wang and J. P. Duarte, "Automatic generation and fabrication of designs," *Automation in Construction*, vol. 11, no. 3, pp. 291 – 302, 2002, rapid Prototyping. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0926580500001126>
- [9] L. Sass and R. Oxman, "Materializing design: the implications of rapid prototyping in digital design," *Design Studies*, vol. 27, no. 3, pp. 325 – 355, 2006, digital Design Digital Design. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0142694X05000864>
- [10] P. Müller, P. Wonka, S. Haegler, A. Ulmer, and L. Van Gool, "Procedural modeling of buildings," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 614–623, Jul. 2006. [Online]. Available: <http://doi.acm.org/10.1145/1141911.1141931>
- [11] Y. I. H. Parish and P. Müller, "Procedural modeling of cities," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '01. New York, NY, USA: ACM, 2001, pp. 301–308. [Online]. Available: <http://doi.acm.org/10.1145/383259.383292>
- [12] P. Prusinkiewicz, M. James, and R. Měch, "Synthetic topiary," in *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '94. New York, NY, USA: ACM, 1994, pp. 351–358. [Online]. Available: <http://doi.acm.org/10.1145/192161.192254>
- [13] A. Runions, M. Fuhrer, B. Lane, P. Federl, A.-G. Rolland-Lagan, and P. Prusinkiewicz, "Modeling and visualization of leaf venation patterns," in *ACM SIGGRAPH 2005 Papers*, ser. SIGGRAPH '05. New York, NY, USA: ACM, 2005, pp. 702–711. [Online]. Available: <http://doi.acm.org/10.1145/1186822.1073251>
- [14] A. Runions, B. Lane, and P. Prusinkiewicz, "Modeling trees with a space colonization algorithm," in *Proceedings of the Third Eurographics Conference on Natural Phenomena*, ser. NPH'07. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2007, pp. 63–70. [Online]. Available: <http://dx.doi.org/10.2312/NPH/NPH07/063-070>
- [15] W. Palubicki, K. Horel, S. Longay, A. Runions, B. Lane, R. Měch, and P. Prusinkiewicz, "Self-organizing tree models for image synthesis," in *ACM SIGGRAPH 2009 Papers*, ser. SIGGRAPH '09. New York, NY, USA: ACM, 2009, pp. 58:1–58:10. [Online]. Available: <http://doi.acm.org/10.1145/1576246.1531364>
- [16] S. Longay, A. Runions, F. Boudon, and P. Prusinkiewicz, "Treesketch: Interactive procedural modeling of trees on a tablet," in *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling*, ser. SBIM '12. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2012, pp. 107–120. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2331067.2331083>
- [17] S. Pirk, O. Stava, J. Kratt, M. A. M. Said, B. Neubert, R. Měch, B. Benes, and O. Deussen, "Plastic trees: interactive self-adapting botanical tree models," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 50:1–50:10, Jul. 2012. [Online]. Available: <http://doi.acm.org/10.1145/2185520.2185546>
- [18] B. De Mozota, *Design Management: Using Design to Build Brand Value and Corporate Innovation*. Skyhorse Publishing Company, Incorporated, 2003. [Online]. Available: [https://books.google.com/books?id=jpy\\_JBhZ7nUC](https://books.google.com/books?id=jpy_JBhZ7nUC)
- [19] P. Shirley and S. Marschner, *Fundamentals of Computer Graphics*, 3rd ed. Natick, MA, USA: A. K. Peters, Ltd., 2009.
- [20] J. Revelles, C. Urea, and M. Lastra, "An efficient parametric algorithm for octree traversal," in *Journal of WSCG*, 2000, pp. 212–219.
- [21] M. Botsch, L. Kobbelt, M. Pauly, P. Alliez, and B. uno Levy, *Polygon Mesh Processing*. AK Peters, 2010.
- [22] E. B. Arisoy, G. Orbay, and L. B. Kara, "Free form surface skinning of 3d curve clouds for conceptual shape design," *Journal of Computing and Information Science in Engineering*, vol. 12, p. 031005, 2012.
- [23] L. B. Kara and K. Shimada, "Sketch-based 3d-shape creation for industrial styling design," *IEEE Computer Graphics and Applications*, vol. 27, pp. 60–71, 2007.
- [24] A. Zoran and L. Buechley, "Hybrid reassemblage: an exploration of craft, digital fabrication and artifact uniqueness," *Leonardo*, vol. 46, no. 1, pp. 4–10, 2013.

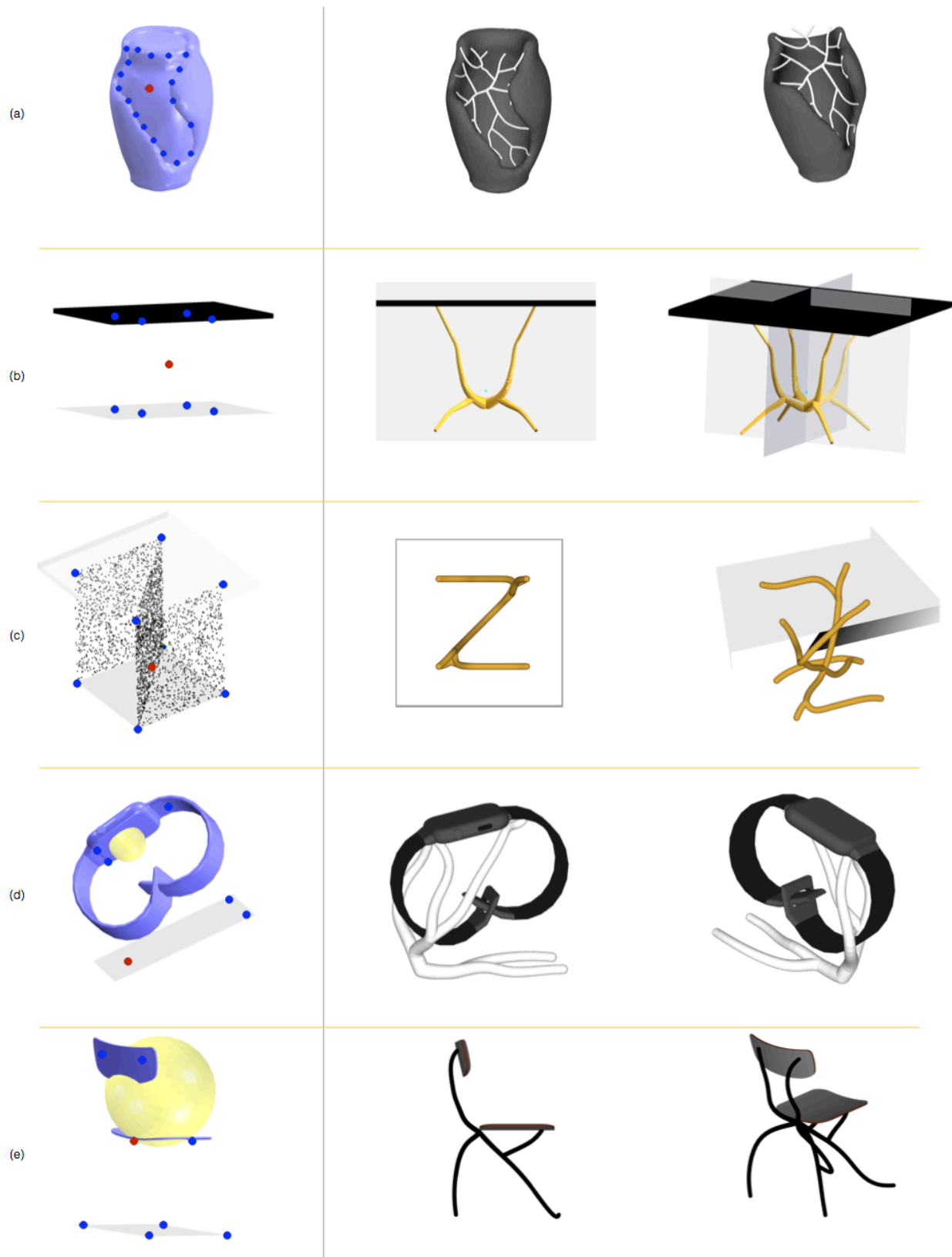


Fig. 12. Designs created with our system. Left: problem setup, right: resulting design from two different views.

# Fast prototyping of visual languages using local context-based specifications

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

*In this paper we present a framework for the fast prototyping of visual languages exploiting their local context based specification.*

*In previous research, the local context specification has been used as a weak form of syntactic specification to define when visual sentences are well formed. In this paper we add new features to the local context specification in order to fully specify complex visual languages such as entity-relationships, use case and class diagrams. One of the advantages of this technique is its simplicity of application and, to show this, we present a tool implementing our framework. Moreover, we describe a user study aimed at evaluating the satisfaction and effectiveness of users when prototyping a visual language.*

**Keywords:** *local context, visual languages, visual language syntax specifications.*

## 1 Introduction

Due to the ever growing evolution of graphical interfaces, visual languages are now a well established form of digital communication in many work and research environments. They are being used extensively by engineers, architects and others whenever there is the need to state and communicate ideas in a standardized way. This is traditionally happening, for example, with software engineering UML graphical notations but it is also catching on in other research fields, such as, for example, synthetic and system biology, [21, 16].

In the 1990s and the early 2000s, the formalization and implementation of visual languages have excited many researchers and many proposals have been defined. In particular, formalisms were defined mainly based on the extension of textual grammar concepts, such as attributed grammars

[13, 17, 9, 23] and graph grammars [3, 14], and on meta-modeling [4].

Lately, a new proposal for the specification and interpretation of diagrams from the only syntactic point of view has been given in [7]. This research is motivated by the need to reduce the complexity of the visual language syntax specification and originated while working on the recognition of sketch languages and on the difficulty of recognizing the symbols of the language. In order to disambiguate sketched symbols, the more knowledge is given on each symbol and on its possible interactions with the others, the better. The methodology, known as *local context-based visual language specification* requires the designer to define the *local context* of each symbol of the language. The local context is seen as the interface that a symbol exposes to the rest of the sentence and consists of a set of attributes defining the local constraints that need to be considered for the correct use of the symbol.

At first, this approach is, then, more lexical than syntactic: the idea is to provide a very detailed specification of the symbols of the language in order to reduce complexity when specifying the language at the sentence level. On the other hand, due to the easy-to-use and intuitiveness requirements, many practical visual languages have a simple syntax that can be completely captured by the local context specification as defined in [7]. To show this, the syntax of the binary trees and of a Turing complete subset of flowcharts have been completely specified through local context in [7].

When considered as a syntax specification, however, the simplicity of the approach is counterbalanced by its low expressiveness, especially with respect to the more powerful grammars formalisms.

In this paper, we define new features to be added to the original local context definition in order to push the expressiveness of the methodology and to allow the complete syntactic specification of complex visual languages such as entity-relationships, use case and class diagrams. We present the tool LoCoMoTiVE (Local Context-based Modeling of 2D Visual language Environments) implementing our framework which basically consists of a simple inter-



face allowing a user, in one screen, to define the symbols of the language and their local context. Moreover, we demonstrate the usability of the tool in a user study, in which participants are asked to define and test a visual language after a short introduction to the tool. Besides the participants' ability to define the visual language, we also administered a System Usability Scale (SUS) [5] questionnaire to evaluate their satisfaction with the tool.

The rest of the paper is organized as follows: the next section refers to related work; Section 3 gives the background information on the "local context-based visual language specification", sketches the three new features and presents a complete syntax specification for the use case diagrams; Section 4 is devoted to describe the tool; the experiment is presented in Section 5; some final remarks and a brief discussion on future work conclude the paper.

## 2 Related Work

In recent years many methods to model a diagram as a sentence of a visual language have been devised. A diagram has been represented either as a set of relations on symbols (the *relation-based approach*) [22] or as a set of attributed symbols with typed attributes representing the "position" of the symbol in the sentence (the *attribute-based approach*) [12]. Even though the two approaches appear to be different, they both consider a diagram (or visual sentence) as a set of symbols and relations among them or, in other words, a spatial-relationship graph [3] in which each node corresponds to a graphical symbol and each edge corresponds to the spatial relationship between the symbols.

Unlike the relation-based approach, where the relations are explicitly represented, in the attribute-based approach the relations must be derived by associating attribute values.

Based on these representations, several formalisms have been proposed to describe the visual language syntax, each associated to ad-hoc scanning and parsing techniques: Relational Grammars [23], Constrained Set Grammars [17], and (Extended) Positional Grammars [10]. (For a more extensive overview, see Marriott and Meyer [18] or Costagliola et al. [8].) In general, such visual grammars are specified by providing an alphabet of graphical symbols with their "physical" appearance, a set of spatial relationships generally defined on symbol position and attachment points/areas, and a set of grammar rules in context-free like format.

A large number of tools exist for prototyping visual languages. These are based on different types of visual grammar formalisms and include, among others, VLDesk [9], DiaGen [19], GenGed [2], Penguin [6], AToM3 [11], and VL-Eli [15].

Despite the context-free like rule format, visual grammars are not easy to define and read. This may explain why there has not been much success in transferring these

techniques from research labs into real-world applications. We observe that several visual languages in use today are simple languages that focus on basic components and their expressive power. These languages do not need to be described with complex grammar rules. Our approach provides a simpler specification for many of them, making it easier to define and quickly prototype visual languages.

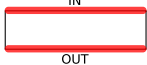
## 3 Local Context Specification of Visual Languages

According to the attribute-based representation, a visual sentence is given by a set of symbols defined by a set of typed attaching points linked through connectors. In [7], the local context specification of a visual language consists in the definition of the sets of graphical elements (named *symbols*) and spatial relations (named *connectors*) composing the language and, for each of them, their attributes. These are:

- Name of the graphical symbol;
- Graphical aspect (for example, specified through an svg-like format);
- Maximum number of occurrences of the symbol in a sentence of the language;
- Attachment areas: an attachment area is a set of points (possibly one) through which symbols and connectors can be connected. For each area the following attributes have been identified:
  - Name of the attachment area;
  - Position: the location of the attachment area on the symbol or connector;
  - Type: an attachment area of a symbol can be connected to an attachment area of a connector only if they have the same type;
  - Local constraints: such as the max number of possible connections for an attachment area.

As a further and sentence level constraint, a local-context based specification may assume that the underlying spatial-relationship graph of any sentence of the specified language is connected.

As an example, Table 1 shows the attributes of the statement symbol STAT and the connector ARROW when considered as alphabetical elements of a particular set of flowcharts. The specification states that STAT has the graphical aspect of a rectangle; it has two attaching areas named IN and OUT corresponding to the upper and lower sides of the rectangle, respectively; it may occur zero or more times in a flowchart; the attachment area IN can only be connected to an attachment area of a connector with type *enter*, while the attachment area OUT can only be connected to an attachment area of a connector with type *exit*. Moreover, IN may participate in one or more connections, while OUT may participate in only one connection. As re-

Symbol name	Graphics	Symbol occurrences	Attachment areas		
			name	type	constraints
STAT		$\geq 0$	<i>IN</i>	enter	$connectNum \geq 1$
			<i>OUT</i>	exit	$connectNum = 1$


Connector name	Graphics		Attachment areas		
			name	type	constraints
ARROW			<i>HEAD</i>	enter	$connectNum = 1$
			<i>TAIL</i>	exit	$connectNum = 1$

Table 1: Attribute specification for the symbol STAT and the connector ARROW.

gards the connector ARROW, its graphical appearance is given by a line with a head and the attachment areas are located to the head (HEAD) and tail (TAIL) of the arrow itself. An arrow can be connected to only one symbol through its head and only one symbol through its tail. In [7], complete local context specifications for a particular set of Turing complete flowcharts and for binary trees are given to show how local context can be used to fully specify the syntax of visual languages.

### 3.1 New Local Context Features

In order to capture as much as possible of the syntax of complex languages other than flowcharts and binary trees and to keep simplicity, new local features need to be added to the original definition of local context. In particular, we introduce the possibility of

- defining *symbol level* constraints involving more than one attaching area of a symbol/connector as opposed to constraints on individual attaching areas;
- assigning multiple types to attaching areas;
- defining constraints to limit connector self loops.

These features allow us to give complete local context specifications of complex languages such as the entity relationship diagrams, class diagrams, and use case diagrams. In the following, we show the practical usefulness of this extension by referring to the local context specifications of the use case diagrams and the entity-relationship diagrams.

*Symbol level constraints.* Table 2 shows the binary version of the *relation* symbol of the well-known entity-relationship (E-R) graphical notation. Each vertex of the symbol has one attaching point (Up, Down, Left or Right) of type *enter*. In order to force its use as an E-R binary relation (as opposed to ternary) the constraints need to set the sum of all their connections to two, apart from requiring that the number of connections to each attaching point be at most one. In this case, the feature simplifies the specification by avoiding that a designer define all the possible ways a binary relation symbol can be attached to the other symbols.

*Multiple types and Connector self loop constraints.* Table 3 shows the complete specification of the use case graphical notation, while Figure 1 shows some examples of correct and incorrect sentences. In the table, the language symbols and connectors are specified in the first and second part while, for sake of completeness, the last row declares, if present, any requirement at sentence level. It can be noted that the attachment area *Border* of the symbol ACTOR (first table row) has two types GenA and AssOut. By considering the Connector part of the table, this means that an ACTOR can be connected through its border to both the head and tail of the GENERALIZATION\_A connector (through GenA) and also to the attaching point *P1:P2* of the connector ASSOCIATION (through AssOut). Moreover, because of the constraint  $numLoop = 0$ , a GENERALIZATION\_A connector cannot be connected to the border of an ACTOR with its head and tail, simultaneously.

In Figure 1, case (b) shows the correct use of connector GENERALIZATION\_Uc, while case (d) shows the correct use of connector GENERALIZATION\_A.

The use of these new features allow the language designer more flexibility in the definition of the language. However, multiple types must be carefully used when dealing with connectors with the same graphical aspect since they may introduce ambiguities in the language.

It is not difficult to see that Table 3 completely specifies the syntax of the use case graphical notation as presented in <http://agilemodeling.com/artifacts/useCaseDiagram.htm> but without the optional “System boundary boxes”.

With respect to a grammar definition, the new specification is basically distributed on the language elements instead of being centralized.

As a final note, the selection of which language elements are symbols and which are connectors is completely left to the language designer. Moreover, connectors may not have a graphical representation (such as the relationships “touching”, “to the left of”).

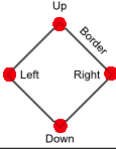
Symbol name	Graphics	Symbol occurrences	name	Attachment points	constraints
BIN_REL		$\geq 0$	$\begin{array}{l} Up \\ Left \\ Right \\ Down \\ Border \end{array}$	$\begin{array}{l} ConA \\ ConA \\ ConA \\ ConA \\ ConB \end{array}$	$\begin{array}{l} connectNum(X) \leq 1 \text{ for } X = \\ Up, Down, Left, Right \wedge (connectNum(Up) + \\ connectNum(Down) + connectNum(Left) + \\ connectNum(Right) = 2) \wedge \\ (connectNum(Border) \geq 0) \end{array}$

Table 2: ER binary relation specification.

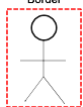




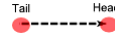
Symbol name	Graphics	Symbol occurrences	name	Attachment points	constraints
ACTOR		$\geq 1$	Border	$\begin{array}{l} GenA \\ AssOut \end{array}$	$\begin{array}{l} connectNum \geq 0 \wedge numLoop = 0 \\ connectNum \geq 0 \end{array}$
USE_CASE		$\geq 1$	Border	$\begin{array}{l} GenUc \\ AssIn \\ Dep \end{array}$	$\begin{array}{l} connectNum \geq 0 \wedge numLoop = 0 \\ connectNum \geq 0 \\ connectNum \geq 0 \wedge numLoop = 0 \end{array}$
Connector	Graphics		name	Attachment points	constraints
ASSOCIATION			$\begin{array}{l} P1:P2 \\ P2:P1 \end{array}$	$\begin{array}{l} AssOut \\ AssIn \end{array}$	$\begin{array}{l} connectNum = 1 \\ connectNum = 1 \end{array}$
GENERALIZATION_A			$\begin{array}{l} Head \\ Tail \end{array}$	$\begin{array}{l} GenA \\ GenA \end{array}$	$\begin{array}{l} connectNum = 1 \\ connectNum = 1 \end{array}$
GENERALIZATION_UC			$\begin{array}{l} Head \\ Tail \end{array}$	$\begin{array}{l} GenUc \\ GenUc \end{array}$	$\begin{array}{l} connectNum = 1 \\ connectNum = 1 \end{array}$
DEPENDENCY			$\begin{array}{l} Head \\ Tail \end{array}$	$\begin{array}{l} Dep \\ Dep \end{array}$	$\begin{array}{l} connectNum = 1 \\ connectNum = 1 \end{array}$
<b>Non local constraint</b>					
<i>the spatial-relationship graph must be connected</i>					

Table 3: Use case diagrams language specifications.

## 4 The tool LoCoMoTiVe

The current implementation of the local context methodology includes the new set of constraints defined in the previous section and is composed of two different modules:

- LoCoModeler: the local context-based specification editor, and
- TiVe: a web-based visual language environment for editing and checking the correctness of the visual sentences.

### 4.1 LoCoModeler

The LoCoModeler module allows designers to create and edit visual language specifications based on local con-

text, quickly and easily. Its output is the formal definition in XML format of the language that will be used during the disambiguation and the recognition of diagrams. Once the language designer has completed the specification, s/he can compile it into a web-based environment (the TiVe module) to allow users to draw sentences and verify their correctness. During language definition, this feature also allows the designer to check the correctness of the specification.

The main view of the LoCoModeller GUI is shown in Figure 2. Its main components are:

- A textbox containing the name of the language and a checkbox to enable/disable the option that diagrams must necessarily be connected;
- A table reporting the main information of symbols and connectors included in the language. It is possible to

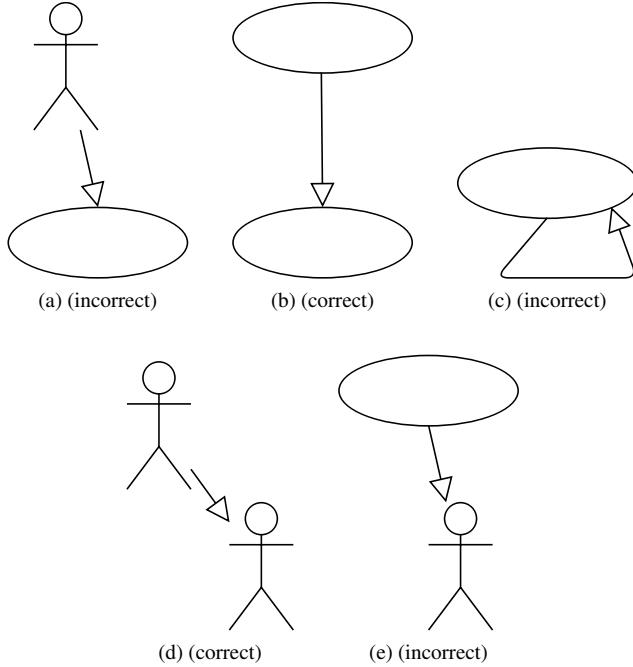


Figure 1: Simple instances of syntactically correct and incorrect use case diagrams.

interact with the widgets in the selected row to edit and/or delete it. The user can add new symbols or connectors by using the buttons below the table.

- A panel (on the right) showing a graphical preview of the symbol/connector selected in the table. It is possible to change the graphical representation of the symbol by using the button *Change*.
- A table (in the center) showing the information related to the selected symbol/connector. Each row specifies the attachment points and their constraints. It is possible to add new rows by using the buttons above the table;
- A textarea to specify the symbol/connector level constraints through C language-like expressions.

#### 4.1.1 Wizard

A new language can be defined by using a wizard interface. Through a sequence of three different views, the user chooses the name of the language, its symbols and connectors (see Figure 3). Symbols and connectors are chosen from a hierarchical repository and their definition already includes some attachment points having default types/constraints. The user can choose whether to keep the default values or modify them.

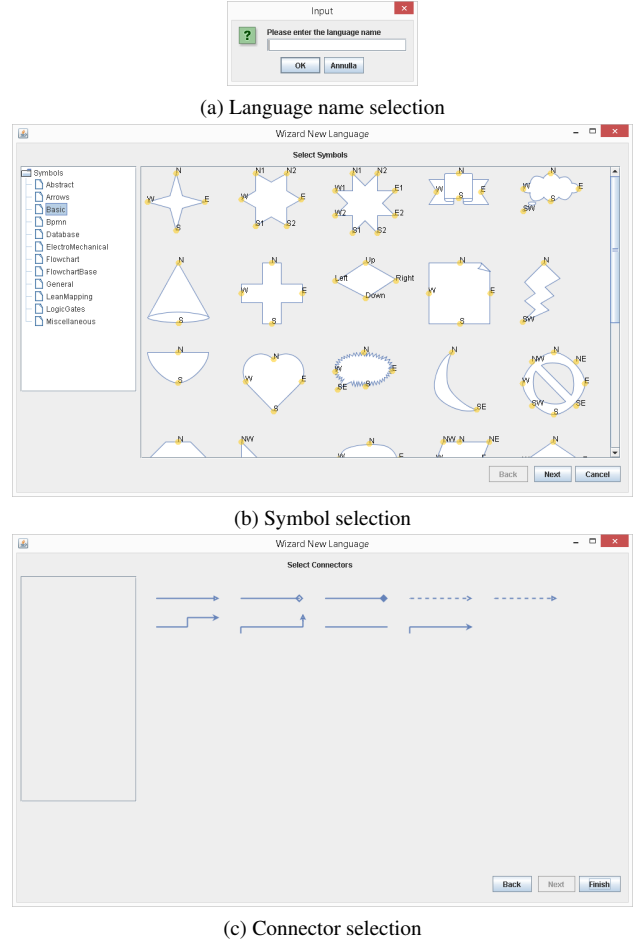


Figure 3: Wizard.

## 4.2 TiVE: the visual language environment

Once the language is defined, the diagrams can be composed by using the symbols and the connectors defined in its specification. This can be done through the graphical editor TiVE, which is a web application enabling diagram composition directly in the web browser and which is created by and can also be launched from the LoCoModeler.

Figure 4 shows the environment. The central component is the working area where the diagrams are composed. The symbols and connectors used for diagram composition are displayed in the sidebar, which contains only those elements included in the definition of the language. An element can be selected and dragged in the central working area.

The upper toolbar provides shortcuts to features such as zoom manipulation, changing fonts, checking diagram correctness, etc.

The correctness of a diagram can be checked at any point of the diagram composition. The diagram in Figure 5 represents an ER diagram with two entities interconnected by

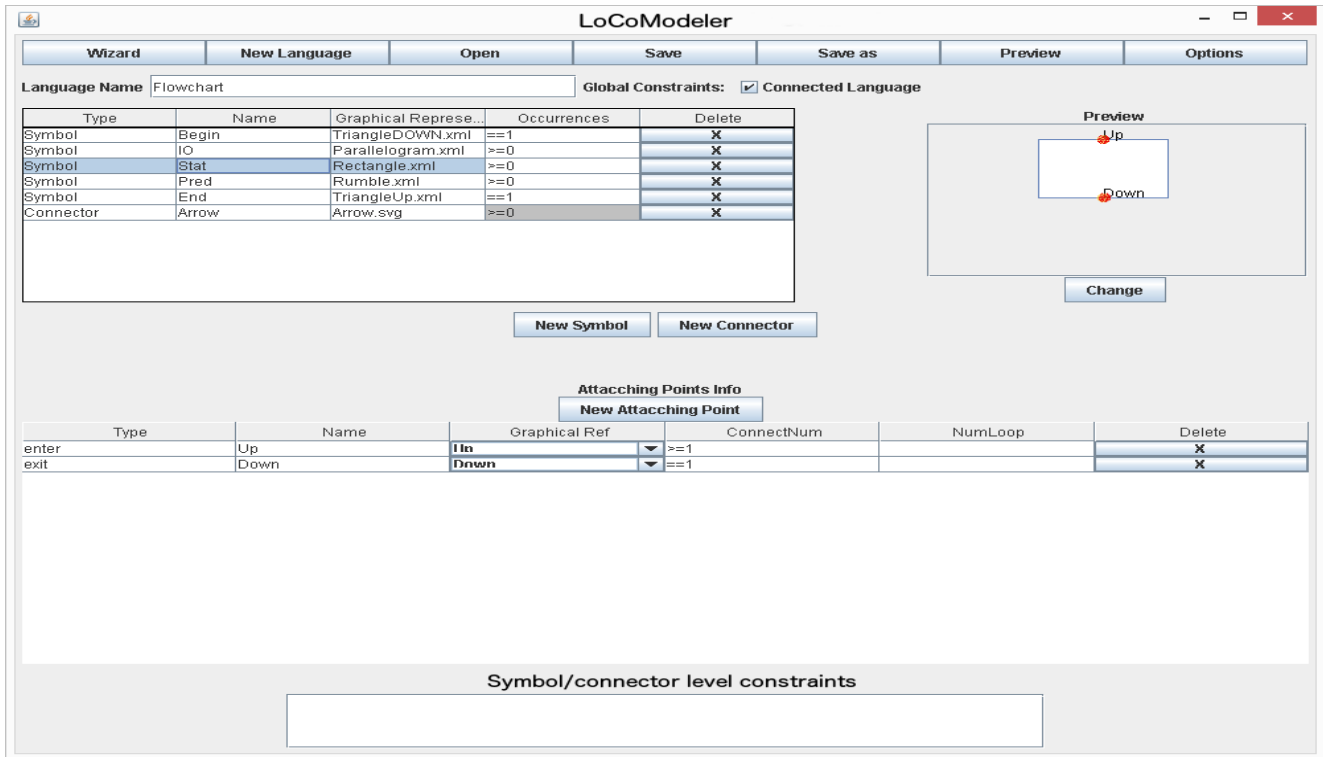


Figure 2: The Local Context-based Modeler.

a binary relation. The diagram is correct and, by launching the check, an alert reports the positive result of the verification. The diagram in Figure 6, instead, represents an incorrect ER diagram, as the relation (the diamond symbol) is connected to a single entity (the rectangle). This violates the local constraints defined for the relation symbol in the language. In fact, in this language, relations must be connected to two or three entities through the diamond's vertices.

### 4.3 Implementation

The LoCoModeler allows the user to produce the language specification in XML format. The specification is used during the removal of ambiguities and the recognition of symbols and connectors.

TiVE is based on Draw.io (<https://www.jgraph.com>), which is a free web application that allows users to create charts directly from the browser and integrates with Google Drive and Dropbox to store data. Draw.io is in turn based on the mxGraph library, which renders the diagram and stores its structure in the form of a graph, where symbols and connectors are its vertices. We modified the library to handle attaching points of symbols and connectors.

In a typical thin-client environment, mxGraph is divided into a client-side JavaScript library and a server side library in one of the two supported languages: .NET and Java. The

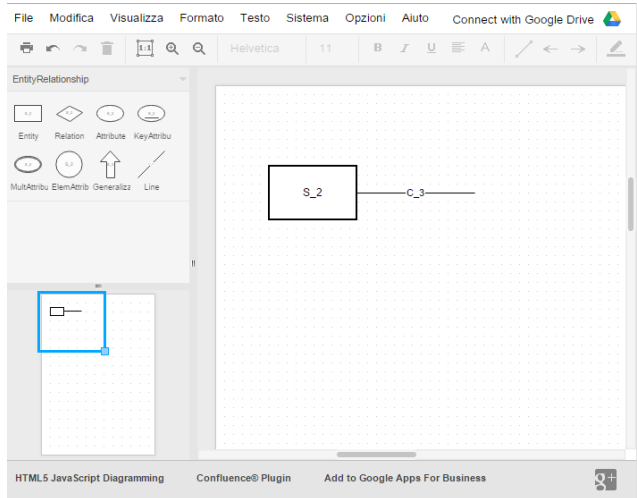


Figure 4: The TiVE home page.

JavaScript library is a part of a larger web application. The JavaScript code uses vector graphics to render the chart. The languages used are SVG (Scalable Vector Graphics) for standard browsers and VML (Vector Markup Language) for Internet Explorer.

An implementation of the tool can be downloaded at the address <http://weblab.di.unisa.it/locomotive>.

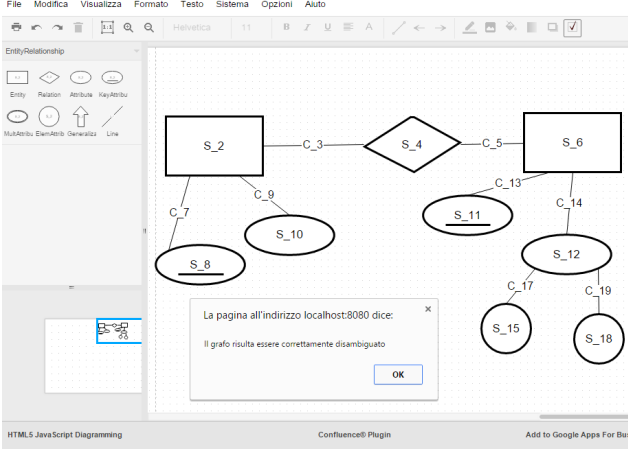


Figure 5: Successful diagram verification (no error found).

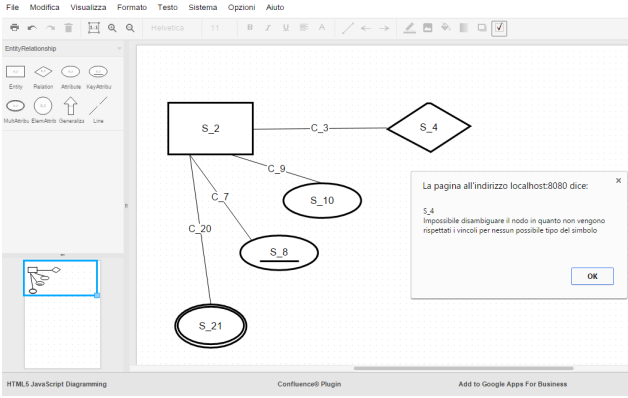


Figure 6: Failed diagram verification (one error found).

## 5 Evaluation

We ran a user study aimed at measuring users' capacity to define visual languages using our tool. Furthermore, we recorded the perceived usability of the system through a questionnaire.

### 5.1 Participants

Our participants were six male and four female Italian university students in computer science (six master students and four phd students), aged between 22 and 45 ( $M = 26.8$ ,  $SD = 6.9$ ), with no previous experience with the system.

Participants were asked to evaluate, with a 5-point Likert scale, their prior knowledge in programming, diagrams, compilers, formal languages, flowchart and other visual languages. The average and standard deviations of the responses are reported in Table 4.

Knowledge	Avg.	St. dev.
Programming	4.40	0.70
Diagrams	3.40	0.84
Compilers	2.90	1.10
Formal languages	3.40	1.07
Flowchart	3.40	0.97
Other visual languages	2.80	1.14

Table 4: Participants prior knowledge evaluation with a 5-point Likert scale.

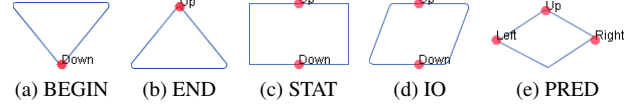


Figure 7: Flowchart symbols.

### 5.2 Apparatus

The experiment was executed on a *Dell Precision T5400* workstation equipped with an *Intel Xeon CPU* at 2.50 GHz running *Microsoft Windows 8.1* operating system, the *Java Run-Time Environment 7*, and the *Firefox* browser.

### 5.3 Procedure

Participants were asked, as a single task of the session, to define a visual language. In particular, they were asked to create a simplified version of flowcharts, as defined in [7], with the following features:

- The language only includes a small set of blocks (start, end, I/O, decision, processing - shown in Figure 7) and an arrow as a connector;
- The handling of text within blocks or arrows is not required;
- An arrow is always directed at the top of a block and comes out from its bottom;

Participants were asked to specify all the constraints necessary to ensure that only well formed flowcharts would pass the correctness check. Furthermore, they were required to carefully check they had defined the language correctly before submitting the task. The time limit for task completion was half an hour.

Before the experimental session, each participant had a brief tutorial phase where an operator (one of the authors) explained him/her the purpose and operation of the system and instructed him/her about the experimental procedure and the task. While showing the operation of the system, the operator also showed participants how to use our tool to define a simple visual language, in this case a simplified version of ER diagrams.

A post-test questionnaire in the form of System Usability Scale (SUS) [5] was administered to participants at the end of the experiment. SUS is composed of 10 statements to which participants assign a score indicating their strength of agreement in a 5-point scale. The final SUS score ranges from 0 to 100. Higher scores indicate better perceived usability. We also gathered some participants' freeform comments.

## 5.4 Results

Two participants out of ten completed the experiment defining the language perfectly, seven completed the experiment with minor inaccuracies in the language definition, while only one of them completed the experiment with major inaccuracies. Here, for minor inaccuracies, we mean small errors that allow user to compose at least one invalid diagram which however satisfied the user's language specification. Typical errors are inaccuracies in defining attachment points cardinality. The participant who committed major errors was unable to compose and correctly compile any diagram. The average task completion time was 25.5 minutes.

The responses given by participants to the statements in the SUS questionnaire are reported in Table 5. In particular the responses to statements 1, 3, 5, 7 and 9 show that participants appreciate the system, that they considered it simple to use and easy to learn even for non-experts of visual languages. Moreover the responses to questions 2, 4, 6, 8 and 10 show that participants did not feel they need support to use the system and did not found the system complex, intricate or inconsistent.

The scores of the questionnaire calculated on the responses of the participants range from 37.5 to 95, with an average value of 80.0, which value indicates a good level of satisfaction [1]. As it can be seen from the data in the table, only a single participant (the one who committed major errors) expressed a negative judgment on the tool.

In addition, participants provided some freeform suggestions for improving the system: most of the criticism was expressed on the editor for diagram composition tool, which was not felt to be very user-friendly. In particular, participants noticed that some basic operations for diagram composition, such as the insertion of connectors, are surprisingly uncomfortable. Furthermore, one participant pointed out that the editor is not well integrated with the VLDE.

## 6 Conclusions

In this paper we have presented a framework for the fast prototyping of visual languages exploiting their local context based specification. We have shown how to define a visual language by extending the local context with three new

features and have presented a simple interface for its implementation LoCoMoTiVE. Moreover, we have described a user study for evaluating the satisfaction and effectiveness of users when prototyping a visual language. At the moment, the user study has been limited to the simpler version of the local context methodology in order to provide us with a first feedback. Given the encouraging results, we are now planning to test the usability of LoCoMoTive with more complex applications.

The local context approach may then greatly help visual language designers to prototype their languages very easily. However, more studies are needed to investigate the computational borders of the approach. Our intention is not too push local context features more than needed, keeping simplicity as a priority. More complex language constructs should then be left to the following phases of the recognition process as it is the case for programming language compiler construction.

As a final goal, we are working on the integration of the local context approach in frameworks for the recognition of hand drawn sketches, as shown in [7].

## References

- [1] A. Bangor, P. T. Kortum, and J. T. Miller. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6):574–594, 2008.
- [2] R. Bardohl. Genged: a generic graphical editor for visual languages based on algebraic graph grammars. In *Visual Languages, 1998. Proceedings. 1998 IEEE Symposium on*, pages 48–55, Sep 1998.
- [3] R. Bardohl, M. Minas, G. Taentzer, and A. Schürr. Handbook of graph grammars and computing by graph transformation. chapter Application of Graph Transformation to Visual Languages, pages 105–180. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 1999.
- [4] P. Bottoni and G. Costagliola. On the definition of visual languages and their editors. In M. Hegarty, B. Meyer, and N. Narayanan, editors, *Diagrammatic Representation and Inference*, volume 2317 of *Lecture Notes in Computer Science*, pages 305–319. Springer Berlin Heidelberg, 2002.
- [5] J. Brooke. Sus: A quick and dirty usability scale. In P. W. Jordan, B. Weerdmeester, A. Thomas, and I. L. McLelland, editors, *Usability evaluation in industry*. Taylor and Francis, London, 1996.
- [6] S. S. Chok and K. Marriott. Automatic construction of intelligent diagram editors. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology, UIST '98*, pages 185–194, New York, NY, USA, 1998. ACM.
- [7] G. Costagliola, M. De Rosa, and V. Fuccella. Local context-based recognition of sketched diagrams. *Journal of Visual Languages & Computing*, 25(6):955–962, 2014. Distributed Multimedia Systems {DMS2014} Part I.



Question		U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	Avg. resp.	St. dev.
S1	I think that if I needed to define a visual language I would use this system	4	5	4	4	4	3	5	3	5	5	4.2	0.79
S2	I found the system unnecessary complex	1	1	1	2	4	2	2	4	1	2	2.0	1.15
S3	I found the system very easy to use	5	5	5	4	4	4	4	2	5	4	4.2	0.92
S4	I think I would need the support of a person who is already able to use the system	2	1	1	1	1	2	2	4	1	1	1.6	0.97
S5	I found the various system functions well integrated	4	5	4	3	3	4	4	3	5	3	3.8	0.79
S6	I found inconsistencies between the various system functions	2	1	1	2	1	1	1	2	1	2	1.4	0.52
S7	I think most people can easily learn to use the system	5	4	5	4	5	4	5	3	5	4	4.4	0.70
S8	I found the system very intricate to use	2	1	1	1	2	2	2	4	1	2	1.8	0.92
S9	I have gained much confidence about the system during use	4	4	4	5	5	4	3	2	4	5	4.0	0.94
S10	I needed to perform many tasks before being able to make the best use of the system	1	1	2	1	1	1	3	4	2	2	1.8	1.03
<b>Score</b>		<b>85</b>	<b>95</b>	<b>90</b>	<b>82.5</b>	<b>80</b>	<b>77.5</b>	<b>77.5</b>	<b>37.5</b>	<b>95</b>	<b>80</b>	<b>80</b>	<b>16.33</b>

Table 5: SUS questionnaire results (5-point Likert scale).

- [8] G. Costagliola, V. Deufemia, and G. Polese. A framework for modeling and implementing visual notations with applications to software engineering. *ACM Trans. Softw. Eng. Methodol.*, 13(4):431–487, Oct. 2004.
- [9] G. Costagliola, V. Deufemia, and G. Polese. Visual language implementation through standard compiler–compiler techniques. *Journal of Visual Languages & Computing*, 18(2):165 – 226, 2007.
- [10] G. Costagliola and G. Polese. Extended positional grammars. In *Proc. of VL '00*, pages 103–110, 2000.
- [11] J. de Lara and H. Vangheluwe. Atom3: A tool for multi-formalism and meta-modelling. In R.-D. Kutsche and H. Weber, editors, *Fundamental Approaches to Software Engineering*, volume 2306 of *Lecture Notes in Computer Science*, pages 174–188. Springer Berlin Heidelberg, 2002.
- [12] E. J. Golin. Parsing visual languages with picture layout grammars. *J. Vis. Lang. Comput.*, 2(4):371–393, Dec. 1991.
- [13] E. J. Golin and S. P. Reiss. The specification of visual language syntax. *J. Vis. Lang. Comput.*, 1(2):141–157, June 1990.
- [14] R. J. and A. Schurr. Defining and parsing visual languages with layered graph grammars. *Journal of Visual Languages & Computing*, 8.1:27–55, 1997.
- [15] U. Kastens and C. Schmidt. VI-eli: A generator for visual languages - system demonstration. *Electr. Notes Theor. Comput. Sci.*, 65(3):139–143, 2002.
- [16] N. Le Novère et al. The systems biology graphical notation. *Nature Biotechnology*, 27:735–741, 2009.
- [17] K. Marriott. Parsing visual languages with constraint multi-set grammars. In M. Hermenegildo and S. Swierstra, editors, *Programming Languages: Implementations, Logics and Programs*, volume 982 of *Lecture Notes in Computer Science*, pages 24–25. Springer Berlin Heidelberg, 1995.
- [18] K. Marriott and B. Meyer. On the classification of visual languages by grammar hierarchies. *Journal of Visual Languages & Computing*, 8(4):375 – 402, 1997.
- [19] M. Minas and G. Viehstaedt. Diagen: A generator for diagram editors providing direct manipulation and execution of diagrams. In *Proceedings of the 11th International IEEE Symposium on Visual Languages, VL '95*, pages 203–, Washington, DC, USA, 1995. IEEE Computer Society.
- [20] I. Plauska and R. Damaševičius. Design of visual language syntax for robot programming domain. *Information and Software Technologies Communications in Computer and Information Science*, 403:297–309, 2013.
- [21] J. Quinn et al. Synthetic biology open language visual (sbol visual), version 1.0.0, 2013. <http://sbolstandard.org/downloads/specification-sbol-visual/> [Online; accessed 4-June-2015].
- [22] J. Rekers and A. Schurr. A graph based framework for the implementation of visual environments. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 148–155, Sep 1996.
- [23] L. Weitzman and K. Wittenburg. Relational grammars for interactive design. In *Visual Languages, 1993., Proceedings 1993 IEEE Symposium on*, pages 4–11, Aug 1993.

# Incremental indexing of objects in pictorial databases

G. Castellano, A.M. Fanelli, M.A. Torsello

Computer Science Department

University of Bari A. Moro

Via Orabona, 4 - 70126 Bari, Italy

(giovanna.castellano, annamaria.fanelli, mariaalessandra.torsello)@uniba.it

## Abstract

*Object indexing is a challenging task that enables the retrieval of relevant images in pictorial databases. In this paper, we present an incremental indexing approach of picture objects based on clustering of object shapes. A semi-supervised fuzzy clustering algorithm is used to group similar objects into a number of clusters by exploiting a-priori knowledge expressed as a set of pre-labeled objects. Each cluster is represented by a prototype that is manually labeled and used to annotate objects. To capture eventual updates that may occur in the pictorial database, the previously discovered prototypes are added as pre-labeled objects to the current shape set before clustering. The proposed incremental approach is evaluated on a benchmark image dataset, which is divided into chunks to simulate the progressive availability of picture objects during time.*

## 1. Introduction

The extensive use of image digital capturing systems in several fields has generated massive amount of digital images that are typically collected in pictorial databases [1]. Most of the past projects on pictorial databases focus on content-based approaches searching images that are visually similar to the query image [2]. Such approaches do not have the capability of assigning textual descriptions automatically to pictures, i.e. they do not perform linguistic indexing.

Linguistic indexing is a difficult task due to the semantic gap problem, i.e. the lack of coincidence among the visual content of images represented by automatically extracted features and the human visual interpretation of the picture content typically expressed by high-level concepts [3]. Learning concepts from images and automatically translating the content of images to linguistic terms can bridge the semantic gap thus resulting in one of the most influential factors in successful image retrieval [4], [5] consequently

broadening the possible usages of pictorial databases.

Different machine-learning methods have been applied to learn associations between the low-level features and the linguistic concepts in a pictorial database [15]. In particular, learning techniques can be used to annotate objects clearly identifiable by linguistic cues. A common approach is to perform classification on the collection of picture objects [6], [14], so that visually similar objects are grouped into the same class and a textual label is associated to each class. Thus, each object is indexed by classifying it into one of the identified classes.

Classification of picture objects can be performed by means of supervised or unsupervised learning methods. Supervised techniques require a lot of training data, and providing these data is a very tedious and error-prone task, especially for large image database. Unsupervised learning techniques overcome these limitations but often they generate inconsistent classes including objects that, although having a similar shape, actually represent different linguistic cues. The presence of objects with ambiguous shapes motivates the use of semi-supervised clustering algorithms that can improve classification by using a combination of both labeled and unlabeled data. In [11] we proposed the use of a semi-supervised clustering algorithm called SSFCM (Semi-Supervised Fuzzy C-Means) to create object classes and prototypes useful for indexing images in a database. However, when new images are added to the database, this static indexing scheme requires rebuilding the prototypes starting from scratch by reprocessing the whole set of objects, i.e. it does not take advantage of the previously created prototypes.

To overcome this limitation, in this paper we propose the use of an incremental version of the SSFCM clustering algorithm, that we call Incremental SSFCM (ISSFCM). The ISSFCM applies SSFCM to chunks of picture objects that are periodically added to the database, thus providing an incremental scheme for picture object indexing.

The paper is organized as follows. Section 2 describes the proposed indexing scheme for pictorial object annota-

tion. In section 3 we provide some preliminary simulation results on a benchmark data set containing picture objects of different shapes. Finally, section 4 concludes the paper.

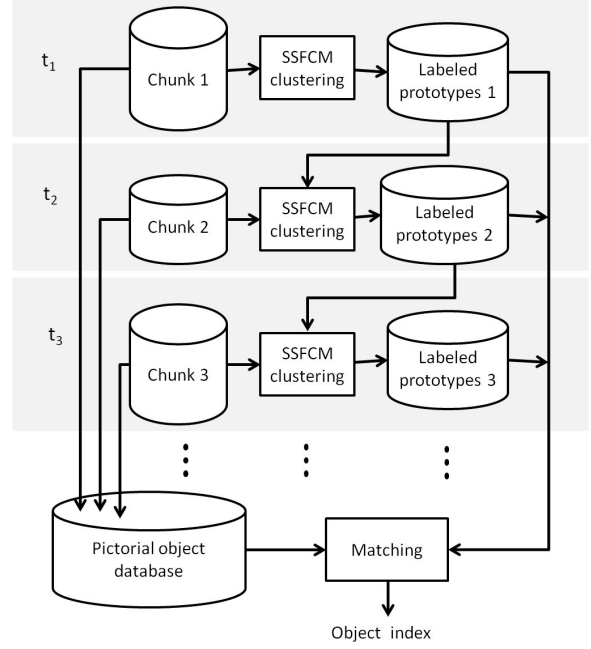
## 2. Incremental scheme for object indexing

We assume that a collection of pictorial objects is available. Each object is described by the contour of its shape. Different shape descriptors could be used to represent object shapes. In this work, each object shape is represented by means of Fourier descriptors that are well-recognized to provide robustness and invariance, obtaining good effectiveness in shape-based indexing and retrieval [8]. The shape of each pictorial object is described by means of  $M$  Fourier descriptors and denoted by a numerical vector  $\mathbf{x} = (x_1, x_2, \dots, x_M)$ .

The proposed scheme for incremental indexing of objects is based on the assumption that sets of object shapes belonging to different semantic classes are available during time and processed as chunks, that is, a chunk of  $N_1$  object shapes is available at time  $t_1$ , a chunk of  $N_2$  shapes is available at  $t_2$  and so on. We denote by  $X_t$  the chunk of picture objects available at time  $t$ . For a correct application of the proposed incremental scheme, all semantic classes should be represented in the early chunks. The chunks of objects are processed as they are added to the database, by applying incrementally the Semi-Supervised FCM (SSFCM) algorithm [11] described in section 2.1. The resulting scheme, called ISSFCM (Incremental SSFCM), is shown in fig. 1. It enables the update of previously derived prototypes when new shapes are continuously available over time. Each time a new chunk of shapes is available, previously created cluster prototypes are used as pre-labeled shapes for the new run of SSFCM. At the end of each SSFCM run, the derived labeled prototypes are used to index all available shapes accumulated in the pictorial database.

The overall scheme of the proposed incremental indexing approach is summarized in algorithm 1. Each time a chunk is available, it is clustered by SSFCM and the resulting clusters are represented by  $K$  prototypes that are manually annotated by textual labels (step 4-7). Then each object is added to the cluster corresponding to the best matching prototype and labeled with the related label (step 8-9). Matching is based on computing Euclidean distance between the Fourier descriptors of the object and the descriptors of prototypes. We chose the Euclidean distance since it is one of the most popular distances in literature that permits to obtain accurate results when matching shapes represented by Fourier descriptors with a low-cost and simple computation [16]. To take into account the evolution of the database, the prototypes discovered from one chunk are added as pre-labeled objects to the next chunk (step 10, step 4).

Precisely, when the first chunk of pictorial objects is



**Figure 1. The scheme of the incremental indexing approach**

available, the algorithm will cluster the chunk into  $K$  clusters and it will derive a set of  $K$  object prototypes that are manually labeled. When a second or later chunk of objects is available, it will be clustered with the labeled prototypes derived from the previous clustered chunks<sup>1</sup>.

Summarizing, our incremental indexing scheme generates a structure of clusters on the basis of chunks which capture the availability of new picture objects during time and reflect physical evolution of the database. The indexing mechanism is incremental in the sense that the cluster prototypes derived from one chunk are used not only for current indexing but also as a starting point for the clustering of successive chunks. The derived prototypes offer an intermediate indexing mechanism that enables automatic linguistic indexing of pictorial objects by requiring manual annotation of a very limited number of objects (namely the prototypes).

### 2.1. Clustering by SSFCM

The SSFCM algorithm works in the same manner as FCM (Fuzzy C-Means) [9], i.e. it iteratively derives  $K$  clusters by minimizing an objective function. Unlike FCM, that performs a completely unsupervised clustering, SSFCM performs a semi-supervised clustering, i.e. it uses

<sup>1</sup>How many chunks of history to use for clustering with a new chunk is predefined by the user.

---

**Algorithm 1** Incremental SSFCM (ISSFCM)

---

**Require:** Chunks of unlabeled objects  $X_1, X_2, \dots$ **Ensure:**  $P$ : set of labeled prototypes;  $X$ : set of annotated objects

```
1:  $H \leftarrow \emptyset$  /* Initialization of history */
2:  $t \leftarrow 1$  /* Initialization of time step */
3: while  $\exists$  non empty chunk  $X_t$  do
4:  $X_t \leftarrow X_t \cup H$  /* Add history to current chunk */
5: Cluster  $X_t$  using SSFCM
6: Derive the set  $P$  of prototypes
7: Annotate manually each prototype in  $P$ 
8: Annotate each object in  $\bigcup_{\tau=1}^t X_\tau$  using the best-
   matching prototype in  $P$ 
9: Update  $X$  with annotated objects
10: Update  $H$  with  $P$ 
11:  $t := t + 1$ 
12: end while
13: return  $P, X$ 
```

---

a set of pre-labeled data to improve clustering results. To embed partial supervision in the clustering process, the objective function of SSFCM includes a supervised learning component, as follows:

$$J = \sum_{k=1}^K \sum_{j=1}^{N_t} u_{jk}^m d_{jk}^2 + \alpha \sum_{k=1}^K \sum_{j=1}^{N_t} (u_{jk} - b_j f_{jk})^m d_{jk}^2 \quad (1)$$

where

$$b_j = \begin{cases} 1 & \text{if object } \mathbf{x}_j \text{ is pre-labeled} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$f_{jk}$  denotes the true membership value of the pre-labeled object  $\mathbf{x}_j$  to the cluster  $k$ ,  $d_{jk}$  represents the Euclidean distance between the object shape  $\mathbf{x}_j$  and the center of the  $k$ -th cluster,  $m$  is the fuzzification coefficient ( $m \geq 2$ ) and  $\alpha$  is a parameter that serves as a weight to balance the supervised and unsupervised components of the objective function. The higher the value of  $\alpha$ , the higher the impact coming from the supervised component is. The second term of  $J$  captures the difference between the true membership  $f_{jk}$  and the membership  $u_{jk}$  computed by the algorithm. The aim to be reached is that, for the pre-labeled objects, these values should coincide.

As described in [12], the problem of optimizing the objective function  $J$  is converted into the form of unconstrained minimization using the standard technique of Lagrange multipliers. By setting the fuzzification coefficient  $m$  equal to 2, the objective function is minimized by updating membership values  $u_{jk}$  according to:

$$u_{jk} = \frac{1}{1 + \alpha} \left[ \frac{1 + \alpha(1 - b_j \sum_{l=1}^K f_{lk})}{\sum_{l=1}^K d_{jk}^2 / d_{lk}^2} \right] + \alpha b_j f_{jk} \quad (3)$$

and the centers of clusters according to:

$$\mathbf{c}_k = \frac{\sum_{j=1}^{N_t} u_{jk}^m \mathbf{x}_j}{\sum_{j=1}^{N_t} u_{jk}^m} \quad (4)$$

The clustering process ends when the difference between the values of  $J$  in two consecutive iterations drops below a prefixed threshold or when the established maximum number of iterations is achieved.

Once the clustering process is completed, a prototype is identified for each cluster by selecting the object shape belonging with the highest membership to that cluster. Then, each prototype is manually associated to a label corresponding to a specific linguistic cue or semantic class.

Summarizing, the result of SSFCM applied to each chunk is a set of  $K$  labeled prototypes  $P = \{p_1, p_2, \dots, p_K\}$  that are used to index objects in the database. Namely, all objects belonging to cluster  $k$  are associated with the text label assigned to prototype  $p_k$ .

### 3. Experimental results

To assess the suitability of the proposed incremental indexing approach, we considered the MPEG-7 Core Experiment CE-Shape-1 data set [8] containing 1400 binary images of object shapes grouped into 70 different classes with each class including 20 samples. Fig. 2 shows a sample image for each class of the considered data set. In order to apply ISSFCM, all images were processed to extract boundaries of shapes and compute Fourier descriptors. Each object shape was represented by a vector of 32 Fourier coefficients (this number was set in our previous experiments on the same dataset).

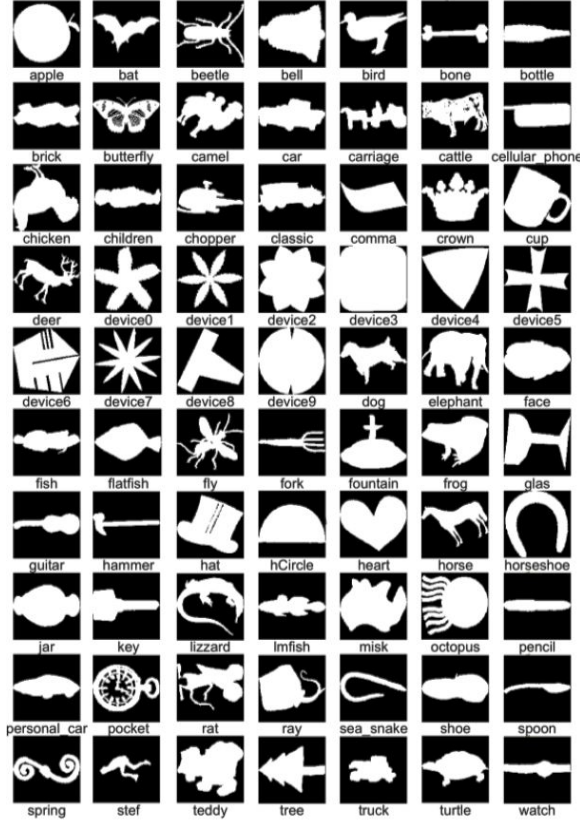
To evaluate the clustering results we used the average purity error, as in [10], defined as follows:

$$pur = 1 - \frac{1}{K} \times \sum_{k=1}^K \frac{|C_k^d|}{|C_k|}$$

where  $K$  denotes the number of clusters,  $|C_k^d|$  denotes the number of objects with the dominant class label in cluster  $k$  and  $|C_k|$  denotes the total number of objects in cluster  $k$ . Intuitively, the purity error measures the purity of the clusters with respect to the true cluster (class) labels that are known for the MPEG-7 dataset.

We performed a suite of experiments in order to analyze the behavior of ISSFCM when varying the percentage  $p$  of pre-labeled shapes ( $p = 20\%$  and  $p = 30\%$ ) and following two different pre-labeling schema:

- scheme A: we assume that each chunk contains a percentage  $p$  of pre-labeled shapes;

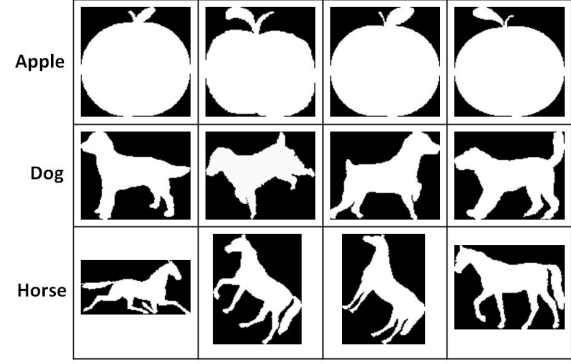


**Figure 2. Sample images from the MPEG-7 Core Experiment CE-Shape-1 data set**

- scheme B: we assume that only the first chunk contains a percentage  $p$  of pre-labeled shapes; in the next chunks the previously derived prototypes represent the pre-labeled shapes.

In all the experiments, the parameters of the ISSFCM algorithm were set as follows: the number of cluster  $K$  was set to the number of classes in the dataset (i.e.  $K = 70$ ), the size of a chunk was set to 280 shapes (hence 5 chunks were built from the whole dataset), the history was set to 1, meaning that only the prototypes extracted from the previous chunk were considered as pre-labeled shapes in the current chunk. Since SSFCM is not deterministic (due to the random initialization of the cluster centers) 10 different runs were performed and the average results are presented.

At the first time step, the SSFCM was applied to the union of the first two chunks in order to obtain more stable and significant initial prototypes to be exploited in the next steps of the incremental clustering process. In this way 4 different time steps were simulated. After clustering a chunk, 70 prototypes were derived and each prototype was manually annotated by a label descriptive of a semantic



**Figure 3. Prototypes derived in each time step for three semantic classes**

**Table 1. Average purity error values**

scheme	percentage of pre-labeled shapes	
	20%	30%
A	0.29	0.18
B	0.28	0.17

class. These prototypes were used to annotate all shapes included in the previous chunks on the basis of a top-matching score. To perform matching we computed the Euclidean distance between descriptors of each shape and descriptors of each prototype. Each shape was annotated with the label of the best-matching prototype. As an example, in fig. 3, we show the prototypes derived for three semantic classes at the end of each time step by applying ISSFCM with the 30% of pre-labeled shapes following scheme B.

The annotation results were evaluated by computing the average purity error. Table 1 reports the average values of the purity error obtained by varying the percentage of pre-labeled shapes and the pre-labeling scheme. It can be seen that, as expected, when the percentage of pre-labeled shapes increases, the quality of the obtained clusters improves. Regardless the pre-labeling percentage, the two pre-labeling schema provide comparable values of the purity error.

The effectiveness of the proposed incremental approach was evaluated by comparing the average purity error obtained in the last step of ISSFCM and the average purity error obtained by applying the SSFCM algorithm in a one-shot way (following the experimental setting described in [11]). To apply the SSFCM in one-shot way the data set was divided into a training set (composed of the shapes included in the first 4 chunks) and a test set (including the 280 remaining shapes). The training set was used to derive the shape prototypes whilst the test set was used to perform annotation by exploiting the derived prototypes. Figure 4a compares the average purity error obtained by applying ISS-

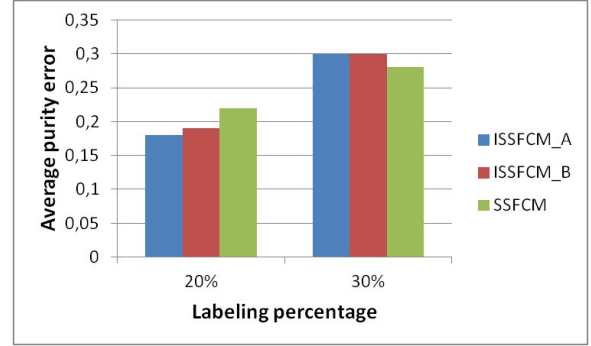
FCM (varying the pre-labeling scheme and the percentage of pre-labeled shapes) and the one-shot SSFCM. We observe that ISSFCM obtains results that are comparable to those obtained by the static one-shot SSFCM with the additional advantage to exploit and update the knowledge discovered in the previous time steps. Finally, we evaluated the annotation accuracy in terms of Precision and Recall and we compared the results obtained by applying the incremental SSFCM the static SSFCM. Figures 4b and 4c show the comparative values of precision and recall, respectively. It can be seen that the incremental indexing approach achieves better annotation accuracy with respect the static one-shot approach thus confirming the benefit of exploiting previously acquired knowledge whenever new picture objects have to be added to the pictorial database.

#### 4. Conclusions

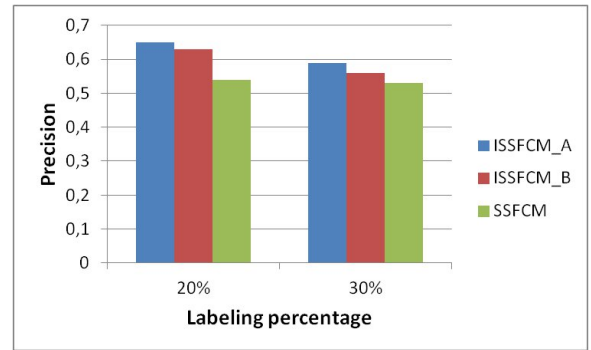
In this paper an incremental scheme for pictorial object indexing has been proposed. The approach exploits a semi-supervised fuzzy clustering algorithm to derive a set of prototypes representative of a number of semantic categories. The derived prototypes are manually annotated by attaching labels related to semantic categories. The use of shape prototypes, which represent an intermediate level of visual signatures, facilitates the annotation process, since only a reduced number of objects need to be manually annotated. Moreover, the use of prototypes simplifies the search process in a pictorial database by reducing time needed to retrieve similar shapes. Indeed, a query is matched only with shape prototypes, thus avoiding unnecessary comparisons with all objects in the database. Annotation results on the MPEG-7 benchmark dataset show that our incremental scheme obtains results which are very similar to those obtained by the one-shot approach with the additional advantage to exploit the previously discovered prototypes thus avoiding the reprocessing of the whole database. These preliminary results encourage the application of the proposed approach to real-world contexts requiring the indexing of evolving collections of pictorial objects.

#### References

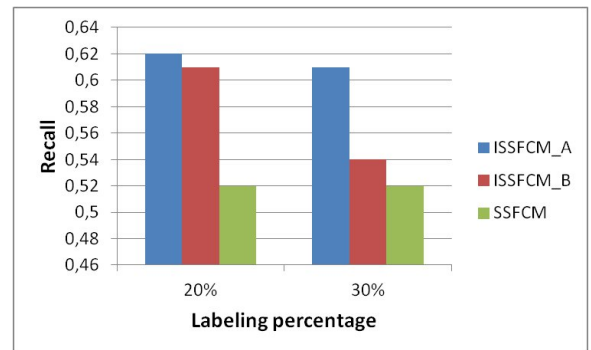
- [1] S.K. Chang and T.L. Kunii. Pictorial data-base systems, IEEE 23. G. Strang, Linear Algebra and Its Applications. Harcourt, Brace, and Computer 14, pp. 13-21, 1981.
- [2] Y. Rui, T. Huang, and S. Chang, Image retrieval: current techniques, promising directions and open issues, J. Visual Commun. Image R. 10(4):39-62, 1999.
- [3] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, Content-based image retrieval at the end



(a)



(b)



(c)

**Figure 4. Comparison between Incremental SSFCM and one-shot SSFCM**

- of the early years, IEEE Trans. Pattern Analysis and Machine Intelligence, 22:1349-1380, 2000.
- [4] W.I. Grosky, and R. Mehrotra. Index-based object recognition in pictorial data management. Computer Vision, Graphics, and Image Processing, 52(3):416-436, 1990.
  - [5] J. Li, and J.Z. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9):1075-1088, 2003.
  - [6] H. Nezamabadi-Pour, and S. Saryazdi. Object-based image indexing and retrieval in DCT domain using clustering techniques. In Proc. of World Academy of Science Engineering and Technology, pp. 207-210, 2005.
  - [7] S. Aghabozorgi, M.R. Saybani, and T.Y. Wah. Incremental clustering of time-series by fuzzy clustering. Journal of Information Science and Engineering 28(4):671-688, 2012.
  - [8] I. Bartolini, P. Ciaccia, and M. Patella. WARP: Accurate retrieval of shapes using phase of Fourier descriptors and Time warping distance. IEEE Trans. on Pattern Analysis and machine Intelligence, 27(1): 142-147, 2005.
  - [9] J.C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, 1981.
  - [10] F. Cao, M. Ester, W. Qian and A. Zhou. Density-based clustering over an evolving data stream with noise. In 2006 SIAM Conference on Data Mining, pp. 328-339, 2006.
  - [11] G. Castellano, A.M. Fanelli and M.A. Torsello. Shape annotation by semi-supervised fuzzy clustering. Information Sciences, 289(24):148-161, 2014.
  - [12] W. Pedrycz and J. Waletzky. Fuzzy clustering with partial supervision. IEEE Transactions on System Man and Cybernetics, 27(5): 787-795, 1997.
  - [13] S. Guha, A. Meyerson, N. Mishra, R. Motwani and L. O'Callaghan. Clustering data streams: Theory and practice, IEEE Trans. on Knowledge and Data Engineering, 15(3):515-528, 2003.
  - [14] D. Stan, and I.K. Sethi. Mapping low-level image features to semantic concepts. In Proc. of the SPIE, pp. 172-179, 2001.
  - [15] D. Zhang, M.M. Islam, and G. Lu, A review on automatic image annotation techniques, Pattern Recogn. 45(1):346-362, 2011.
  - [16] D. Zhang and G. Lu, Shape-based image retrieval using generic Fourier descriptor, Signal Processing: Image Communication, 17(10):825-848, 2002.



# RankFrag: A Machine Learning-Based Technique for Finding Corners in Hand-Drawn Digital Curves

Gennaro Costagliola\*, Mattia De Rosa\*, Vittorio Fortino†, Vittorio Fuccella\*

\*Dipartimento di Informatica, University of Salerno, Via Giovanni Paolo II, 84084 Fisciano (SA), Italy

†Unit of Systems Toxicology, Finnish Institute of Occupational Health (FIOH), Helsinki, Finland

{gencos, matderosa, vfuccella}@unisa.it, vittorio.fortino@ttl.fi

## Abstract

We describe RankFrag: a technique which uses machine learning to detect corner points in hand-drawn digital curves. RankFrag classifies the stroke points by iteratively extracting them from a list of corner candidates. The points extracted in the last iterations are said to have a higher rank and are more likely to be corners. The technique has been tested on three different datasets described in the literature. We observed that, considering both accuracy and efficiency, RankFrag performs better than other state-of-art techniques.

Keywords: corner finding, stroke segmentation, fragmentation, sketch recognition, machine learning, RankFrag

## 1 Introduction

The research on hand-drawn sketch recognition has had a recent boost due to the diffusion of devices (smartphones and tablets) equipped with touch screens. Sketched diagrams recognition raises a number of issues and challenges, including both low-level stroke processing and high-level diagram interpretation [11]. A low-level problem is the *segmentation* (also known as *fragmentation*) of input strokes. Its objective is the recognition of the graphical primitives (such as lines and arcs) composing the strokes. Stroke segmentation can be used for a variety of objectives, including symbol [16, 4] and full diagram [3] recognition.

Most approaches for segmentation use algorithms for finding corners, since these points represent the most noticeable discontinuity in the graphical strokes. Some other approaches [1] also find the so called *tangent vertices* (smooth points separating a straight line from a curve or parting two curves). Besides stroke segmentation, the identification of corners has other applications, including gesture recognition [9] and gestural text entry [6, 5].

A high accuracy and the possibility of being performed

in real time are crucial features for segmentation techniques. Tumen and Sezgin [26] also emphasize the importance of the adaptation to user preferences and drawing style and to the particular domain of application. Adaptation can be achieved by using machine learning-based techniques. Machine learning has also proven to improve accuracy. In fact, almost all of the most recent segmentation methods use some machine learning-based technique.

The technique presented here, which we call *RankFrag*, uses machine learning to decide if a candidate point is a corner. Our technique is inspired by previous work. In particular, the work that mostly influenced our research is that of Ouyang and Davis [20], which introduced a *cost* function expressing the likelihood that a candidate point is a corner. We adopt their cost function, but our corner finding procedure is different. The technique works by iteratively removing points from a list of candidate corners. At each iteration, the point minimizing the cost function is classified and, in the case it is not a corner, it is removed. As a point is removed from the list, it is assigned a *rank*, which is a progressively decreased integer value. Points with a higher rank (a lower integer value) are more likely to be corners. Another important characteristic of RankFrag is the use of a variable “region of support” for the calculation of some local features, which is the neighborhood of the point on which the features are calculated. Most of the features used for classification are taken from several previous works in the literature [23, 15, 27, 21, 20]. Four novel features are introduced.

We tested our technique on three different datasets previously introduced and already used in the literature to evaluate existing techniques. We compared the performance of RankFrag to other state-of-art techniques [28, 26].

Summarizing, this paper introduces and evaluates:

1. a novel iterative procedure for finding corners in digital curves;
2. the use of four previously untested features for corner classification.

The rest of the paper is organized as follows: the next section contains a brief survey on the main approaches for

sketch segmentation; in Section 3 we describe our technique; Section 4 presents the evaluation of the performance of our technique in comparison to those of existing techniques, while the results are reported in Section 5; lastly, some final remarks and a brief discussion on future work conclude the paper.

## 2 Related Work

According to a widely accepted classification [24], the methods for corner detection in digital curves can be divided in two categories: those that perform a classification of the points and those that compute a piecewise approximation of the curves.

The former methods evaluate some features on the points of the stroke, after they have possibly been resampled, e.g. at a uniform distance. Curvature and speed are the features that have been used first. In particular, the corners are identified by looking at maxima in the curvature function or at minima in the speed function. Lately, methods based on machine learning have begun to consider a broader range of features.

One of the first methods proposed in the literature is [24], which assesses the curvature through three different measures. The authors also propose an advanced method for the determination of the region of support for local features. One of the first methods based on the simple detection of speed minima is [14]. Given the inaccuracy of curvature and speed taken individually, it was decided to evaluate them both in combination: [22] uses a hybrid fit by combining the set of candidate vertices derived from curvature data with the candidate set from speed data.

A method introducing a feature different from curvature and speed is *ShortStraw* [27]. It uses the *straw* of a point, which is the segment connecting the endpoints of a window of points centered on the considered point. The method gave good results in detecting corners in polylines by selecting the points having a straw of length less than a certain threshold. Subsequently, the method has been extended by Xiong and LaViola [28] to work also on strokes containing curves.

One of the first methods to use machine learning for corner finding is the one described in [20]. It is used to segment the shapes in diagrams of chemistry. A very recent one is *ClassySeg* [13], which works with generic sets of strokes. The method firstly detects candidate segment windows containing curvature maxima and their neighboring points. Then, it uses a classifier trained on 17 different features computed for the points in each candidate window to decide if it contains a corner point.

The approaches for computing a piecewise approximation of digital curves try to fit lines and curves sequentially in a stroke; the dominant points then correspond to the in-

tersections of adjacent substrokes. The problem of finding the optimal subset of the  $n$  points of the stroke has an exponential complexity. Nevertheless, almost all the algorithms that implement this approach use dynamic programming to reduce the exponential runtime complexity to  $O(n^2)$ . The first work [2] dates back to 1961. This algorithm fixes the number of segments and finds the solution minimizing the error. An algorithm proposed later [8] fixes the error and minimizes the number of segments. The algorithms also differ for the norm they use to measure the approximation error. A recent method, called DPFRag [26] learns primitive-level models from data, in order to adapt fragmentation to specific datasets and to user preferences and sketching style.

Lastly, there are hybrid methods, which use both the approaches mentioned above. *SpeedSeg* [12] and *TCVD* [1] are examples of such methods. *TCVD* is also able to find both the corners and the points where there is a significant change in curvature (referred to as “tangent vertices” in [1]). In order to detect corners, the former method mainly relies on pen speed while the latter uses a curvature measure. Tangent vertices are found through piecewise approximation by both methods.

## 3 The Technique

Our technique segments an input stroke in primitives by breaking it in the points regarded as corners. As a preliminary step, a *Gaussian smoothing* [10] is executed on the raw points in order to reduce the resampled stroke noise. Then, the stroke is processed by resampling its points to obtain an equally spaced ordered sequence of points  $P = (p_1, p_2, \dots, p_n)$ , where  $n$  varies depending on a fixed space interval and on the length of the stroke.

In order to identify the corners, the following three steps are then executed:

1. Initialization;
2. Pruning;
3. Point classification.

The initialization step creates a set  $D$  containing  $n$  pairs  $(i, c)$ , for  $i = 1 \dots n$  where  $c$  is the (*initial*) *cost* of  $p_i$  and is calculated through Eq. 1 derived, through some simplification steps, from the *cost* function defined in [20].

$$Icost(p_i) = \begin{cases} [dist(p_i; p_{i-1}, p_{i+1})]^2 & \text{if } i \in \{2, \dots, n-1\} \\ +\infty & \text{if } i = 1 \text{ or } i = n \end{cases} \quad (1)$$

In the above equation, the term  $dist(p_i; p_{i-1}, p_{i+1})$  indicates the minimum distance between  $p_i$  and the line segment formed by  $(p_{i-1}, p_{i+1})$ . Since  $p_1$  and  $p_n$  do not have a preceding and successive point, respectively, they are treated as special cases and given the highest cost.

The pruning step iteratively removes  $n-u$  elements from  $D$  in order to make the technique more efficient. The value  $u$  is the number of candidate corners not pruned in this step and depends on the complexity of the strokes in the target dataset. Its value has no effect on the accuracy of the method, provided that it is conservatively chosen so that no corner is eliminated in the pruning step. However, too high a value for this parameter may affect its efficiency.

At each iteration, the element  $m$  with the lowest cost is removed from  $D$  and the costs of the closest preceding points  $p_{pre}$  in  $P$  and the closest successive point  $p_{suc}$  in  $P$  of  $p_m$ , with  $pre$  and  $suc$  occurring in the set  $\{i : (i, c) \in D\}$ , are updated through Eq. 2 derived from the *cost* function defined in [20].

$$Cost(p_i) = \begin{cases} \sqrt{mse(S; p_{ipre}, p_{isuc})} \times dist(p_i; p_{ipre}, p_{isuc}) & \text{if } i \in \{2, \dots, n-1\} \\ +\infty & \text{if } i = 1 \text{ or } i = n \end{cases} \quad (2)$$

In the above equation,

- the points  $p_{ipre}$  and  $p_{isuc}$  are, respectively, the closest preceding and successive points of  $p_i$  in  $P$ , with  $ipre$  and  $isuc$  occurring in the set  $\{i : (i, c) \in D\}$ ;
- $S = \{p_{ipre}, \dots, p_{isuc}\}$  is the subset of points between  $p_{ipre}$  and  $p_{isuc}$  in the resampled stroke  $P$ ;
- $mse(S; p_{ipre}, p_{isuc})$  is the mean squared error between the set  $S$  and the line segment formed by  $(p_{ipre}, p_{isuc})$ ;
- the function *dist* is defined as for Eq. 1.

The point classification step returns the list of points recognized as corners by further removing from  $D$  all the pairs with indices of the points that are not recognized as corners. This is achieved by the following steps:

1. find the current element in  $D$  with minimum cost (if  $D$  contains only pairs with indices 1 and  $n$ , return an empty list);
2. calculate the features of the point corresponding to the current element and invoke the binary classifier, previously trained with data.
  - if the classifier returns false, delete the element from  $D$ , make the necessary updates and go to 1.
  - if the classifier returns true, proceed to consider as current the next element in  $D$  in ascending cost order. If the corresponding point is one of the endpoints of the stroke, return the list of points corresponding to the remaining elements in  $D$  (except for 1 and  $|P|$ ), otherwise go to 2.

In Fig. 1, the function DETECTCORNERS() shows the pseudocode for the initialization, pruning and point classification steps. In the pseudocode,  $D$  is the above described set with the following functions:

- INIT( $L$ ) initialize  $D$  with all the  $(i, c)$  pairs contained in  $L$ ;
- FINDMINC() returns the element of  $D$  with the lowest cost;
- PREVIOUSI( $i$ ) returns  $j$  such that  $(j, c')$  is the closest preceding element of  $(i, c)$  in  $D$ , i.e.,  $j = \max\{k \mid (k, c) \in D \text{ and } k < i\}$ ;
- SUCCESSIVEI( $i$ ) returns  $j$  such that  $(j, c')$  is the closest successive element of  $(i, c)$  in  $D$ , i.e.,  $j = \min\{k \mid (k, c) \in D \text{ and } k > i\}$ ;
- SUCCESSIVEC( $i$ ) returns the successive element of  $(i, c)$  in  $D$  with respect to the ascending cost order;
- REMOVE( $i$ ) removes  $(i, c)$  from  $D$ ;
- UPDATECOST( $i, c$ ) updates the cost  $c'$  to  $c$  for  $(i, c')$  in  $D$ .

DETECTCORNERS() calls a CLASSIFIER( $i, P, D$ ) function that computes the features (described in Section 3.2) of the point  $P[i]$ , and then uses them to determine if  $P[i]$  is a corner by using a binary classifier previously trained with data (described in Section 3.3).

### 3.1 Complexity

The complexity of the function DETECTCORNERS() in the previous section depends on the implementation of the data structure  $D$ . We will base our calculation by implementing  $D$  with an array and a pointer: the  $i$ th element of the array refers to the node that contains the pair  $(i, c)$  (or *nil* if the node does not exist) while the pointer refers to the node with the minimum  $c$ . Each node has 3 pointers: one that points to the successive node in ascending  $c$  order, one that points to the successive node in ascending  $i$  order and one that points to the previous node in ascending  $i$  order. Based on this implementation, the FINDMINC(), PREVIOUSI(), SUCCESSIVEI(), SUCCESSIVEC() and REMOVE() functions are all executed in constant time, while the UPDATECOST() function is  $O(|D|)$  (where  $|D|$  is the number of nodes referred in  $D$ ) and the INIT( $L$ ) function is  $O(|L| \log |L|)$  (by using an efficient sorting algorithm). In the following we will show that the DETECTCORNERS() complexity is  $O(n^2)$ , where  $n = |P|$ .

It is trivial to see that: the complexity of the ICOST() function is  $O(1)$ ; the complexity of COST() is  $O(n)$  in the

**Input:** an array  $P$  of equally spaced points that approximate a stroke, a number  $u$  of not-to-be-pruned points, and the `CLASSIFIER()` function.

**Output:** a list of detected corners.

```

1: function DETECTCORNERS( $P, u, \text{CLASSIFIER}$ )
2:   # initialization
3:   for  $i = 1$  to  $|P|$  do
4:      $c \leftarrow \text{ICOST}(i, P)$            # computes Eq. 1
5:     add  $(i, c)$  to  $\text{TempList}$ 
6:   end for
7:    $D.\text{INIT}(\text{TempList})$ 

8:   # pruning
9:   while  $|D| > u$  do
10:     $(i_{\min}, c) \leftarrow D.\text{FINDMINC}()$ 
11:     $\text{REMOVEANDUPDATE}(i_{\min}, P, D)$ 
12:  end while

13:  # point classification
14:  while  $|D| > 2$  do
15:     $(i_{\text{cur}}, c) \leftarrow D.\text{FINDMINC}()$ 
16:    loop
17:       $\text{isCorner} \leftarrow \text{CLASSIFIER}(i_{\text{cur}}, P, D)$ 
18:      if  $\text{isCorner}$  then
19:         $(i_{\text{cur}}, c) \leftarrow D.\text{SUCCESSIVEC}(i_{\text{cur}})$ 
20:        if  $i_{\text{cur}} \in \{1, |P|\}$  then
21:          for each  $(i, c)$  in  $D$  such that
22:             $(i \neq 1 \wedge i \neq |P|)$ 
23:            add  $P[i]$  to  $\text{CornerList}$ 
24:          return  $\text{CornerList}$ 
25:        end if
26:      else
27:         $\text{REMOVEANDUPDATE}(i_{\text{cur}}, P, D)$ 
28:        break loop
29:      end if
30:    end loop
31:  end while
32:  return  $\emptyset$ 
33: end function

34: procedure REMOVEANDUPDATE( $i, P, D$ )
35:   $i_{\text{pre}} \leftarrow D.\text{PREVIOUSI}(i)$ 
36:   $i_{\text{suc}} \leftarrow D.\text{SUCCESSIVEI}(i)$ 
37:   $D.\text{REMOVE}(i)$ 
38:
39:   $c \leftarrow \text{COST}(i_{\text{pre}}, P, D)$            # computes Eq. 2
40:   $D.\text{UPDATECOST}(i_{\text{pre}}, c)$ 
41:
42:   $c \leftarrow \text{COST}(i_{\text{suc}}, P, D)$ 
43:   $D.\text{UPDATECOST}(i_{\text{suc}}, c)$ 
44: end procedure

```

Figure 1: The implementation of the initialization, pruning and corner classification steps.

worst case and, consequently, the complexity of `REMOVEANDUPDATE()` is  $O(n)$ ; and the complexity of `CLASSIFIER()` is  $O(n)$  since some features need  $O(n)$  time in the worst case to be calculated.

The complexity of each of the three steps is then:

1. Initialization: `ICOST()` is called  $n$  times and `D.INIT()` one time, consequently the complexity of the initialization step is  $O(n \log n)$ .
2. Pruning: `D.FINDMINC()` and `REMOVEANDUPDATE()` are called  $n - u$  times each, consequently the complexity of this step is  $O(n(n - u))$ .
3. Point classification: the *while* loop (in line 14) will be executed at most  $k = |D| - 2 \leq u - 2$  times. In the loop (in line 16), `CLASSIFIER()` will be called at most  $k$  times, `D.SUCCESSIVEC()` at most  $k - 1$  times, and `REMOVEANDUPDATE()` at most once. Thus, in this step, they will be called less than or equal to  $k^2$ ,  $k^2$  and  $k$  times, respectively.

The complexity of the `CLASSIFIER()` calls can be calculated by considering that for each point, if none of its features changes, the result of `CLASSIFIER()` can be retrieved in  $O(1)$  by caching its previous output. Since the execution of the `REMOVEANDUPDATE()` function involves the changing of the features of two points, `CLASSIFIER()` will be executed at most  $3k$  times in  $O(n)$  (for a total of  $O(k \times n)$ ) and the remaining times in  $O(1)$  (for a total of  $O(k^2)$ ), giving a complexity of  $O(k \times n)$ .

Furthermore, the complexity of the `D.SUCCESSIVEC()` calls is  $O(k^2)$ , while the complexity of the `REMOVEANDUPDATE()` calls is  $O(k \times n)$ .

Thus, since  $k < n$ , the point classification step is in the worst case  $O(k \times n)$ , or rather  $O(n \times u)$ .

It is worth noting that the final  $O(n^2)$  complexity does not improve even if a better implementation of  $D$  providing an  $O(\log |D|)$  `UPDATECOST()` function is used.

### 3.2 Features

Most of the features used in our classification are derived from previous research in the field. In particular, we have three different classes of features:

- *Stroke features*: features calculated on the whole stroke;
- *Point features*: local features calculated on the point. These features are calculated using a fixed region of support and their values remain stable throughout the procedure;

- *Rank-related features*: dynamically calculated local features. The region of support for the calculation of these features is the set of points from the predecessor  $p_{pre}$  and the successor  $p_{suc}$  of the current point in the candidate list. Their value can vary during the execution of the *Point classification* step.

Some features are parametric. In particular, they can adopt two different types of parameters:

- An integer parameter  $w$ , defining the width of the (fixed) region of support used to calculate point features;
- A boolean parameter *norm*, indicating whether a normalization is applied in the calculation of the feature.

### 3.2.1 Stroke Features

The features calculated on the whole stroke can be useful to the classifier, since a characteristic of the stroke can interact in some way with a local feature. For instance, the length of a stroke may be correlated to the number of corners in it: it is likely that a long stroke has more angles than a short stroke. We derived two stroke features from [20]: the length of the stroke and the diagonal length of its bounding box. These features are called *Length* and *Diagonal*, respectively. In Figure 2a the bounding box (light gray) and the diagonal (dark gray) of a hand drawn diamond (black) are shown. Furthermore, we added a feature telling how much the stroke resembles an ellipse (or a circle), called *EllipseFit*. The use of this feature prevents that corners are accidentally inserted in strokes resembling circles or ellipses. It is calculated by measuring the average Euclidean distance of the points of the stroke to an ideal ellipse, normalized by the length of the stroke. Figure 2b shows the *EllipseFit* calculation for a hand-drawn diamond. In particular, the figure shows the segments (dark gray) connecting the diamond (black) and the ellipse (light gray), of which we calculate the average measure.

### 3.2.2 Point Features

The *point features* are local characteristics of the points. The speed of the pointer and the curvature of the stroke at a point have been regarded as very important features from the earliest research in corner finding. Here, the speed at  $p_i$  is calculated as suggested in [23], i.e.,  $s(p_i) = \|p_{i+1}, p_{i-1}\| / (t_{i+1} - t_{i-1})$ , where  $t_i$  represents the timestamp of the  $i$ -th point. We also have a version of the speed feature where a min-max normalization is applied in order to have as a result a real value between 0 and 1; the *Curvature* feature used here is calculated as suggested in [15].

A feature that has proven useful in previous research is the *straw*, proposed in [27]. The straw at the point  $p_i$  is the length of the segment connecting the endpoints of a window of points centered on  $p_i$ . Thus we define  $Straw(p_i, w) = \|p_{i+w}, p_{i-w}\|$ , where  $w$  is the parameter defining the width of the window. An example of straw is shown in dark gray in Figure 2d.

A simple feature to evaluate if a point is a corner, is the magnitude of the angle formed by the segments  $(p_{i-w}, p_i)$  and  $(p_i, p_{i+w})$ , defined here as  $Angle(p_i, w)$ . An example is shown in Figure 2e. A useful feature to distinguish the curves from the corners is what we call *AlphaBeta*, derived from [28]. Here we use as a feature the difference between *alpha* and *beta*, the magnitudes of two angles in  $p_i$  using different segment lengths, one three times the other:  $AlphaBeta(p_i, w) = Angle(p_i, 3w) - Angle(p_i, w)$ . An example of the two angles is shown in Figure 2f.

Lastly, in this research we introduce two point features that, as far as we know, have never been tested so far for corner detection. One feature is the position of the point within the stroke. Its use tends to prevent that corners are inserted in uncommon positions of the stroke. The position is calculated as the ratio between the length of the stroke from  $p_0$  to  $p_i$  and the total length of the stroke. We call this feature *Position*( $p_i$ ). The other feature is the difference of two areas: the former is the one of the polygon delimited by the points  $(p_{i-w}, \dots, p_i, \dots, p_{i+w})$  and the latter is the one of the triangle  $(p_{i-w}, p_i, p_{i+w})$ . The rationale for this feature is that its value will be positive for a curve, approximately 0 for an angle and even negative for a cusp. We call it *DeltaAreas*( $p_i, w$ ). Figure 2g shows an example that highlights the difference between the two areas.

### 3.2.3 Rank-Related Features

The *rank-related features* are local characteristics of the points. The difference with the *point features* is that their region of support varies according to the rank of the point: the considered neighborhood is between the closest preceding and successive points of  $p_i$ , which we have called  $p_{ipre}$  and  $p_{isuc}$ , respectively. The *Cost* function defined in Equation (2) is an example of feature from this class. It tends to assume higher values at the corners. A distinguishing feature of our approach, strictly related to the *Cost*, is the *Rank*. We define the *Rank* of a point  $p = P[i]$  with respect to  $D$ , as the size of  $D$  resulting from the removal of  $(i, c)$  from  $D$ . As already explained, this feature is a good indicator of whether a point is a corner and it is useful to associate it to the cost function, to improve classification.

A simple feature derived from [20] is *MinDistance*, representing the minimum of the two distances  $\|p_{ipre}, p_i\|$  and  $\|p_i, p_{isuc}\|$ , respectively. We also used a normalized version, obtained by dividing the minimum by  $\|p_{ipre}, p_{isuc}\|$ .

Feature	Class	Parameters	Ref.
$Length(S)$	Stroke	/	[20]
$Diagonal(S)$	Stroke	/	[20]
$EllipseFit(S)$	Stroke	/	
$Speed(p, norm)$	Point	$norm = T, F$	[23]
$Curvature(p)$	Point	/	[15]
$Straw(p, w)$	Point	$w = 4$	[27]
$Angle(p, w)$	Point	$w = 1, 2$	[28]
$AlphaBeta(p, w)$	Point	$w = 3, 4, 6, 15$	[28]
$Position(p)$	Point	/	
$DeltaAreas(p, w)$	Point	$w = 11$	
$Rank(p)$	Rank-Related	/	
$Cost(p)$	Rank-Related	/	[20]
$MinDistance(p, norm)$	Rank-Related	$norm = T, F$	[20]
$PolyFit(p)$	Rank-Related	/	[21]
$CurveFit(p)$	Rank-Related	/	[21]

Table 1: The features used in our classifier. Features without a reference are defined for the first time in this paper.

As in previous research, we try to fit parts of the stroke with beautified geometric primitives. The following two features are similar to the ones defined in [21]:  $PolyFit(p_i)$  fits the substroke  $(p_{ipre}, \dots, p_i, \dots, p_{isuc})$  through the polyline  $(p_{ipre}, p_i, p_{isuc})$ , while  $CurveFit(p_i)$  uses a bezier curve to approximate the points. The return value is the average point-to-point euclidean distance normalized by the length of the stroke. Examples of the two aforementioned features are shown in Figures 2i and 2c, respectively.

Table 1 summarizes the set of features used by RankFrag in the CLASSIFIER function. The table reports the name of the feature, its class, the values of the parameters (if present) with which it is instantiated and the reference paper from which we derived it. The presence of more than one parameter value means that some features are used multiple times, instantiated with different parameter values. The set of features has been chosen by performing a two-step feature selection method. In the first step, bootstrapping along with RF algorithm was used to measure the importance of all the features and produce stable feature importance (or rank) scores. Then, all the features were grouped into clusters using correlation, and those with the highest ranking score from each group were chosen to form the set of relevant and non-redundant features.

### 3.3 Classification method

The binary classifier used by RankFrag in the CLASSIFIER function to classify corner points is based on *Random Forests* (RF) [17]. Random Forests are an ensemble machine learning technique that builds forests of classification trees. Each tree is grown on a bootstrap sample of the data, and the feature at each tree node is selected from a random subset of all features. The final classification is determined by using a voting system that aggregates the classification results from all the trees in the forest. There are many ad-

vantages of RF that make their use an ideal approach for our classification problem: they run efficiently on large data sets; they can handle many different input features without feature deletion; they are quite robust to overfitting and have a good predictive performance even when most predictive features are noisy.

### 3.4 Implementation

RankFrag was implemented as a Java application. The classifier was implemented in *R* language, using the *randomForest* package [18]. The call to the classifier from the main program is performed through the *Java/R Interface* (JRI), which enables the execution of *R* commands inside Java applications.

## 4 Evaluation

We evaluated RankFrag on three different datasets already used in the literature to evaluate previous techniques. We repeated 30 times a 5-fold cross validation on all of the datasets. For all datasets, the strokes were resampled at a distance of three pixels, while a value of  $u = 30$  was used as a parameter for pruning. Since there is no single metric that determines the quality of a corner finder, we calculated the performance of our technique using the various metrics already described in the literature. The results for some metrics were averaged in the cross validation and were summed for others.

The hosting system used for the evaluation was a laptop equipped with an *Intel<sup>TM</sup> Core<sup>TM</sup> i7-2630QM* CPU at 2.0 GHz running *Ubuntu 12.10* operating system and the *OpenJDK 7*.

### 4.1 Model validation

Here we describe the process of assessing the prediction ability of the RF-based classifiers. The accuracy metrics were calculated by repeating 30 times the following procedure individually for each dataset and taking the averages:

1. the data set  $DS$  is randomly partitioned into 5 parts  $DS_1, \dots, DS_5$  with an equal number of strokes (or nearly so, if the number of strokes is not divisible by 5);
2. for  $i = 1 \dots 5$ :  $DSt_i = DS \setminus DS_i$  is used as a training set, and  $DS_i$  is used as a test set.
  - RankFrag is executed on  $DSt_i$  in order to produce the training data table. In  $DS$ , the correct corners had been previously marked manually. For each point extracted from the candidate list the input feature vector is calculated, while the



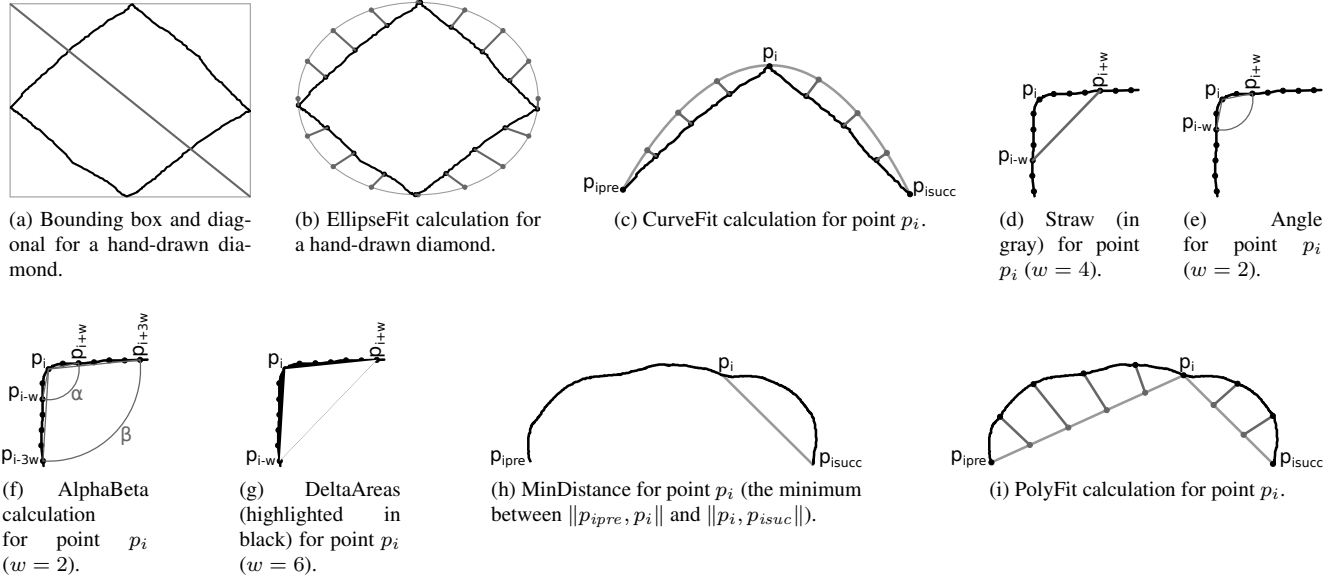


Figure 2: Examples for the features used in our classifier.

output parameter is given by the boolean value indicating whether the point is marked or not as a corner. The training table contains both the input and output parameters;

- a random forest is trained using the table;
- RankFrag is executed on  $DS_i$ , using the trained random forest as a binary classifier;
- In order to generate the accuracy metrics, the corners found by the last run of RankFrag are compared with the manually marked ones. A corner found by RankFrag is considered to be correct if it is within a certain distance from a marked corner.

3. In order to get aggregate accuracy metrics, for each of them the average/sum (depending on the type of the metric) of the values obtained in the previous step is calculated.

## 4.2 Accuracy Metrics

A corner finding technique is mainly evaluated from the points of view of accuracy and efficiency. There are different metrics to evaluate the accuracy of a corner finding technique. We use the following, already described in the literature [27, 13]:

- **False positives and false negatives.** The number of points incorrectly classified as corners and the number of corner points not found, respectively;

- **Precision.** The number of correct corners found divided by the sum of the number of correct corners and false positives: 
$$precision = \frac{correct\ corners}{correct\ corners + false\ positives};$$
- **Recall.** The number of correct corners found divided by the sum of the number of correct corners and false negatives: 
$$recall = \frac{correct\ corners}{correct\ corners + false\ negatives}.$$
 This value is also called **Correct corners accuracy**;
- **All-or-nothing accuracy.** The number of correctly segmented strokes divided by the total number of strokes;

The presence of the angle is determined by human perception. Obviously, different operators can perform different annotations on a dataset. The task of judging whether a corner is correctly found should also be done by a human operator. In our case, the human judgment is unfeasible due to the very high number of tests. Thus, we just checked whether the found corner was at a reasonable distance from the marked corner. In particular, we adopted as a tolerance the fixed distance of 20 pixels already used in literature for tests on the same datasets [13].

## 4.3 Datasets

Two of the three datasets used in our evaluation, the *Sezgin-Tumen COAD Database* and *NicIcon* datasets, are associated to a specific domain, while the *IStraw* dataset is not associated to any domain, but was produced for benchmarking purposes by Xiong and LaViola [28]. Some fea-

Dataset	No. of classes	No. of symbols	No. of strokes	No. of drawers	Source
COAD	20	400	1507	8	[25]
NicIcon	14	400	1204	32	[19]
ISraw	10	400	400	10	[28]

Table 2: Features of the three data sets.

tures of the three datasets are summarized in Table 2. The table reports, for each of them, the number of different classes, the total number of symbols and strokes, the number of drawers and a reference to the source document introducing it.

The symbols in the *Sezgin-Tumen COAD Database* (called only COAD, for brevity, in the sequel) dataset are a subset of those used in the domain of *Military Course of Action Diagrams* [7], which are used to depict battle scenarios. A set of 620 symbols was firstly introduced by Tirkaz et al. [25] to measure the performance of a multi-stroke symbol recognizer. Here we use a subset of 400 symbols annotated by Tumen and Sezgin and used to evaluate a technique for finding corners [26].

The *NicIcon Database of Handwritten Icons* [19] is a set of symbols, drawn by 32 different subjects, gathered for assessing pen input recognition technologies, representing images for emergency management applications. Here we use the subset of 400 multi-stroke symbols, annotated by Tumen and Sezgin [26].

The *ISraw* dataset is referred to as an *out-of-context* dataset, i.e., it is not linked to a domain. It was one of the datasets used to test the homonymous technique [28] and DPFRag [26]. It contains both line and arc primitives belonging to 400 unistroke symbols, drawn by 10 different subjects.

Figure 3 shows one random sample from each class of the three symbol set.

## 5 Results

In this section we report the results of our evaluation. As for the accuracy, we calculated all of the metrics described in the previous section. Furthermore, RankFrag’s accuracy is compared to that of other state-of-art methods by using the All-or-nothing metric. It is worth noting that, due to the unavailability of working prototypes, we did not directly test the other methods: we only report the performance declared by their respective authors.

The accuracy achieved by RankFrag on the three datasets is reported in Table 3. The results are averaged over the 30 performed trials.

Table 4 shows a comparison of the accuracy of RankFrag with other state-of-art methods. The methods considered

Metrics	COAD	NicIcon	ISraw
<i>Corners manually marked</i>	2271	867	1795
<i>Corners found</i>	2260.67	774.03	1790.80
<i>Correct corners</i>	2254.20	730.90	1784.33
<i>False positives</i>	6.47	43.13	6.47
<i>False negatives</i>	16.80	136.10	10.67
<i>Precision</i>	0.9972	0.9441	0.9964
<i>Recall / Correct corners accuracy</i>	0.9926	0.8428	0.9940
<i>All-or-nothing accuracy</i>	0.9870	0.8657	0.9572

Table 3: Average accuracy results of RankFrag on the three datasets.

Dataset	RankFrag	DPFRag	ISraw
COAD	0.99	0.97	0.82
NicIcon	0.87	0.84	0.24
ISraw	0.96	0.96	0.96

Table 4: Comparison of RankFrag with other methods on the All-or-nothing accuracy metric.

here are DPFRag [26] and ISraw [28]. Due to the unavailability of other data, we only report the results related to the All-or-nothing metric. As we can see, RankFrag outperforms the other two methods on two out of three datasets.

As for efficiency, we report that the average time needed to process a stroke is  $\sim 390$  ms. Our prototype is rather slow, due to the inefficiency of the calls to R functions. We also produced a non-JRI implementation by manually exporting the created random forest from R to Java (avoiding the JRI calls). With this implementation, the average execution time was lowered to  $\sim 18$  ms, enabling real-time runs.

## 6 Discussion and Conclusion

We have introduced RankFrag, a technique for segmenting hand-drawn sketches in the corner points. RankFrag has a quadratic asymptotic time complexity with respect to the number of sampled points in an input stroke. This complexity is the same reported in the literature for many other methods and, to the best of our knowledge, there is no method with a lower complexity. The technique was evaluated on three different datasets. The datasets were specifically produced for evaluating corner detection algorithms or were already used previously for this purpose.

We compared the results obtained by RankFrag with those already available in the literature for two different techniques: DPFRag [26] and ISraw [28]. With respect to the latter, our results show a clear advantage in accuracy on two datasets for RankFrag. With respect to DPFRag, our technique has a comparable accuracy, with a slight advantage on two of the three datasets. Nevertheless, compared to DPFRag, our technique has the additional advantage that

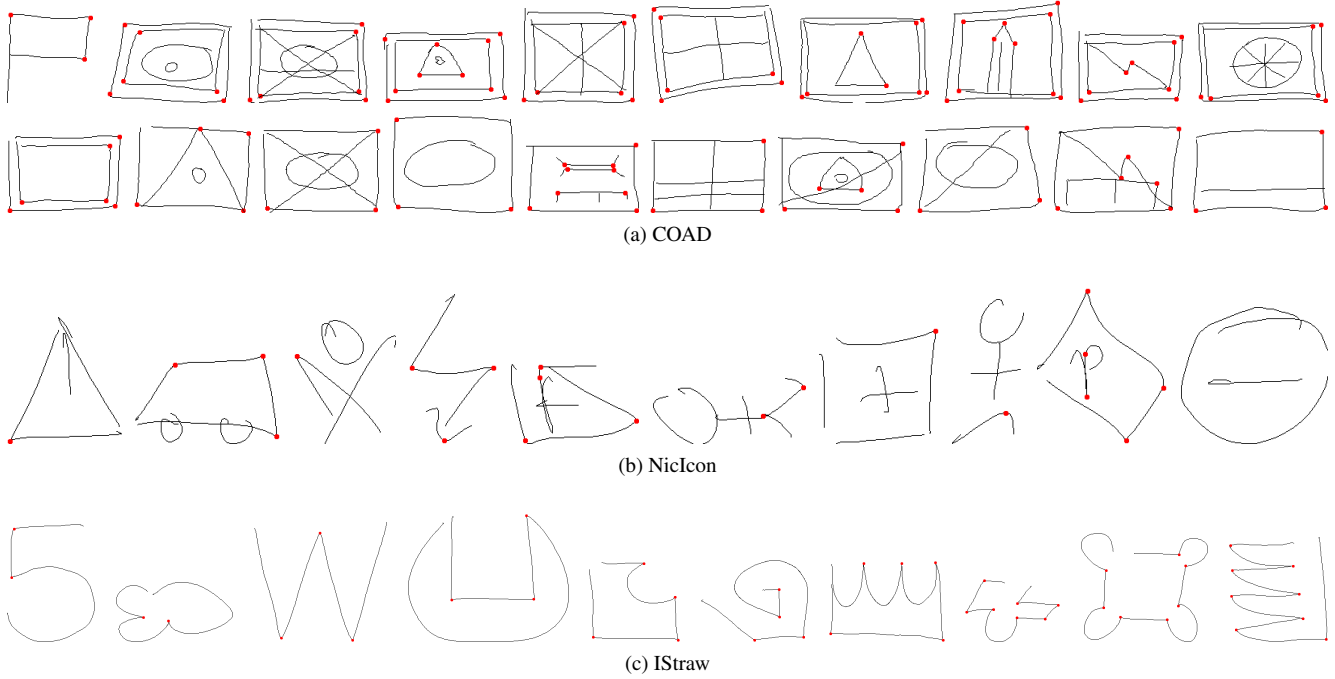


Figure 3: One random sample from each class of the three symbol set. The manually annotated corners are highlighted with a red circle.

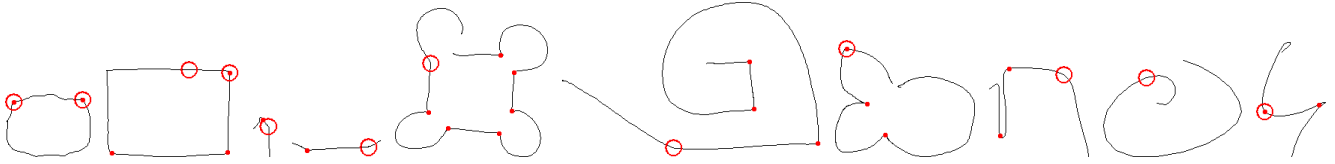


Figure 4: Examples of misclassification by RankFrag. Detected corners are represented through a red dot, while classification errors are represented through rings.

it can be performed in real time on all the tested data, regardless of the complexity of the input strokes. The chart reported in [26] (Figure 9) shows that this is not guaranteed for DPFRag and that its running time grows with the number of corners in the stroke.

RankFrag can be considered a significant improvement to the segmentation technique presented in [20]. In that technique, the classifier is used as a stop function: when the classifier decides to stop, all the remaining points (those with a higher cost) are classified as corners. We found that such a technique is not appropriate for strokes containing curves, since the cost function alone is not a reliable indicator and gives many false positives. Thus, we decided to invoke the classifier within a more complex, but still efficient, procedure, which performs further checks to establish whether a point is an angle. We also more profitably use a larger set of features, some of which have a variable region of support. Lastly, in our analyses the random forest

seemed to have better performance with respect to the other classifiers which we preliminarily tested, such as SVM and Neural Networks.

It is worth noting that, although further accuracy improvements are possible, it is very difficult to get a score close to 100% due to the procedure used in our tests: the decision of the classifier was compared to an earlier annotation made by a human operator. Some decisions are debatable and the annotation process is not free from errors. Figure 4 shows some examples of corner misclassification by RankFrag on the three datasets, including both false positives (dots inside a ring) and false negatives (rings). Although annotation errors are evident in some of the strokes reported in the figure, we decided not to alter the original annotation in order to obtain a more faithful comparison with the other methods.

RankFrag has only been tested for finding corner points and not *tangent vertices*, as done by other techniques

[12, 1]. It can be directly used in various structural methods for symbol recognition. However in some methods an additional step to classify the segments in lines or arcs may be required.

The non-JRI version of our implementation is able to produce the segmentation of a stroke in real time on a sufficiently powerful device. Future work will aim to achieve further implementation improvements, in order to further reduce the execution time and make the technique applicable in real time on more strokes at once (e.g., an entire diagram) or on mobile devices with low computational power. For testing purposes, our implementation can be downloaded at <http://weblab.di.unisa.it/rankfrag/>.

## References

- [1] F. Albert, D. Fernández-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] R. Bellman. On the approximation of curves by line segments using dynamic programming. *Commun. ACM*, 4(6):284, June 1961.
- [3] G. Costagliola, M. De Rosa, and V. Fuccella. Local context-based recognition of sketched diagrams. *Journal of Visual Languages & Computing*, 25(6):955 – 962, 2014.
- [4] 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(0):101 – 116, 2014.
- [5] G. Costagliola, V. Fuccella, and M. D. Capua. Interpretation of strokes in radial menus: The case of the keyscetch text entry method. *Journal of Visual Languages & Computing*, 24(4):234 – 247, 2013.
- [6] G. Costagliola, V. Fuccella, and M. Di Capua. Text entry with keyscetch. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*, IUI '11, pages 277–286, New York, NY, USA, 2011. ACM.
- [7] T. D.U. Commented APP-6A - Military symbols for land based systems, 2005.
- [8] J. Dunham. Optimum uniform piecewise linear approximation of planar curves. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(1):67–75, Jan 1986.
- [9] V. Fuccella and G. Costagliola. Unistroke gesture recognition through polyline approximation and alignment. In *Proceedings of CHI '15*, pages 3351–3354, New York, NY, USA, 2015. ACM.
- [10] R. Haddad and A. Akansu. A class of fast Gaussian binomial filters for speech and image processing. *Signal Processing, IEEE Transactions on*, 39(3):723–727, Mar 1991.
- [11] T. Hammond, B. Eoff, B. Paulson, A. Wolin, K. Dahmen, J. Johnston, and P. Rajan. Free-sketch recognition: Putting the chi in sketching. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 3027–3032, New York, NY, USA, 2008. ACM.
- [12] J. Herold and T. F. Stahovich. Speedseg: A technique for segmenting pen strokes using pen speed. *Computers & Graphics*, 35(2):250–264, 2011.
- [13] J. Herold and T. F. Stahovich. A machine learning approach to automatic stroke segmentation. *Computers & Graphics*, 38(0):357 – 364, 2014.
- [14] C. F. Herot. Graphical input through machine recognition of sketches. In *Proceedings of the 3rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '76, pages 97–102, New York, NY, USA, 1976. ACM.
- [15] D. H. Kim and M.-J. Kim. A curvature estimation for pen input segmentation in sketch-based modeling. *Computer-Aided Design*, 38(3):238 – 248, 2006.
- [16] 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.
- [17] B. Leo. Random forests. *Machine Learning*, 45(1):5–32, dec. 2001.
- [18] A. Liaw and M. Wiener. Classification and regression by randomForest. *R News*, 2(3):18–22, 2002.
- [19] R. Niels, D. Willems, and L. Vuurpijl. The nicicon database of handwritten icons. 2008.
- [20] T. Y. Ouyang and R. Davis. Chemink: a natural real-time recognition system for chemical drawings. In *Proceedings of the 16th international conference on Intelligent user interfaces*, IUI '11, pages 267–276, New York, NY, USA, 2011. ACM.
- [21] B. Paulson and T. Hammond. Paleosketch: accurate primitive sketch recognition and beautification. In *Proceedings of the 13th international conference on Intelligent user interfaces*, IUI '08, pages 1–10, New York, NY, USA, 2008. ACM.
- [22] T. M. Sezgin, T. Stahovich, and R. Davis. Sketch based interfaces: Early processing for sketch understanding. In *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, PUI '01, pages 1–8, New York, NY, USA, 2001. ACM.
- [23] T. F. Stahovich. Segmentation of pen strokes using pen speed. In *AAAI Fall Symposium Series*, pages 21–24, 2004.
- [24] C.-H. Teh and R. Chin. On the detection of dominant points on digital curves. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(8):859–872, Aug 1989.
- [25] C. Tirkaz, B. Yanikoglu, and T. M. Sezgin. Sketched symbol recognition with auto-completion. *Pattern Recognition*, 45(11):3926–3937, 2012.
- [26] R. S. Tumen and T. M. Sezgin. Dpfrag: Trainable stroke fragmentation based on dynamic programming. *IEEE Computer Graphics and Applications*, 33(5):59–67, 2013.
- [27] 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.
- [28] 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.

# WiSPY: A Tool for Visual Specification and Verification of Spatial Integrity Constraints

Vincenzo Del Fatto  
Faculty of Computer Science  
Free University of Bozen-Bolzano  
39100 Bolzano, ITALY  
vincenzo.delfatto@unibz.it

Luca Paolino  
Department of Research  
Link Campus University  
00162 Roma, ITALY  
l.paolino@unilink.it

Vincenzo Deufemia  
Department of Computer Science  
University of Salerno  
84084 Fisciano (SA), ITALY  
deufemia@unisa.it

Sara Tumiatì  
South Tyrolean Municipality Consortium  
39100 Bolzano, ITALY  
sara.tumiatì@gvcc.net

## Abstract

*Nowadays, most of tools for spatial data manipulation allow to edit information on maps without performing any integrity verification. On the other hand, data repositories such as the DBMS only permit few constraints to be defined by means of their Data Definition Languages and leave programmers to implement procedures for complex constraints. In this work we present the WiSPY system, a plugin of the GIS tool uDig for visually specifying and verifying complex spatial integrity constraints. WiSPY includes a visual environment for defining spatial data models with integrity constraints and for automatically generating the constraint checker. The latter is used by the WiSPY tool to verify the integrity of the data produced during the map editing process. The system has been validated on a real case study concerning the current regulation of the Public Illumination Plan (PIP) managed by an Italian municipality.*

## 1 Introduction

The management of spatial data is one of the fields where companies and researchers have invested much money and time in the last decade. The result is that commercial products such as Autodesk Autocad Map 3D<sup>1</sup>, Bentley Microsta-

tion<sup>2</sup>, ESRI ArcGIS<sup>3</sup>, or free and open source products such as Google Map<sup>4</sup>, QGIS<sup>5</sup>, GRASS GIS<sup>6</sup>, uDig [22], have become part of the daily life not only for GIS users. This is because, they offer a large amount of features for spatial data manipulation, spatial analysis and reasoning functionalities that in many cases may support or simplify our activities. Despite the large amount of available features, these products lack of an adequate control during the editing phase, not allowing a solid constraint check. Also in the Database Management Systems (DBMS) field, commonly used product as Oracle and PostgreSQL, which offer spatial extensions, only allow basic functionalities to support constraint checking. Indeed, their Data Definition Languages (DDL) only support the management of simple topological constraints while advanced controls need to be coded. In this context, it appears to be desirable to provide a significant support in this phase in order to improve the quality of data and minimizing the implementation activities which often are annoying and repetitive.

In order to increase the dataset quality, correction operations can be performed both during the editing phase (on the fly) or after a data manipulation session (*a posteriori*). Both such approaches have pros and cons. On one hand, checking correctness during the editing phase has a direct effect on the data entry process, since the feedback is immediate, but the check can be performed only on a subset

<sup>1</sup><http://www.autodesk.it/products/autocad-map-3d/overview>

<sup>2</sup><http://www.bentley.com/it-IT/products/microstation/>

<sup>3</sup><http://www.esri.com/software/arcgis>

<sup>4</sup><https://maps.google.com/>

<sup>5</sup><http://www.qgis.org/en/site/>

<sup>6</sup><http://grass.osgeo.org/>

of data. On the other hand, an a posteriori check is performed on the whole dataset or on a selected subset of data, giving the possibility to apply a complete verification and to globally re-adjust geographic data. However, in this case the system returns feedback to the user just at the end of the manipulation process, making more difficult to handle possible errors.

To guarantee the effective verification of map constraints according to the designer requirements, in [7] a visual language parsing approach for constraint checking of input spatial data during the editing phase is presented. The integrity of data produced during the map editing process is guaranteed by a constraint checker automatically generated from a visual language grammar. In order to reduce the efforts for defining the constraints to be checked, a high-level data model is used to specify the user needs.

In this paper we propose a software system, named WiSPY, which consists of two uDig plugins that allow the user to:

- specify geographic models by means of the OMT-G visual modeling language;
- automatically translate the OMT-G models to the corresponding grammars;
- validate geographic incoming data against the constraint checker generated from the grammars.

The rest of the paper is organized as follows. Section 2 presents the case study based on the Public Illumination Plan which we used to validate the proposed system. Section 3 introduces the OMT-G modeling language used for specifying spatial integrity constraints (SICs, for short). Section 4 describes the WiSPY tool, while Section 5 shows its application to the considered case study. Section 6 discusses the work existing in literature related with our proposal. Conclusions and future work are given in Section 7.

## 2 Case Study

The case study presented in this section represents the current regulation of the Public Illumination Plan (PIP) managed by the South Tyrolean Municipalities Consortium (STMC), in South Tyrol, Italy. The STMC<sup>7</sup> is a cooperative founded in 1954 that includes among its members all the south Tyrolean municipalities and is mainly focused on legal practice, administrative training, labor legislation, and ICT services. The PIP is a complex set of regulations that can be difficult to interpret and apply correctly. In particular, these regulations define a complex set of lighting categories depending on the type of road the lamp is placed (urban or extraurban roads, pedestrian zone, bicycle paths),

if the road is heavily busy or not, if there is a pedestrian passage, if there are crossing roads, and so on. The plan must be “safe for people and things” and implemented in order to limit light pollution. Light pollution is considered as a misdirected, excessive or obtrusive artificial light, causing a serious degradation of the natural nocturne light. A public administration must intervene in order to prevent such situations. In addition, a illumination system that involves wrong light bulbs, wrong lamp types or has an overestimation of the lamp power could create economical issues. The proposed WiSPY tool can help domain experts to better understand and effectively manage such a complex real world scenario.

The verification of the PIP is a typical task that a public administration is faced with and it is a complex task for different reasons, such as the difficulty of managing many types of geographic data involved in, as well as, the tight connection to the context they are inserted in. In fact, the simple containment spatial relationship is not sufficient to check a wide range of constraints that could depend on the context on which a lamp is being placed (type of road or area), the type of lamp itself, the type of illumination based on both lumen and lux. In addition, checking the correctness of the geographic data related to a street lamp could be tricky, because a lamp is normally represented as a point in space, while the constraints may need a polygon to be successfully checked. For example, checking the correct distribution of lamps along a road is not sufficient to compute the distance between the points of the lamps, it is also necessary to calculate the amplitude of the radiation given by the lux value.

The case study consists in the automatic verification of real geographic data related to the PIP of the South Tyrolean Municipalities. The data includes basic cartographic data (boundaries, hydrography, vegetation), roads, buildings, and of course the PIP. Based on the current regulations, a set of constraints suited to the validation of the chosen municipality’s PIP can be specified. The road types in the municipal boundaries are of type C (secondary extra urban road) and type F (local roads and bicycle paths). The following steps are necessary to determine which configuration is suitable for each road type:

- determine whether the road is of type C or F, including the speed limit;
- determine which technical illumination class is related to the road, in order to have the right luminance values;
- determine the type of lamp;
- determine the height of the poles;
- determine how to place the poles at the side of the road:
  - unilateral;

<sup>7</sup><http://www.gvcc.net>



- bilateral with alternate center;
- bilateral with opposite center;
- double centred (between the two carriageways);
- determine the distance of the poles.

All these factors must be taken into account during the verification process, and they depend on each others; determining the distance of the poles is the step that depends more on the other steps, whilst the first two steps are those that influence more the decision.

### 3 Visual Modeling Geographic Data embedding Integrity Constraints with OMT-G

Although existing data modeling approaches and tools can be adapted for geographic database design, most of them do not support certain aspects of the modeling process, such as the treatment of SICs. A visual language for modeling geographic data must be able to visualize different aspects of the data structure including numerous types of representations, such as point, line, polygon as well as non-spatial data; conventional as well as geo-referenced classes; different types of spatial relations, spatial constraints, spatial aggregation relationships.

Object Modeling Technique for Geographic Applications (OMT-G) is an object-oriented approach to model data for geographic information. Its notation is based on the classic OMT class diagram notation [5], and further extended to embrace also Unified Modeling Language (UML) concepts and notations [4]. OMT-G provides three types of primitives, based on the UML primitives for class diagrams, to model the geometry and topology of geographic data, providing support for topologic structures, network structures, multiple views of objects, and spatial relationships. These types are classes, relationships and SICs. These primitives allow also for the specification of alphanumeric attributes and associated methods for each class.

OMT-G offers three types of diagrams: the class diagram, that represents the classes involved in the model, as well as their relations; the transformation diagram, that permits the description of the transformation process of a class, if the class diagram indicates the need of multiple representation of it; the presentation diagram which describes how to represent the visual aspects of objects in the visualization.

OMT-G class diagrams are composed of conventional and geo-referenced classes. The first behave as UML classes and have no geographical properties. The latter include a geographical representation alternative, which specializes in two types of representations: discrete, associated with real world elements (geo-objects), or continuously distributed over the space (geo-fields). Geo-objects are represented with points, lines, polygons or network elements,

whereas geo-fields correspond to variables such as soil type, relief and temperature. The relationships of a OMT-G class diagrams can be conventional, e.g. UML relationships, or georeferenced. The latter include topological relations (e.g. touch, in, cross, overlap, and disjoint), arc-node network relations and spatial aggregations.

OMT-G class diagram permits the derivation of the set of SICs that must be observed in the implementation. SICs can be classified in: *topological* (the geometrical properties), *semantic* (the semantic of the geographic feature), and *user-defined* integrity constraints, “business rules” and all those controls that are non-spatial. Topological integrity constraints include spatial dependencies, spatial associations, connectivity, and geo-field rules. Semantic integrity constraints include spatial association and disjunction rules. User-defined integrity constraints are obtained from methods that can be associated to the classes.

## 4 The WiSPY Tool

In this section we present WiSPY (Visual specification & Verification of SPatial integritY constraints), an extension of the uDig GIS tool [22] for enabling users to visually model geographic applications, also embedding SICs, and to verify the correctness of the input geographic data. WiSPY has been implemented by means of two plugins. In the following we present the architecture of the WiSPY tool and provide details about the implemented plugins.

### 4.1 The Architecture

Figure 1 shows the architecture of the proposed WiSPY tool. It has been implemented on top of uDig, which is a software program based on the Eclipse platform featuring full-layered Open Source GIS. In particular, uDig provides a complete Java solution for viewing, editing, and accessing GIS data. Since it is built on top of the Eclipse “Rich Client Platform”, WiSPY has been developed in Java as two uDig plugins, namely OMT-G Editor and Constraint Checker.

The OMT-G Editor provides three different environments, one for each diagram the OMT-G data model provides. In the canvas of the class diagram editor, users specify their schema, adding classes and relationships chosen from the tool palette. The relationships selected from the palette can be inserted by clicking over the source class and dragging a line to the target class. As an example, Figure 2 shows simple OMT-G class diagram modeling *containment* constraints among Municipality, Lamp, and Road objects.

The OMT-G editor includes a function to derive a visual grammar modeling the SICs specified in the OMT-G data model. The Constraint Checker generated from the grammar by using the ANTLR parser generator<sup>8</sup> can be activated

<sup>8</sup><http://www.antlr.org/>

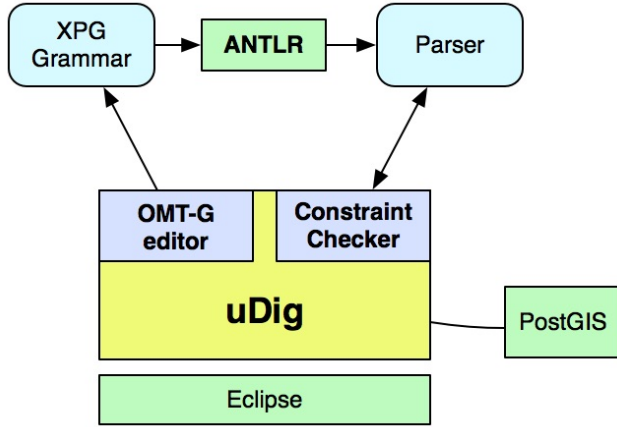


Figure 1: The architecture of WiSPY tool.

by the user during map editing phase. In particular, the input of WiSPY is a set of geographic data whose type is defined in the OMT-G data model. The output is the validation of the input data with respect to the SICs specified between the classes of the OMT-G data model. If a SIC is violated then a suitable error message is shown to user with information to recover from the violation.

## 4.2 Constraint Checker Generation

The WiSPY tool automatically generates a constraint checker able to verify the SICs defined by a OMT-G class diagram. In particular, WiSPY exploits the visual language compiler-compiler technique proposed in [6] for deriving a visual language parser through standard compiler-compilers, like YACC [11].

The OMT-G Editor allows users to develop their schema, adding classes and relationships chosen from the tool palette. The relationships can be annotated with SICs, which impose restrictions on the input data. In particular, the classes of the model represents the geographical objects that the user can place on the map, while the relationships specified between two classes define SICs on their instances. Such constraints are defined on the attributes of the involved classes. The editor provides the set of standard OMT-G spatial integrity rules (e.g., *contain relation*, *coincide relation*, *cross relation*, *touch*, *in*) as well as standard processes such as generalization and specialization. Moreover, the users can define new rules by specifying a set of conditions on the classes' attributes.

The constraint checker generation process consists of mapping the OMT-G class model into a visual grammar. To this end, we use the XPG grammar formalism [6], which is similar to context-free string grammars, where more general relations other than concatenation are allowed. In particular, an XPG textually describes a diagram by grammar

productions that alternate (terminal and nonterminal) symbols with relations defined on the symbol attributes. Thus, the idea is to map the classes defined in the OMT-G schema, which represent the spatial objects to be placed on the map, into terminal symbols of the grammar, while the SICs defined between the spatial objects are modeled in terms of spatial relations among them [6]. In this way, the user can analyze the SICs specified for a particular application domain and, eventually, customize some of them interacting with the editor.

For instance, the containment constraint between *Road* and *Lamp* in Fig. 2 is modeled by the production:

$Roads \rightarrow ROAD \langle contains \rangle LAMP$

where *contains* is an empty production with associated a semantic action that verifies the satisfiability of the relationship [6]. *ROAD* and *LAMP* are terminal symbols having associated the set of attributes defined in the corresponding classes of the OMT-G model, e.g., *type* for *ROAD*. Such attributes are used by the *contains* production to verify the spatial constraint. The nonterminal symbol *Roads* has associated a set of attributes whose value is synthesized from the attribute values of *ROAD* and *LAMP*.

The WiSPY tool provides the implementation of semantic actions for a predefined set of constraints. However, as said above, the tool enables users to define their own constraints. In particular, the user can annotate a OMT-G relationship connecting two classes *A* and *B* with a boolean condition on the attributes of *A* and *B*. As an example, for the OMT-G diagram in Figure 2, a user could define the following illuminance constraint: *the lightning of lamps associated to urban highways is greater than 40SB<sup>2</sup>*. This constraint is named *illuminates* and is defined by the following boolean expression:

$(Road.type = 'Highway' \wedge Lamp.lightning > 40)$ .

The constraint checker is obtained by giving as input to a compiler-compiler the grammar automatically generated from the OMT-G model. Since WiSPY uses the ANTLR parser generator to perform this task, we represent the XPG grammar into a format compatible with ANTLR. The use of

Fig. 3 shows the grammar constraint checker editor embedded into the uDig interface. In particular, in the right side of the interface (label D), the palette contains all the suitable tools for grammar checking, and the "Select Feature Set" operator is activated. By using this operator, users can select geographic features into the area of interest by using a simple rectangle selection tool on the standard uDig map view. After this operation the geographic data of interest are selected and highlighted in yellow in the map (see label A). In this example, the highlighted polygons represent areas, while the highlighted points represent lamps. In this case, only the selected geographic features are involved in the constraint check process. At the bottom left side of the interface (label B), the details about the selected features

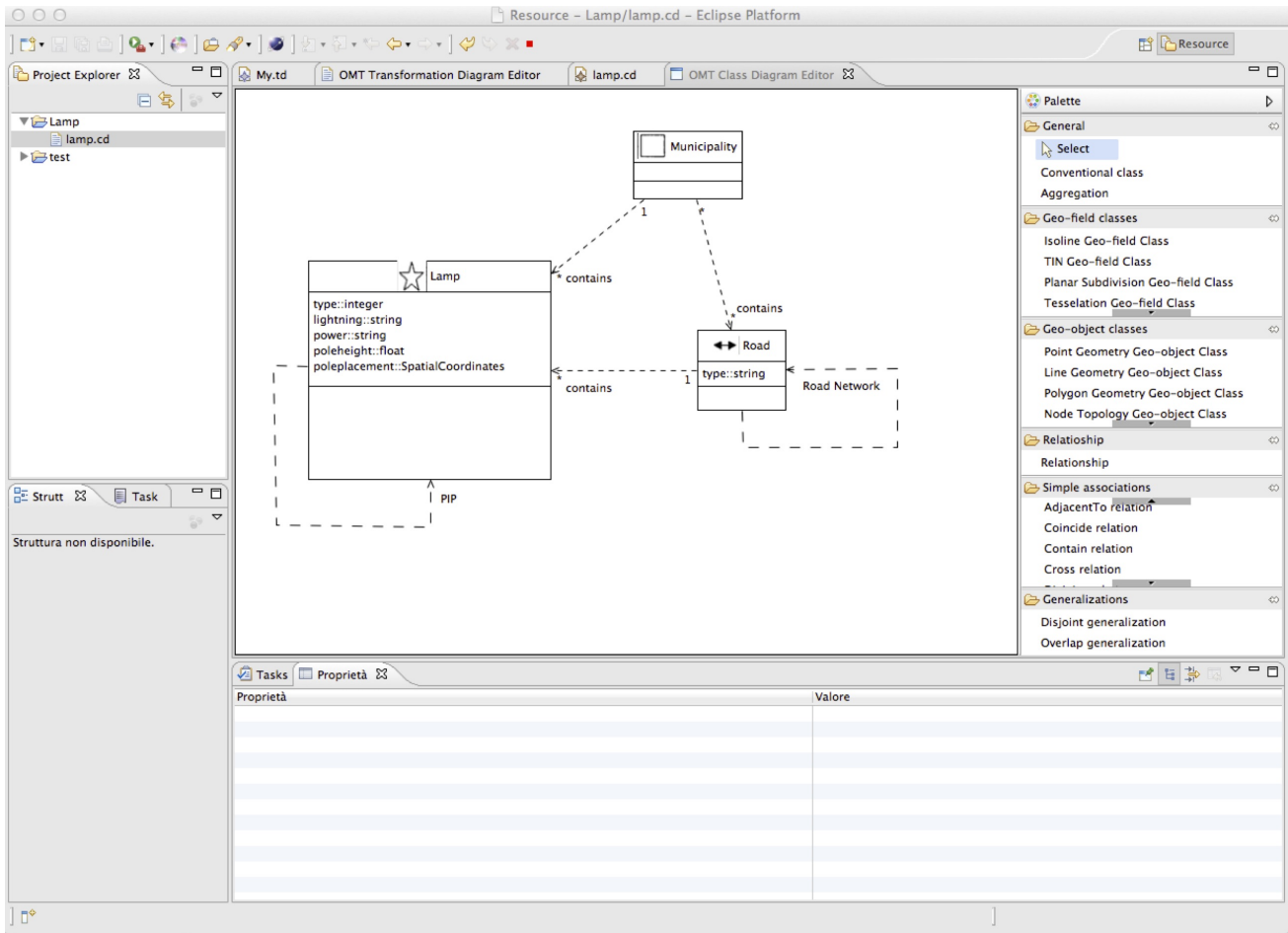


Figure 2: OMT-G interface for specifying the spatial integrity constraints.

are shown. Finally, at the bottom right side of the interface (label C), the output console of the validation process is shown.

### 4.3 Verification of SICs

The parser generated with ANTLR is used by WiSPY to validate the input spatial data against the SICs specified in the OMT-G model. In particular, the parser analyzes the spatial objects positioned on a map driven by the relationships specified in the grammar. If the spatial objects violates a SIC then it yields a parse error.

Fig. 4 shows the result of the constraint checking process in the WiSPY interface. In this example, points representing lamps are highlighted by using the “Select Feature Set” operator. Executing the constraint checking, the steps performed by the validation process are listed into the console view, located at the bottom of the interface. If an error occurs, it is reported to the user in the console view.

## 5 Checking SICs for Public Illumination Plans

The best way to illustrate how to apply our system to real problems is through an example on PIP case study. In this domain, lamps are spatial objects having associated the following information: localization of the lamp ( municipality, hamlet, street, GPS coordinates), number of light points for every lamp, lamp type, type of light source, number of light sources for light point, electric power for each light source, year, overall electric power of the lamp, mounting typology (wall or pole), pole type, pole height, circuit and electric power panel, road classification and technical illumination classification. The roads are classified as:

- Category A: highways;
- Category B: high-speed extraurban roads;
- Category C: secondary extraurban roads;
- Category D: urban arterial roads;

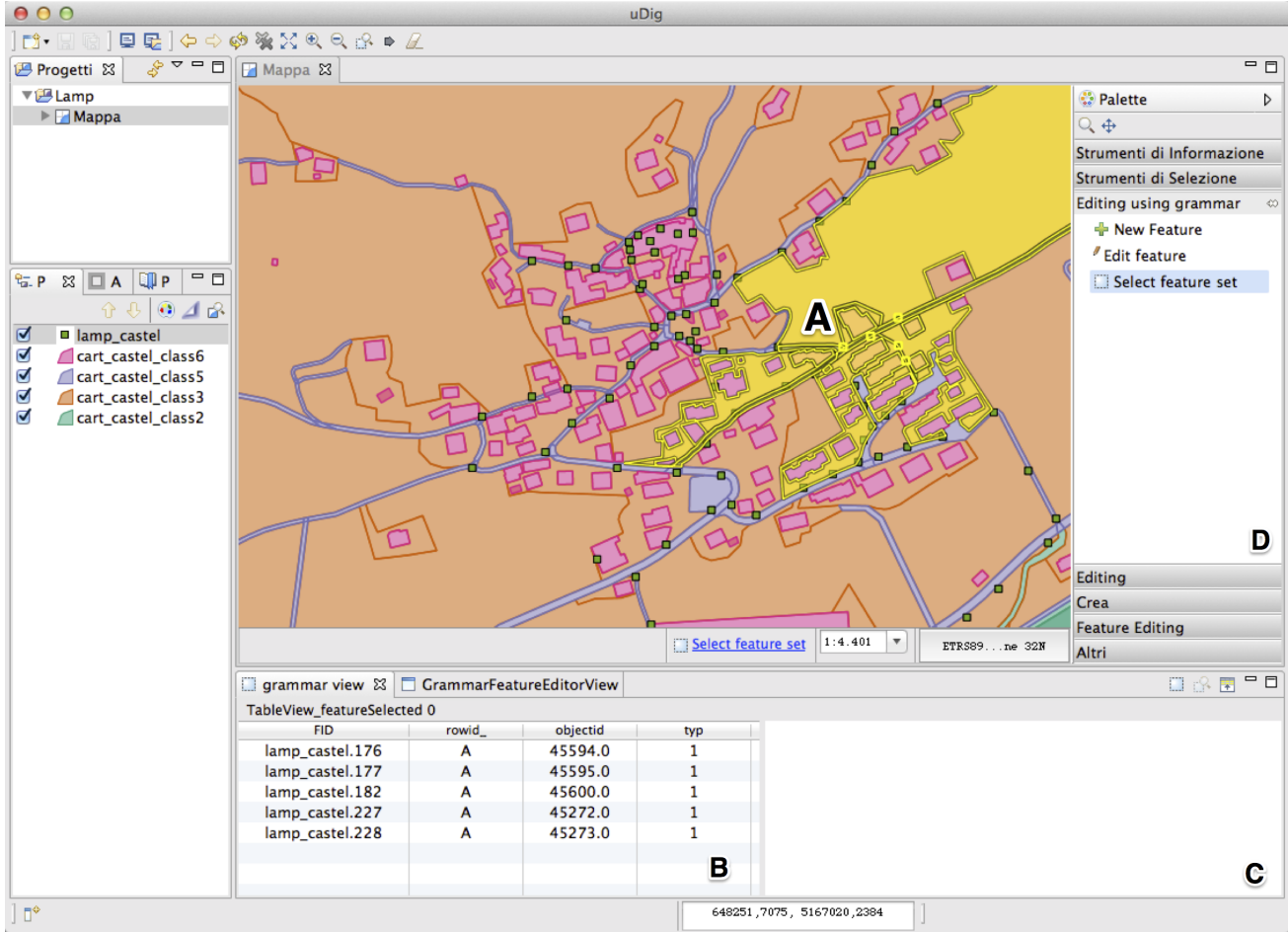


Figure 3: WiSPY main window with the additional tools for grammar parsing.

- Category E: urban district roads;
- Category F: local roads.

Road illumination varies depending on road classification and on the related technical illumination classification. This classification is the same specified in the UNI EN 13201 European normative, which states that the illumination level is based on the traffic intensity of the road and on the daytime. Therefore the technical illumination classification of a road may vary during the daytime. Since the geographic data used for the prototype are coming from a municipality far away from highways, the classification used in WiSPY is simplified as reported in Tables 1 and 2. The illumination parameters reported in Table 2 refers to:

- $\bar{L}(cd/m^2)$  is the average road surface luminance of a carriageway of a road expressed in candelas per square meter;
- $U_o$  is the overall uniformity of road surface luminance;

- $U_l$  is the longitudinal uniformity of road surface luminance;
- $TI$  is the threshold increment, which measures the loss in percentage of visibility caused by the disability glare of the luminaries of a road lighting installation;
- $SB^2$  is the surround ratio of illumination of a carriageway of a road
- $\bar{E}$  is the hemispherical illuminance averaged over a road area expressed in lux.

Road Type	Road description	Speed limit (km/h)	Tech. Illum. Class.
C - Secondary extraurban roads	Extraurban road	70-90	ME3a
	Local extraurban roads	50	ME4b
F - Local roads	Local urban roads	30	S3
	Historic town center	-	CE4
F - Pedestrian and bicycle routes	-	-	S3

Table 1: UNI EN 13201 road classification (subset).

In order to illustrate how WiSPY is able to identify violations in the public illumination plan, in the following we

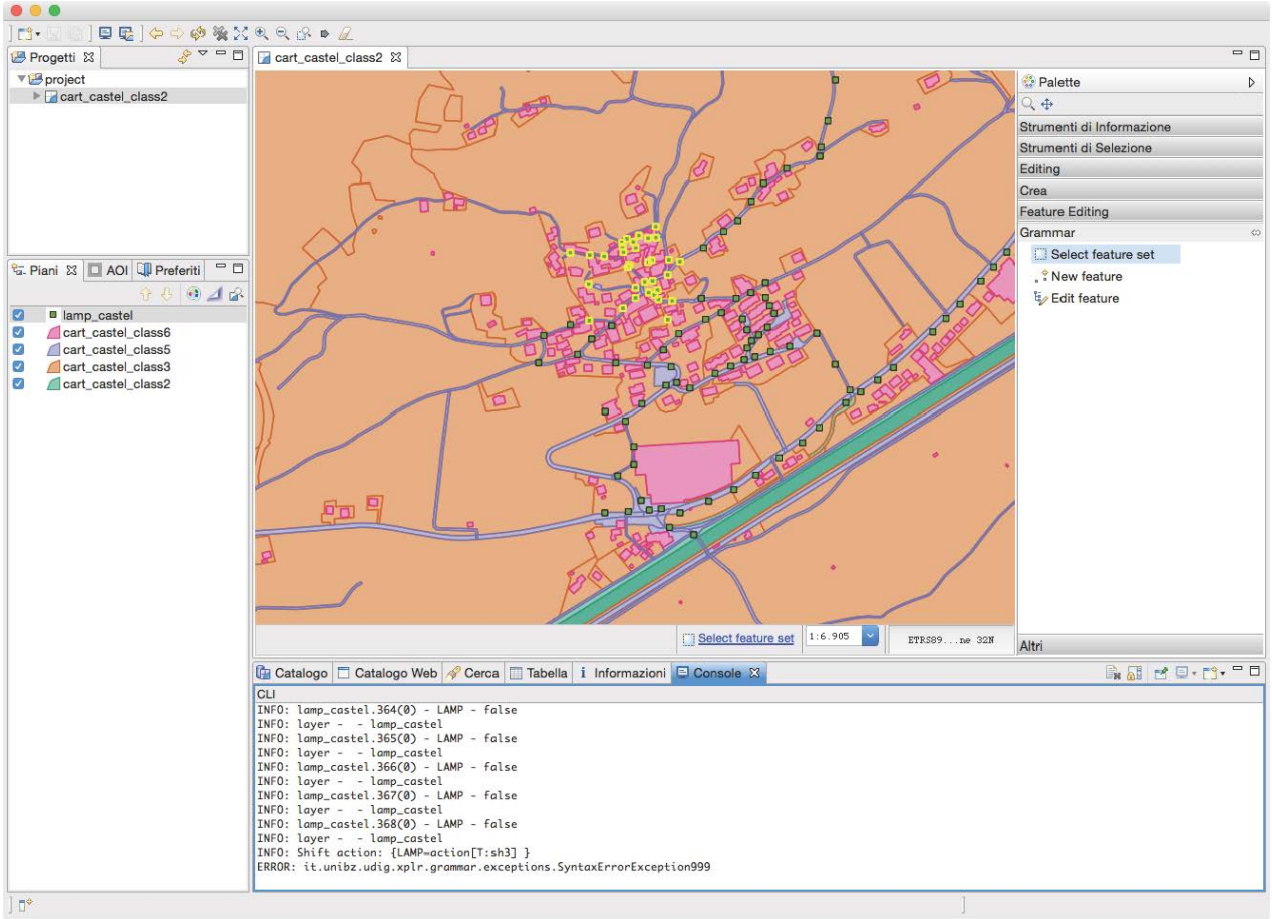


Figure 4: WiSPY interface showing the result of the constraint checking process.

Class	Detail type	Detail values
ME3a	Luminance of the road surface of the carriageway	$1,0 \bar{L}(\text{cd}/\text{m}^2)$
		$0,4 U_o$
		$0,7 U_i$
	Disability glare	$15 \text{ TI in } \%$
ME4b	Lighting of surroundings	$0,5 \text{ SB}^2$
	Luminance of the road surface of the carriageway	$0,75 \bar{L}(\text{cd}/\text{m}^2)$
		$0,4 U_o$
		$0,6 U_i$
S3	Disability glare	$15 \text{ TI in } \%$
	Lighting of surroundings	$0,5 \text{ SB}^2$
	Horizontal illuminance	$7,5 \bar{E} \text{ in lx}$
CE4	Horizontal illuminance	$1,5 \bar{E}_{min} \text{ in lx}$
		$10 \bar{E} \text{ in lx}$
		$0,4 U_o$

Table 2: Considered UNI EN 13201 technical illumination classes.

sketch the grammar derived from the OMT-G model in Figure 2. In particular, the terminal symbols of the grammar correspond to the classes of the model, i.e., MUNICIPALITY, LAMP, and ROAD, while the productions are generated according to the class relationships specified in the OMT-G model:

1.  $\text{Map} \rightarrow \text{MUNICIPALITY} \langle \text{contains} \rangle \text{Objects}$ ;
2.  $\text{Objects} \rightarrow \text{Lamps} \langle \text{union} \rangle \text{Roads}$ ;
3.  $\text{Lamps} \rightarrow \text{Lamp} \langle \text{pip} \rangle \text{Lamps}$
4.  $\text{pip} \rightarrow \epsilon$   
**SemanticAction:** {  
**if**  $\text{distance}(\text{Lamp}, \text{Lamps}) < \text{MIN\_LAMP\_DIST}$   
**then**  
 $\text{parse.alert}(\text{'CONSTRAINT VIOLATION'}$ ,  
 $\text{'Lamp'} + \text{Lamp.pos} + \text{' is too close to other lamps'}$ );  
**}**
5.  $\text{Lamps} \rightarrow \text{Lamp}$



6. Lamp  $\rightarrow$  LAMP
7. Roads  $\rightarrow$  Road  $\langle touches \rangle$  Roads
8. Roads  $\rightarrow$  Road
9. Road  $\rightarrow$  ROAD
10. Road  $\rightarrow$  ROAD  $\langle contains, isIlluminated \rangle$  Lamp
11. *isIlluminated*  $\rightarrow \epsilon$   
**SemanticAction:** {  
**if** (ROAD.type='C'  $\wedge$  Lamp.lightning $\neq$ 0.5)  $\vee$  (...)   
**then**  
*parse.alert('CONSTRAINT VIOLATION',*  
*'Lamp'+Lamp.position + ' violates the '*  
*'illumination of road ' + ROAD.name);*  
**}**

The productions have associated semantic actions that analyze the values of the attributes associated to the spatial symbols and verify whether the PIP constraints are satisfied. In particular, the first production indicates that a map is composed of a municipality symbol (visually defined in Figure 2 with a polygon) containing within its area other objects. The latter can be Lamps and/or Roads as defined in production 2. The set of lamps in the municipality has to satisfy the constraint associated to the *PIP* relationship in Figure 2, which defines the compatibility constraints among lamps. As an example, the semantic action associated to production 4 checks if the lamps are positioned too close. In this case, a message is shown to the user. Thus, productions 3-6 define the nonterminal Lamps as a set of LAMP positioned within a MUNICIPALITY according to compatibility constraints. Similarly, productions 7-9 define the nonterminal Road as a set of ROAD symbols positioned within a MUNICIPALITY and related through a *touch* relationship. Productions 10 and 11 define the compatibility constraint between roads and lamps according to the classifications reported in Tables 1 and 2. In particular, each lamp is associated to road and has to satisfy the user-defined constraint *isIlluminated*. Notice that, for readability of productions, we have omitted the semantic actions that synthesize the attributes for the LHS nonterminal from the attributes of the RHS (non)terminals.

The parser automatically generated from the previous grammar is able to analyze the municipality, road, and lamp symbols positioned by the user on a map, as shown in Figure 4, and verify whether the previously described SICs are violated. As an example, when the parser analyzes the lamps positioned on a map it applies productions 3 and 4 trying to reduce the LAMP terminal symbols into Lamps nonterminal symbols. If a lamp is too close to a lamp already analyzed (the spatial coordinates of the lamps previously analyzed by the parser are associated to Lamps' nonterminal symbol) then violation message is shown to the user.

When a SIC is violated by two or more geographical objects WiSPY shows a message with the information on the objects involved in the violation and the type of violation.

Thw WiSPY approach simplifies the specification and verification of SICs since the geographic application domain can be easily modeled with OMT-G class diagrams, the SICs can be specified as visual relationships between classes and customized using boolean conditions on attribute values, and the constraint checker can be automatically obtained from the annotated OMT-G model. In this way, the user can easily customize/add new SICs and rapidly prototyping new constraint checkers.

## 6 Related Work

The quality of spatial databases is an open problem in the field of geographic information systems and, in the last few decades, many efforts have been done to deal with implementation and management issues [15, 23]. In the following, we highlight the most important features of these works.

A constraint solver of spatial data based on programming logic has been presented in [1, 2]. The constraint system is able to handle the basic spatial types such as points, lines and polygons as well as the constraints in terms of equalities and inequalities, memberships, metric, topological and structural constraints. The system also provides a suitable theory for managing constraints and a set of transformation rules. The latter handle a special kind of constraints used for consistency checking, enabling an optimized and efficient resolution of spatial constraints.

In [12] a dimension graph representation is used for maintaining the spatial constraints among objects in an Euclidean space. The constraint consistency checking problem is transformed into a graph cycle detection problem on dimension graph.

The process for discovering inconsistencies in geographical dataset described in [19] consists of three steps: error definition, error checking, and error correction. Basically, the first step consists of the execution of some computational geometry algorithms, while the third one is solved by applying the first order calculus predicates.

In [14] a system developed for automatically maintaining topological constraints in a geographic database is presented. This system is based on extending to spatial data the notion of standard integrity maintenance through active databases. Topological relationships, defined by the users, are transformed into SICs, which are stored in the database as production rules. A similar approach is also introduced in [3].

An automated constraint checking procedure has been introduced by Udagepola *et al.* [21] to check constraint violations at compiling time before updating the database. It

is based on a data structure called Semantic Spatial Outlier R-Tree (SSRO-Tree).

In [17] Rigaux *et al.* presented Dedale, a constraint-based spatial database system relied on a linear constraints logical model. This system provides both an abstract data model and a user declarative query language based on SQL in order to represent and manipulate geometric data in arbitrary dimension. A different approach which combines relational and constraint data models is used in [10], where a three-tier constraint database architecture is presented. The latter increases the level of abstraction between the physical data and its semantics by introducing an additional layer to the classical relational data model architecture (logical and physical layer), which allows to manage both constraint-based and geometric data representations in the same layer of abstraction, in opposition to the pure constraint databases, where all data are represented in terms of constraints.

The framework presented in [20] allows the definition of hierarchical descriptions of abstract regions. To this aim, the framework exploits attributed grammars which can be translated by a compiler of compiler to a parser for abstract regions. Once generated, the parsers can be used for evaluating whether the incoming regions are consistent with the specified patterns. Basically, the abstract region candidates that were identified by the parsing rules can be evaluated to check if they conform to the definition provided by the user.

On the commercial side, Oracle® Spatial<sup>9</sup> allows spatial constraint checking by using either the PL/SQL language or by defining the constraint within the table procedure. ArcGIS<sup>10</sup> provides users with a button bar where it is possible to visually define simple constraints. More complex constraints have to be implemented by specific languages.

A significant part of the proposed WiSPY tool concerned with the visual definition of spatial constraints. The choice we made for this purpose is using the OMT-G modelling language. Similar to other approaches, it uses some visual formalisms for describing the spatial objects composing the geodatabase and others for connecting the objects specifying the relationships existing among them. We have chosen OMT-G [4] for its capability of explicitly specifying the constraints in associations and attributes [9], which is a limitation of the models extending UML [18], such as Ext. UML [16] and GeoFrame [8]. Moreover, OMT-G seems to be the most simply and user-friendly notation for non-expert constraint designers. Along this line, in [13] Lizardo and Davis presented a tool which provides various consistency checks on the integrity of the defined schema, and includes a function that maps OMT-G geographic concep-

tual schemas into physical schemas, including the SICs. Although, it seems very similar to our approach, it is based on SQL constraints which considerably limits the power of constraint checking.

## 7 Conclusions

In this paper we have proposed a system to support users in the automatic verification of SICs in geographic applications by exploiting visual language parsing. We have demonstrated, by implementing the WiSPY tool, that the visual language parsing is suitable for identifying violation in the PIP case study and for solving ambiguities that may arise in their interpretation. We have motivated our choice of having an entirely visual system, and highlighted its advantages. This choice represents the major difference between our proposal and the related work.

Our future work will focus on the extension of the current prototype in a fully functional product. Moreover, we will concentrate our efforts on finding appropriate solutions to present the information provided to the user, feedback and solutions, in a flexible and supportive manner.

## References

- [1] J. M. Almendros-Jiménez. Constraint logic programming over sets of spatial objects. In *Proceedings of the 2005 ACM SIGPLAN Workshop on Curry and Functional Logic Programming*, WCFLP '05, pages 32–42, New York, NY, USA, 2005. ACM.
- [2] J. M. Almendros-Jiménez and A. Corral. Solving constraints on sets of spatial objects. In M. V. Hermenegildo and D. Cabeza, editors, *Practical Aspects of Declarative Languages, 7th International Symposium, PADL 2005, Long Beach, CA, USA, January 10-11, 2005, Proceedings*, volume 3350 of *Lecture Notes in Computer Science*, pages 158–173. Springer, 2005.
- [3] A. Belussi, E. Bertino, and B. Catania. Manipulating spatial data in constraint databases. In M. Scholl and A. Voisard, editors, *Advances in Spatial Databases, 5th International Symposium, SSD'97, Berlin, Germany, July 15-18, 1997, Proceedings*, volume 1262 of *Lecture Notes in Computer Science*, pages 115–141. Springer, 1997.
- [4] K. A. V. Borges, C. A. Davis, and A. H. F. Laender. OMT-G: an object-oriented data model for geographic applications. *GeoInformatica*, 5(3):221–260, 2001.
- [5] K. A. V. Borges, A. H. F. Laender, and C. A. Davis. Spatial data integrity constraints in object oriented geographic data modeling. In C. B. Medeiros, editor, *ACM-GIS '99, Proceedings of the 7th International Symposium on Advances in Geographic Information Systems, November 2-6, 1999, Kansas City, USA*, pages 1–6. ACM, 1999.
- [6] G. Costagliola, V. Deufemia, and G. Polese. Visual language implementation through standard compiler-compiler techniques. *J. Vis. Lang. Comput.*, 18(2):165–226, 2007.

<sup>9</sup>[https://docs.oracle.com/cd/E18283\\_01/appdev.112/e11830/sdo\\_intro.htm#insertedID0](https://docs.oracle.com/cd/E18283_01/appdev.112/e11830/sdo_intro.htm#insertedID0)

<sup>10</sup><https://sites.google.com/site/ochaimwiki/geodata-preparation-manual/how-to-check-topology-using-arcgis>

- [7] V. D. Fatto, V. Deufemia, and L. Paolino. Map integrity constraint verification by using visual language parsing. *JDIM*, 6(4):332–341, 2008.
- [8] J. L. Filho and C. Iochpe. Specifying analysis patterns for geographic databases on the basis of a conceptual framework. In *Proceedings of the 7th ACM International Symposium on Advances in Geographic Information Systems*, GIS '99, pages 7–13, New York, NY, USA, 1999. ACM.
- [9] A. Friis-Christensen, N. Tryfona, and C. S. Jensen. Requirements and research issues in geographic data modeling. In W. G. Aref, editor, *ACM-GIS 2001, Proceedings of the Ninth ACM International Symposium on Advances in Geographic Information Systems, Atlanta, GA, USA, November 9-10, 2001*, pages 2–8. ACM, 2001.
- [10] D. Q. Goldin. Taking constraints out of constraint databases. In B. Kuijpers and P. Z. Revesz, editors, *Constraint Databases, Proceedings of the 1st International Symposium on Applications of Constraint Databases, CDB'04, Paris, June 12-13, 2004*, volume 3074 of *Lecture Notes in Computer Science*, pages 168–179. Springer, 2004.
- [11] S. Johnson. *YACC: Yet Another Compiler Compiler*. Bell Laboratories, Murray Hills, NJ, 1978.
- [12] X. Liu, S. Shekhar, and S. Chawla. Consistency checking for euclidean spatial constraints: a dimension graph approach. In *12th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2000), 13-15 November 2000, Vancouver, BC, Canada*, page 333. IEEE Computer Society, 2000.
- [13] L. E. O. Lizardo and C. A. D. Jr. OMT-G designer: A web tool for modeling geographic databases in OMT-G. In M. Indulska and S. Purao, editors, *Advances in Conceptual Modeling - ER 2014 Workshops, ENMO, MoBiD, MReBA, QMMQ, SeCoGIS, WISM, and ER Demos, Atlanta, GA, USA, October 27-29, 2014. Proceedings*, volume 8823 of *Lecture Notes in Computer Science*, pages 228–233. Springer, 2014.
- [14] C. B. Medeiros and M. Cilia. Maintenance of binary topological constraints through active databases. In *Proceedings of the 3rd ACM International Workshop on Advances in Geographic Information Systems, Baltimore, Maryland, December 1-2, 1995, in conjunction with CIKM 1995.*, page 127, 1995.
- [15] L. Plümer and G. Gröger. Achieving integrity in geographic information systems maps and nested maps. *GeoInformatica*, 1(4):345–367, 1997.
- [16] R. Price, N. Tryfona, and C. S. Jensen. Extended spatiotemporal UML: motivations, requirements and constructs. *J. Database Manag.*, 11(4):14–27, 2000.
- [17] P. Rigaux, M. Scholl, L. Segoufin, and S. Grumbach. Building a constraint-based spatial database system: model, languages, and implementation. *Inf. Syst.*, 28(6):563–595, 2003.
- [18] J. Rumbaugh, I. Jacobson, and G. Booch. *Unified Modeling Language Reference Manual, The (2Nd Edition)*. Pearson Higher Education, 2004.
- [19] S. Servigne, T. Ubeda, A. Puricelli, and R. Laurini. A methodology for spatial consistency improvement of geographic databases. *GeoInformatica*, 4(1):7–34, 2000.
- [20] J. Steinhauer, T. Wiese, C. Freksa, and T. Barkowsky. Recognition of abstract regions in cartographic maps. volume 2205 of *Lecture Notes in Computer Science*, pages 306–321. Springer, 2001.
- [21] K. P. Udagepola, L. Xiang, L. H. Wei, and Y. X. Zong. Efficient management of spatial databases by data consistency and integrity constraints. *WSEAS Transactions on Computers*, 5(2):447–454, 2006.
- [22] uDig. User-friendly desktop internet gis. <http://udig.refractions.net/>, 2004.
- [23] M. F. Worboys. A unified model for spatial and temporal information. *Comput. J.*, 37(1):36–34, 1994.

# GO-Bayes Method for System Modeling and Safety Analysis

Guoqiang Cai, Limin Jia, Hui Zhen,  
Mingming Zheng, Shuai Feng  
State Key Lab of Rail Traffic Control & safety  
Beijing Jiaotong University, Beijing, China  
guoqiangcai@163.com

MengChu Zhou  
Department of Electrical and Computer Engineering  
New Jersey Institute of Technology, Newark, NJ 07102,  
USA  
zhou@njit.edu

**Abstract**—Safety analysis ensuring the normal operation of an engineering system is important. The existing safety analysis methods are limited to relatively simple fact description and statistical induction level. Besides, many of them enjoy poor generality, and fail to achieve comprehensive safety evaluation given a system structure and collected information. This work describes a new safety analysis method, called a GO-Bayes algorithm. It combines structural modeling of the GO method and probabilistic reasoning of the Bayes method. It can be widely used in system analysis. The work takes a metro vehicle braking system as an example to verify its usefulness and accuracy. Visual implementation by Extendsim software shows its feasibility and advantages in comparison with the Fault Tree Analysis (FTA) method.

**Keywords:** Safety analysis, GO-Bayes method, and Reliability

## I. INTRODUCTION

Safety evaluation technologies were originated in the 1930s. In the 1960s, the needs from the US military field's engineering system safety theory and applications promoted their rapid development<sup>[1]</sup>. As people's awareness of safety continues to grow and system safety engineering becomes a mature discipline, a system safety engineering approach is gradually extended to aviation, nuclear industry, petroleum, chemical, and manufacturing areas. Researchers have proposed new theories and methods, such as safety checklist<sup>[2]</sup>, safety analysis<sup>[3]</sup> and evaluation methods<sup>[4]</sup>, event trees<sup>[5]</sup>, fault trees<sup>[6]</sup> and risk assessment techniques<sup>[7-8]</sup>, mode evaluation, six-stage safety and other risk index evaluation method, artificial neural networks and other technologies.

The GO method has commonly been used since 1980s<sup>[9]</sup>. Several improved methods for quantitative analysis are proposed in signal processing<sup>[10]</sup>. This work intends to improve the GO algorithm based on Bayes reasoning<sup>[11-13]</sup> and names the new method as a GO-Bayes algorithm. It has the following innovative characteristics:

First, the structural modular reliability analysis of the GO method is applied to analyze the operational status of a safety analysis assessment system; Second, the Bayes probability theory is used in a safe state probability parameter extraction process to each basic unit of the model; Third, the Bayes

inference is integrated into the system GO graph model, reversing fault reasoning analysis and evaluation, thereby achieving simpler quantitative analysis. The proposed GO-Bayes method combines the structural modeling of the GO method and probabilistic reasoning of the Bayes method<sup>[14]</sup>, which can be used in situations where one has a large amount of system fault information. Its use can help one prevent and diagnose faults in a timely fashion, thus ensuring the safe operation of an entire system.

## II. GO-BAYES METHOD

The proposed GO-Bayes method is system-unit-component failures oriented. It combines basic unit models and logic analysis models according to a flow chart to establish the analysis model, in accordance with certain rules to calculate reliability parameters. Besides we adopt Bayes methods to deduce system failure and solve inverse probability, in order to achieve a comprehensive system safety evaluation. The GO-Bayes method's operators are shown in Figure 1.

### A. Modeling method

The GO-Bayes method inherits graph modeling ideas, e.g., schematic diagrams, flow charts and other drawings. First, we summarize the basic model elements, and explain the unit algorithm. Second, we build a system model according to a system structure and data flows among its units. We use system modeling algorithms to process raw input data and then obtain system outputs according to the working mechanism and fault conditions.

### B. Bayes theory based on information fusion

Information fusion research based on the Bayes theory is mainly used for system internal self-monitor and self-test information. The information (hereinafter collectively referred to as detection information) plays a strong role in system operation safety analysis. The GO-Bayes method is based on the description and information of each component and subsystem, and the model is systematically analyzed. Fault information related to the system reliability and safety is integrated for system reliability analysis.

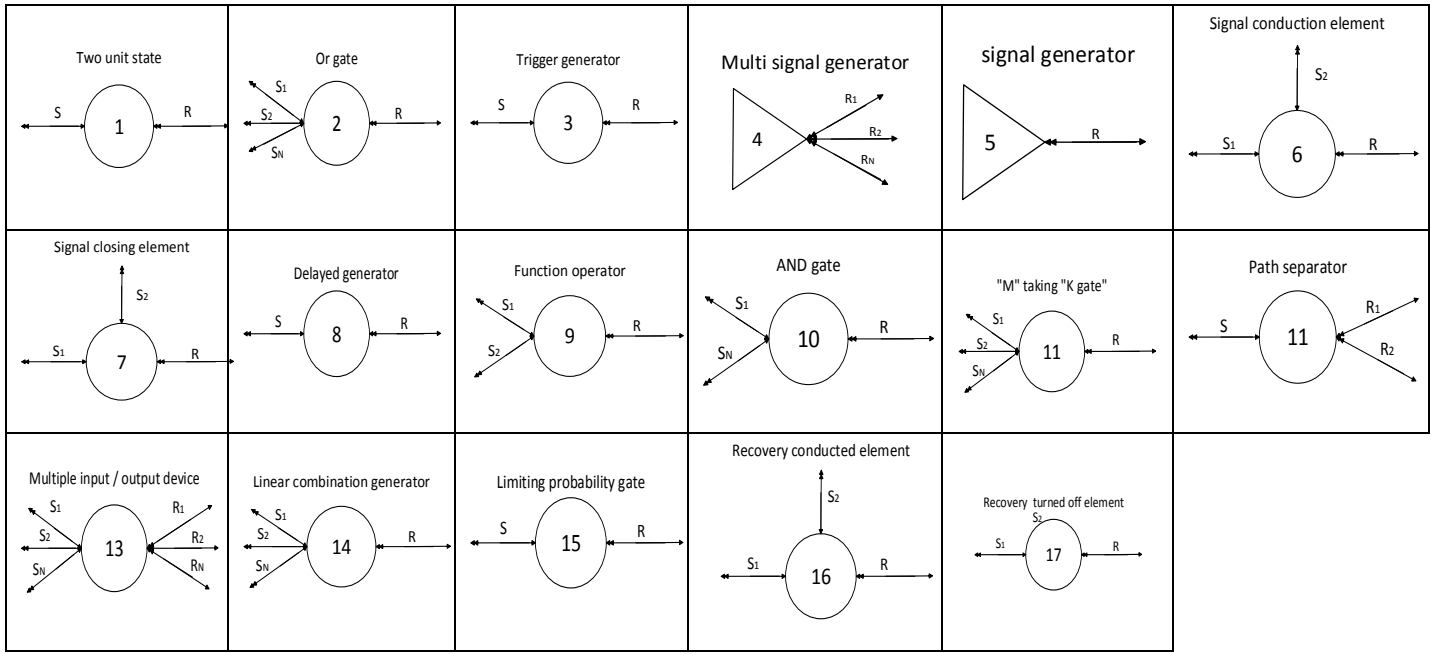


Figure 1. The operators used in the GO-Bayes method

### C. Probabilistic Inference based on GO-Bayes

System probabilistic safety evaluation can be realized in two ways. First, from components to systems, based on the probability parameters of component parts, we solve probability parameters of a system, such as normal work probability and failure probability. Second, from the system to the components, based on known system state information and component probability parameters, we reason a system's various safety status probabilities, i.e., "inverse probability".

### III. INTRODUCTION TO UNIT MODEL

When a basic unit is described, its probability data follows the following principles [15-16]. We use the following notation:  $S$  is the data unit subscript, like  $R_s$ ,  $F_s$  and  $P_s$ ;  $I$  is the input data subscript, like  $R_i$ ,  $F_i$  and  $P_i$ ; and  $O$  is the output data subscript, like  $R_o$ ,  $F_o$  and  $P_o$ .

#### A. Signal generating unit

A signal generating unit means an input to a system, external event or signal independent of the system. It can represent a generator, power, environmental impact and human factors. It has two states, normal or faulty. Its safety probability parameter comprises, Unreliability  $F(1)$ , inverse probability  $P(1)$ . Its single arrow output indicates an unreliability output, double arrow indicates an inverse probability input, satisfying:

$$F_o(1) = F_s(1) \quad (1)$$

$$P_s(1) = P_i(1) \quad (2)$$

Figure 2 means a signal generating unit model, and Figure 3 means a signal generator unit.

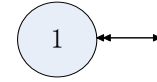


Figure 2. A signal generating unit model

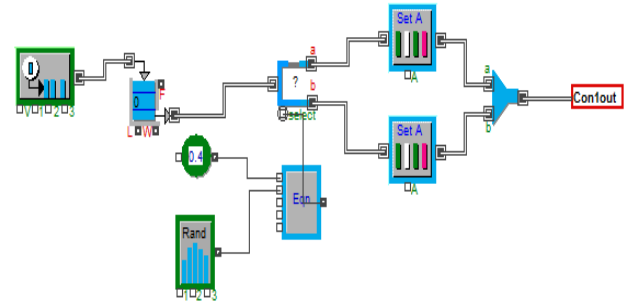


Figure 3. A Signal generator unit

#### B. Two state unit

As shown in Figures 4-5, a two-state unit is the most common unit, whose two states are normal and faulty ones. It has input and output data, and can represent resistors, switches, and valves. Its unreliability value is calculated based on the reliability theory,

$$F_o(2) = 1 - [1 - F_s(2)][1 - F_i(2)] \quad (3)$$

Two-state unit output failure results from either input fault or its own fault. They form a series logical relationship with the inverse probability

$$P_s(2) = \frac{F_s(2)P_i(2)}{F_o(2)} \quad (4)$$

$$P_o(2) = \frac{F_i(2)P_i(2)}{F_o(2)} \quad (5)$$

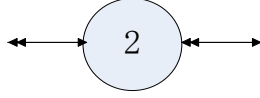


Figure 4. A Two-state unit model

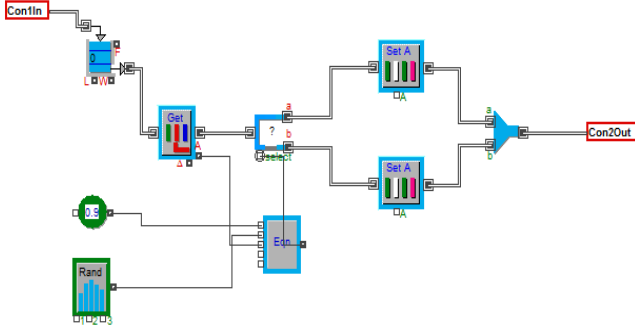


Figure 5. A two-state unit

### C. Conditional control unit

A conditional control unit, as shown in Figures 6-7, requires two inputs, the working status input, with a subscript label 1, and the control state input, with a subscript label 2. Its output represents their safety status. A conditional control unit may represent relay and mechanical control valves and so on. Its probability parameter calculation rules as follows:

$$F_o(3) = 1 - [1 - F_{i1}(3)][1 - F_{i2}(3)][1 - F_s(3)] \quad (6)$$

$$P_s(3) = \frac{F_s(3)P_i(3)}{F_o(3)} \quad (7)$$

$$P_{o1}(3) = \frac{F_{i1}(3)P_i(3)}{F_o(3)} \quad (8)$$

$$P_{o2}(3) = \frac{F_{i2}(3)P_i(3)}{F_o(3)} \quad (9)$$

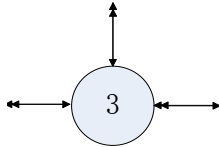


Figure 6. A Conditional control unit model

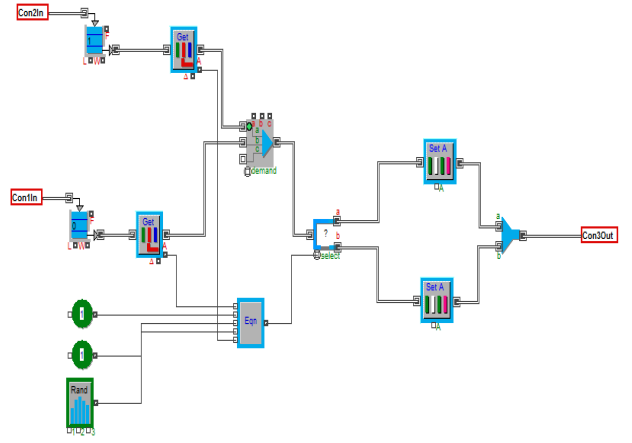


Figure 7. A Conditional control unit

### D. AND gate

An AND gate unit is shown in Figures 8-9. It can rely on several reliability input data items (with subscript labels being 1, 2, 3, ..., n), to compute one reliability output. It yields an output only when multiple inputs simultaneously are available. It does not have its own data. It does not stand for an internal system component, but is used to connect different units. Its reverser fault data is expressed as an input and multiple output. Obviously, an AND gate unit represents a parallel logical relationship. Its probability parameters can be computed:

$$F_o(4) = F_{i1}(4)F_{i2}(4)...F_{in}(4) \quad (10)$$

$$P_{o1}(4) = P_i(4) \quad (11)$$

$$P_{o2}(4) = P_i(4) \quad (12)$$

$$P_{on}(4) = P_i(4) \quad (13)$$

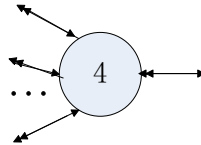


Figure 8. An AND gate unit model

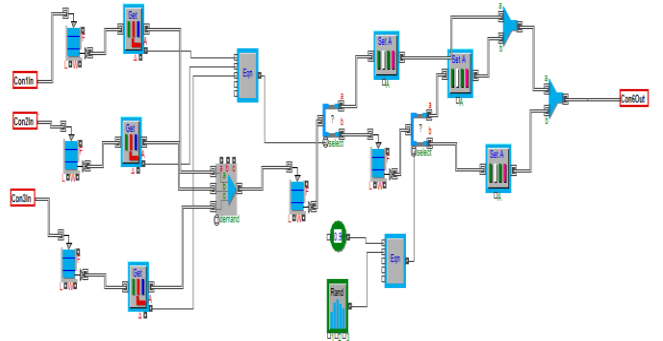


Figure 9. An AND gate unit



### E. OR gate

An OR gate unit relies on several reliability input data (with a subscript label 1, 2, 3, ..., n), to compute reliability output data, as shown in Figures 10-11. When one of the multiple inputs occurs, it can yield an output. It does not have its own data, and stand for no internal system component, but can be used to connect multiple units. Its reverseer fault data is expressed as an input and multiple outputs. Probability parameters are calculated as follows:

$$F_o(5) = 1 - [1 - F_{I1}(5)][1 - F_{I2}(5)] \dots [1 - F_{In}(5)] \quad (14)$$

$$P_{o1}(5) = \frac{F_{I1}(5)P_I(5)}{F_o(5)} \quad (15)$$

$$P_{o2}(5) = \frac{F_{I2}(5)P_I(5)}{F_o(5)} \quad (16)$$

$$P_{on}(5) = \frac{F_{In}(5)P_I(5)}{F_o(5)} \quad (17)$$

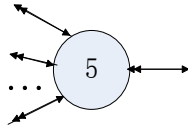


Figure 10. An OR gate unit model

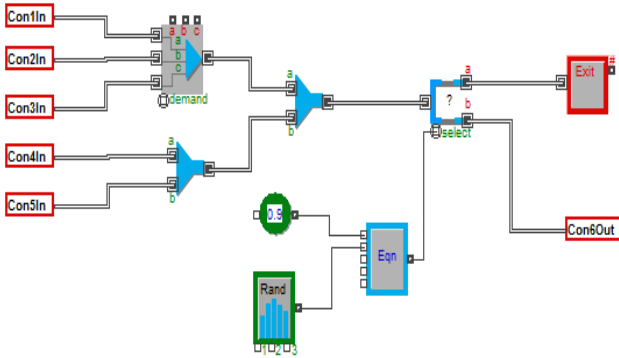


Figure 11. An OR gate unit

### F. Voting gate

A voting gate unit as shown in Figures 12-13 has several reliability input data items (with a subscript label 1, 2, 3, ..., n), and one output. It produces an output only when more than k inputs are present at the same time. It does not have its own data, and stands for no internal system component, but it can be used to connect multiple units. Its reverseer fault data is expressed as an input and multiple outputs. It represents a parallel and series logical relationship. It can be divided into a combination of AND gate units and OR gate units. For example taking 2 from 4 has  $C_4^2=6$  options, two AND gate units connect to one OR gate unit, meaning that two or more input failures lead to system output failure.

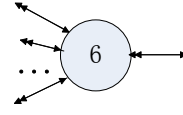


Figure 12. A Voting gate unit model

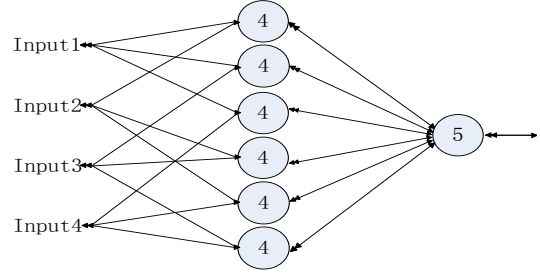


Figure 13. A 4-input series-parallel voting gate model

Its probability parameters can be derived from those for AND gate and OR gate units. We derive an algorithm for a GO-Bayes basic voting model.

Take 2 from 4 as an example in Figure 13. It can be divided into a combination of 6 AND gate units and 1 OR gate unit.

Assuming that the probabilities of inputs 1-4 are  $R_{I1}(6)=x_1$ ,  $R_{I2}(6)=x_2$ ,  $R_{I3}(6)=x_3$ ,  $R_{I4}(6)=x_4$ , we can derive the probability formula for the 2/4 voting gate  $F_o(6)$ ,

$$F_o(6) = 1 - [1 - (1 - x_1) \cdot (1 - x_2)] \cdot [1 - (1 - x_1) \cdot (1 - x_3)] \cdot [1 - (1 - x_1) \cdot (1 - x_4)] \cdot [1 - (1 - x_2) \cdot (1 - x_3)] \cdot [1 - (1 - x_2) \cdot (1 - x_4)] \cdot [1 - (1 - x_3) \cdot (1 - x_4)] \quad (18)$$

thus

$$F_o(6) = 1 - (x_1 \cdot x_2 \cdot x_3 + x_1 \cdot x_2 \cdot x_4 + x_1 \cdot x_3 \cdot x_4 + x_2 \cdot x_3 \cdot x_4 - 3x_1 \cdot x_2 \cdot x_3 \cdot x_4) \quad (19)$$

Since the jointly signal has no effect on the calculation of the reverse probability, according to the reverse probability formula of an AND gate unit (label 4) and OR gate unit (label 5), we can derive the following reverse probability of the gate:

$$P_I(1 \cdot 2) = \frac{F_I(1 \cdot 2)P_I(6)}{F_o(6)} = \frac{(1 - x_1) \cdot (1 - x_2) \cdot P_I(6)}{F_o(6)} \quad (20)$$

$$P_I(1 \cdot 3) = \frac{F_I(1 \cdot 3)P_I(6)}{F_o(6)} = \frac{(1 - x_1) \cdot (1 - x_3) \cdot P_I(6)}{F_o(6)} \quad (21)$$

$$P_I(1 \cdot 4) = \frac{F_I(1 \cdot 4)P_I(6)}{F_o(6)} = \frac{(1 - x_1) \cdot (1 - x_4) \cdot P_I(6)}{F_o(6)} \quad (22)$$

$$P_I(2 \cdot 3) = \frac{F_I(2 \cdot 3)P_I(6)}{F_o(6)} = \frac{(1 - x_2) \cdot (1 - x_3) \cdot P_I(6)}{F_o(6)} \quad (23)$$

$$P_i(2 \cdot 4) = \frac{F_i(2 \cdot 4)P_i(6)}{F_o(6)} = \frac{(1-x_2) \cdot (1-x_4) \cdot P_i(6)}{F_o(6)} \quad (24)$$

$$P_i(3 \cdot 4) = \frac{F_i(3 \cdot 4)P_i(6)}{F_o(6)} = \frac{(1-x_3) \cdot (1-x_4) \cdot P_i(6)}{F_o(6)} \quad (25)$$

$$P_{i1}(6) = \frac{F_{i1}(6)}{(1-x_1) \cdot (1-x_2)} \cdot \frac{(1-x_1) \cdot (1-x_2) \cdot P_i(6)}{F_o(6)} = \frac{F_{i1}(6) \cdot P_i(6)}{F_o(6)} \quad (26)$$

$$P_{i2}(6) = \frac{F_{i2}(6)}{(1-x_1) \cdot (1-x_2)} \cdot \frac{(1-x_1) \cdot (1-x_2) \cdot P_i(6)}{F_o(6)} = \frac{F_{i2}(6) \cdot P_i(6)}{F_o(6)} \quad (27)$$

$$P_{i3}(6) = \frac{F_{i3}(6)}{(1-x_1) \cdot (1-x_3)} \cdot \frac{(1-x_1) \cdot (1-x_3) \cdot P_i(6)}{F_o(6)} = \frac{F_{i3}(6) \cdot P_i(6)}{F_o(6)} \quad (28)$$

$$P_{i4}(6) = \frac{F_{i4}(6)}{(1-x_1) \cdot (1-x_4)} \cdot \frac{(1-x_1) \cdot (1-x_4) \cdot P_i(6)}{F_o(6)} = \frac{F_{i4}(6) \cdot P_i(6)}{F_o(6)} \quad (29)$$

Then we get a 2/4 vote gate algorithm. We can obtain the similar results for other voting gates.

#### IV. SAFETY ANALYSIS OF VISUAL UNIT METRO VEHICLES BRAKING SYSTEM

We now show how to use the proposed GO-Bayes method to analyze an urban rail transit vehicle air braking part.

##### A. Basic composition of air braking

Air braking portion of a braking system's basic components include air compressor and filtration device (as shown in A5 of Figure 14), total duct, air spring devices and pneumatic part (beginning with L in Figure 14), parking braking device (B7), braking control section (B13), braking airline (beginning with B), foundation brakes (beginning with C), and electronic anti-skid devices (beginning with G).

We build the visual system for an urban metro vehicle braking system as shown in Figure 14.

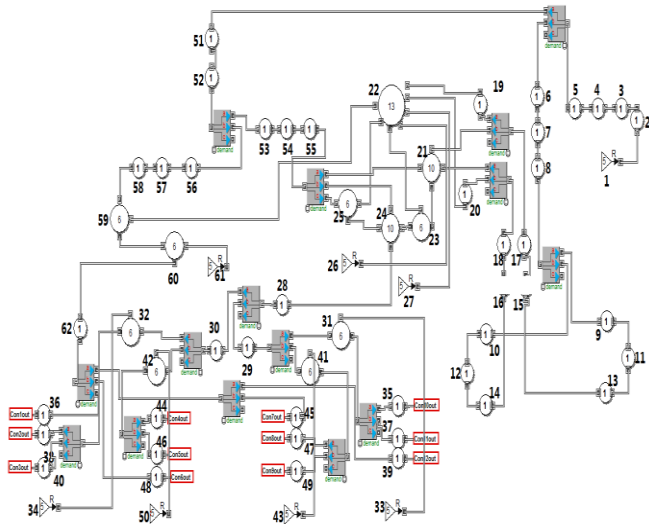


Figure 14. Visual system of an urban metro vehicle braking system

##### B. GO-Bayes modeling method for a braking system

The braking system has a complex structure and many components. In order to display and analyze it fully, this work uses a hierarchical modeling method by dividing the braking system into two layers. Its first layer has six functional sections as shown in Figure 15.

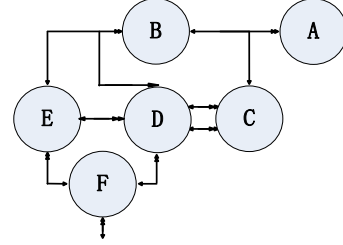


Figure 15. First layer structure model of a braking system

In Figure 15, node A represents an air supply device, B the line along which braking air passes through, C the air spring suspension, D the braking control device, E parking braking control, and F the foundation braking.

In the second layer of the model structure as shown in Figure 16, since the number of components is big, we label them according to the labels in the first parts and the position in the device. Numbers on the left of the dash represents unit types, and those on the right side correspond to the system unit.

- (1) An air supply device is shown in Figure 16 and Table 1.

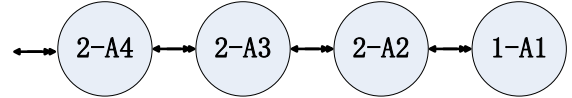


Figure 16. Structure model of an air supply device

TABLE 1 Units in an air supply device model

Code	Corresponding component
1-A1	Drive motor
2-A2	Air compressor
2-A3	Drying tower
2-A4	total air cylinder

- (2) A braking air route is given in Figure 17 and Table 2.

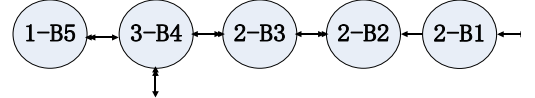


Figure 17. Braking air route structure model

TABLE 2 Components in a braking air route structure model

Code	Corresponding component
2-B1	Total air duct
2-B2	Cut-off valve
2-B3	Safety valve
3-B4	Braking reservoir cylinder
1-B5	Exhaust valve

(3) An air spring suspension device is shown in Figure 18 and Table 3.

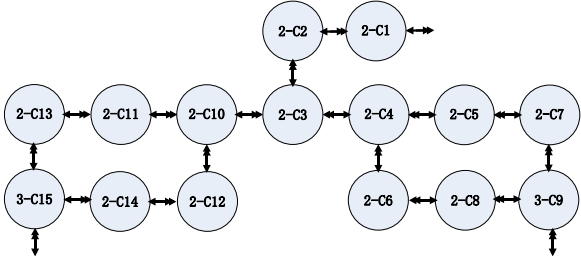


Figure 18. Structure model of an air spring suspension device

TABLE 3 Components in an air spring suspension device model

Code	Corresponding component
2-C1	Cut-off valve
2-C2	Filter
2-C3	Air spring cylinder
2-C4	Cut-off valve
2-C5	Left height valve
2-C6	Right height valve
2-C7	Air spring
2-C8	Air spring
3-C9	Pressure valve
2-C10	Cut-off valve
2-C11	Left height valve
2-C12	Right height valve
2-C13	Air spring
2-C14	Air spring
3-C15	Pressure valve

(4) A braking control device inner has its structure and components shows in Figure 19 and Table 4.

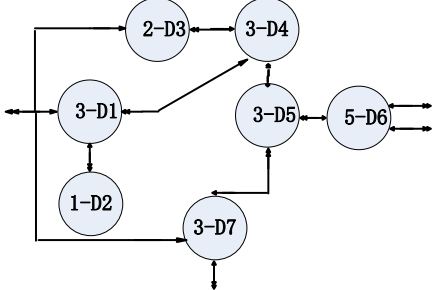


Figure 19. Structure model of a braking control system

TABLE 4 Components in a braking control system model

Code	Corresponding component
3-D1	Analog converter
1-D2	ECU code
2-D3	Emergency solenoid valve
3-D4	Pressure Switch
3-D5	Weighing valve
5-D6	OR gate
3-D7	Relay valve

(5) A parking braking device has its structure and components in Figure 20 and Table 5.

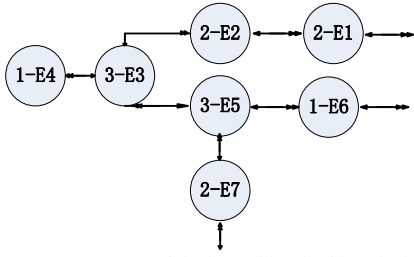


Figure 20. Structure model of a parking braking device

TABLE 5 Components in a braking device model

Code	Corresponding component
2-E1	Cut off valve
2-E2	Pressure Switch
3-E3	Parking braking solenoid valve
1-E4	Parking braking code
3-E5	Pulse valve
1-E6	Two-way valve
2-E7	Check

(6) A Foundation Braking is given in Figure 21 and Table 6.

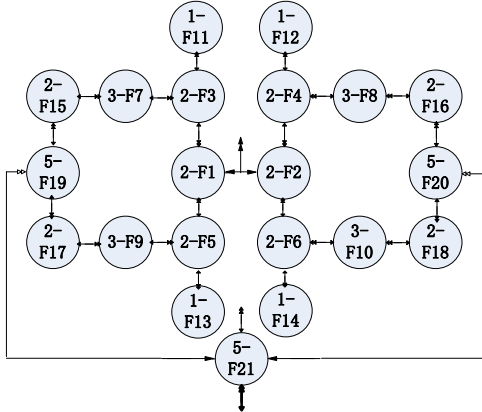


Figure 21. Structure model of a foundation braking device

TABLE 6. Components in a foundation braking device model

Code	Corresponding component
2-F1	Cut off valve
2-F2	Cut off valve
3-F3	Slip solenoid valve
3-F4	Slip solenoid valve
3-F5	Slip solenoid valve
3-F6	Slip solenoid valve
2-F7	Braking air reservoir
2-F8	Braking air reservoir
2-F9	Braking air reservoir
2-F10	Braking air reservoir
1-F11	Speed Sensor
1-F12	Speed Sensor
1-F13	Speed Sensor
1-F14	Speed Sensor
2-F15	Slipper
2-F16	Slipper
2-F17	Slipper
2-F18	Slipper
5-F19	Or gate
5-F20	Or gate
5-F21	Or gate

Another feature of the hierarchical model is that each of its modules can be individually analyzed. During the entire system analysis procedure, the correlations among modules have to be paid attention to.

### C. Calculation of probability indicators

Safety probability indicators are computed based on the GO-Bayes model of the braking system. First, obtain the fault parameters for each component by statistically analyzing historical operating cumulative data of the system. Component fault rate is the total number of the system failures divided by the total number of components, and then divided by the time duration. Secondly, we can calculate the component fault probability and normal work probability at time  $k$  according to the correlations among failure rate indices.

#### (1) Original data

Component failure rate data for each component comes mainly from the historical operation statistics. Assume steady-state operation 100h as shown in Table 7,

TABLE 7 Initial data of components in the model

Code	Failure Rate(10E-06/h)	Fault Rate
1-A1	3	2.9996E-04
2-A2	4.5	4.4990E-04
2-A3	6	5.9982E-04
2-A4	0.5	4.9999E-05
2-B1	1	9.9995E-05
2-B2	1.2	1.1999E-04
2-B3	2.5	2.4997E-04
3-B4	0.7	6.9998E-05
1-B5	8	7.9968E-04
2-C1	3	2.9996E-04
2-C2	1	9.9995E-05
2-C3	0.5	4.9999E-05
2-C4	3	2.9996E-04
2-C5	10	9.9950E-04
2-C6	10	9.9950E-04
2-C7	1.5	1.4999E-04
2-C8	1.5	1.4999E-04
3-C9	5	4.9988E-04
2-C10	3	2.9996E-04
2-C11	10	9.9950E-04
2-C12	10	9.9950E-04

2-C13	1.5	1.4999E-04
2-C14	1.5	1.4999E-04
3-C15	5	4.9988E-04
3-D1	2	1.9998E-04
1-D2	42	4.1912E-03
2-D3	9	8.9960E-04
3-D4	4	3.9992E-04
3-D5	0.8	7.9997E-05
3-D7	2	1.9998E-04
2-E1	3	2.9996E-04
2-E2	4	3.9992E-04
3-E3	3.5	3.4994E-04
1-E4	1	9.9995E-05
3-E5	3	2.9996E-04
1-E6	7	6.9976E-04
2-E7	1.2	1.1999E-04
2-F1	3	2.9996E-04
2-F2	3	2.9996E-04
3-F3	3.5	3.4994E-04
3-F4	3.5	3.4994E-04
3-F5	3.5	3.4994E-04
3-F6	3.5	3.4994E-04
2-F7	1	9.9995E-05
2-F8	1	9.9995E-05
2-F9	1	9.9995E-05
2-F10	1	9.9995E-05
1-F11	15	1.4989E-03
1-F12	15	1.4989E-03
1-F13	15	1.4989E-03
1-F14	15	1.4989E-03
2-F15	9	8.9960E-04
2-F16	9	8.9960E-04
2-F17	9	8.9960E-04
2-F18	9	8.9960E-04

#### (2) Calculation results

Based on the above model structure, we calculate the reliability and unreliability of each component's output, and then reverse reasoning to obtain each component's input probabilities. We can obtain the system output reliability that is 9.7079E-01, unreliability is 2.9205E-02. In Table 8, the inverse probability of the following components is relatively larger. That is to say, (1-D2, 1.4351E-01), (2-D3, 3.0803E-02), (1-F11, 1-F12, 1-F13, 1-F14, 5.1322E-02), (2-F15, 2-F16, 2-F17, 2-F18, 3.0803E-02) indicate the component fault will most likely lead to system fault. Hence, we have to focus on tracking them.

TABLE 8 Safety analysis results of braking system

Code	Output unreliability (Cumulative probability of fault)	Output reliability (Normal work probability)	Component input inverse Probability	Component inverse Probability
1-A1	2.9996E-04	9.9970E-01	1.0271E-02	1.0271E-02
2-A2	7.4972E-04	9.9925E-01	2.5671E-02	1.5405E-02
2-A3	1.3491E-03	9.9865E-01	4.6194E-02	2.0538E-02
2-A4	1.3990E-03	9.9860E-01	4.7903E-02	1.7120E-03
2-B1	1.4989E-03	9.9850E-01	5.1322E-02	3.4239E-03
2-B2	1.6187E-03	9.9838E-01	5.5425E-02	4.1086E-03
2-B3	1.8683E-03	9.9813E-01	6.3970E-02	8.5591E-03
3-B4	2.7362E-03	9.9726E-01	9.3691E-02	2.3968E-03
1-B5	7.9968E-04	9.9920E-01	2.7382E-02	2.7382E-02
2-C1	1.6986E-03	9.9830E-01	5.8160E-02	1.0271E-02
2-C2	1.7984E-03	9.9820E-01	6.1578E-02	3.4239E-03
2-C3	1.8483E-03	9.9815E-01	6.3287E-02	1.7120E-03
2-C4	2.1477E-03	9.9785E-01	7.3538E-02	1.0271E-02
2-C5	3.1450E-03	9.9685E-01	1.0769E-01	3.4224E-02
2-C6	3.1450E-03	9.9685E-01	1.0769E-01	3.4224E-02
2-C7, 2-C8	4.1414E-03	9.9586E-01	1.4180E-01	5.1357E-03

3-C9	4.9378E-03	9.9506E-01	1.6907E-01	1.7116E-02
2-C10	2.1477E-03	9.9785E-01	7.3538E-02	1.0271E-02
2-C11, 2-C12	3.1450E-03	9.9685E-01	1.0769E-01	3.4224E-02
2-C13, 2-C14	4.1414E-03	9.9586E-01	1.4180E-01	5.1357E-03
3-C15	4.9378E-03	9.9506E-01	1.6907E-01	1.7116E-02
3-D1	7.1146E-03	9.9289E-01	2.4361E-01	6.8474E-03
1-D2	4.1912E-03	9.9581E-01	1.4351E-01	1.4351E-01
2-D3	3.6334E-03	9.9637E-01	1.2441E-01	3.0803E-02
3-D4	8.0078E-03	9.9199E-01	2.7419E-01	1.3694E-02
3-D5	1.5056E-02	9.8494E-01	5.1551E-01	2.7391E-03
5-D6	6.6279E-03	9.9337E-01	2.2694E-01	0
3-D7	1.5252E-02	9.8475E-01	5.2226E-01	6.8474E-03
2-E1	3.0354E-03	9.9696E-01	1.0393E-01	1.0271E-02
2-E2	3.4341E-03	9.9657E-01	1.1759E-01	1.3694E-02
3-E3	3.8824E-03	9.9612E-01	1.3294E-01	1.1982E-02
1-E4	9.9995E-05	9.9990E-01	3.4239E-03	3.4239E-03
3-E5	1.7367E-02	9.8263E-01	5.9467E-01	1.0271E-02
1-E6	1.5942E-02	9.8406E-01	5.4585E-01	2.3960E-02
2-E7	1.7485E-02	9.8251E-01	5.9871E-01	4.1086E-03
2-F1,2-F2	1.7780E-02	9.8222E-01	6.0880E-01	1.0271E-02
3-F3, 3-F4, 3-F5, 3-F6	1.9595E-02	9.8040E-01	6.7096E-01	1.1982E-02
2-F7, 2-F8, 2-F9, 2-F10	1.9693E-02	9.8031E-01	6.7432E-01	3.4239E-03
1-F11, 1-F12, 1-F13, 1-F14	1.4989E-03	9.9850E-01	5.1322E-02	5.1322E-02
2-F15, 2-F16, 2-F17, 2-F18	2.0575E-02	9.7942E-01	7.0451E-01	3.0803E-02
5-F19	2.3363E-02	9.7664E-01	7.9996E-01	0
5-F20	2.3363E-02	9.7664E-01	7.9996E-01	0
5-F21	2.9205E-02	9.7079E-01	1.0000E+00	0

#### D. Experimental Analysis

We can conclude from the above safety analysis:

(1) When a system shows abnormal conditions, we have to obtain real-time inverse probability through the fault backward reasoning method. The inverse probability of components (3-B4, 1.3070E-02), (1-B5, 1.4932E-01), (3-D1, 3.7340E-02), (1-D2, 7.8258E-01), and (1-F11, 4.7087E-01) is significantly larger than the others', which shows that these parts may be abnormal. We should thus focus on tracking them. In addition by using the system diagram model to analyze 3-B4, 1-B5, 3-D1, and 1-D2, which are working parts connected together, the abnormal output of 3-D1 indicates that the failure possibility of these four components is very large, and the failure possibility of (1-D2, 7.8258E-01) is the highest. It represents Electronic Control Unit instruction, error rate of which is

higher, because it has many electronic circuit components. While (1-F11, 4.7087E-01) is an independent failure, in fact, it represents the speed sensor with a self-resetting function. Its false detection occurs frequently. If an abnormal event is detected when its probability of failure is less than 1/2, we should check and maintain them.

(2) Traditional fault probability calculation depends on the forward deduction of historic data. By contrast, the GO-Bayes method provides structural models of a system and inverse reasoning probability. The models' output and inverse probability reflect more accurately the system's reliability than traditional fault probability. Figure 22 is a metro train's braking system based on FTA. Table 9 shows GO-Bayes' advantage compared with FTA.

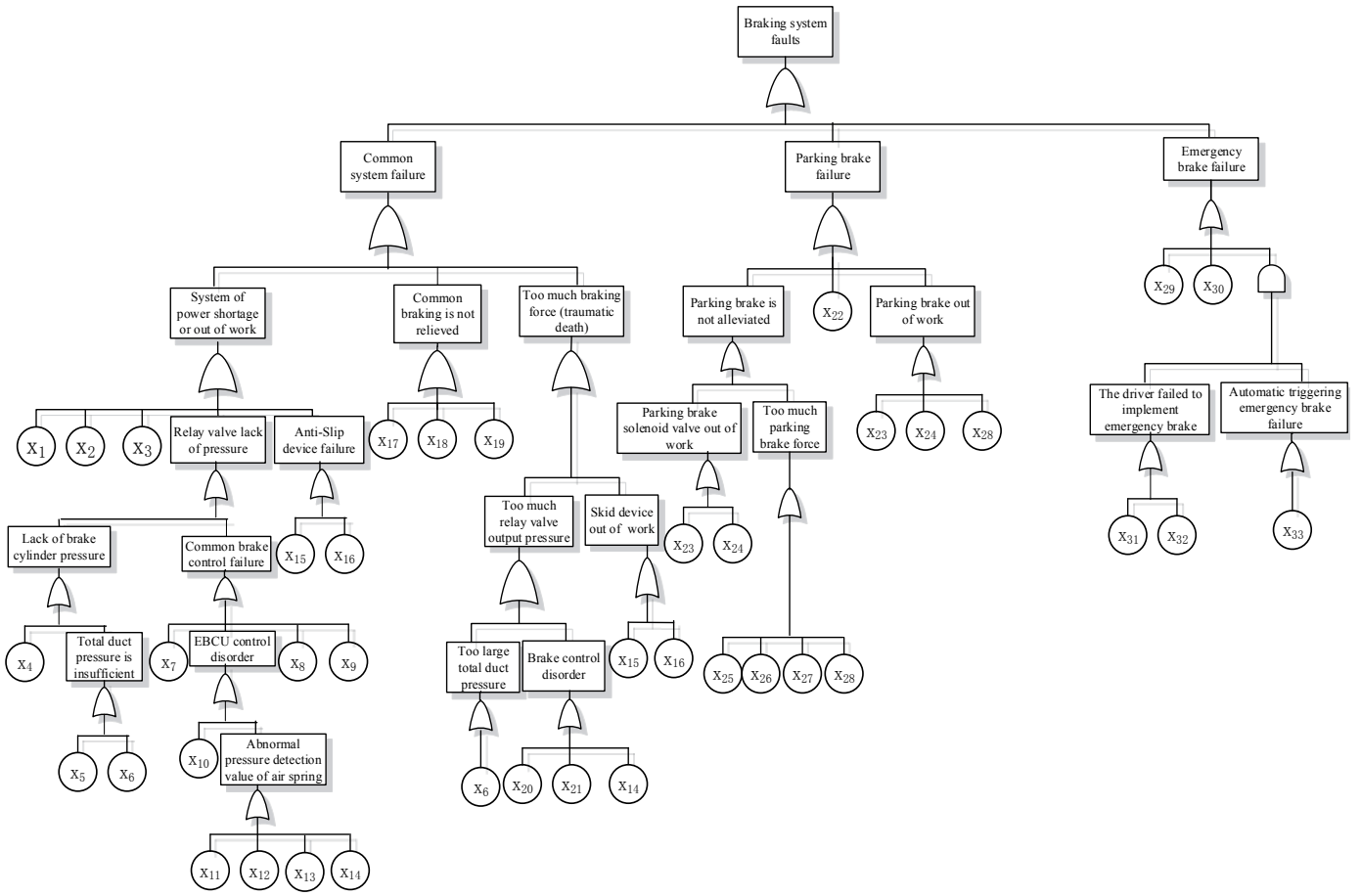


Figure 22 A metro train's braking system based on FTA

TABLE 9 GO-Bayes compared with FTA

Model Feature	GO- Bayes	FTA
Modeling oriented	success	Failure
Modeling method	bi-direction induction	deductive
Modeling consistency	basically identical	differences according to everyone's understanding
Structure	similar principle diagram	hierarchical logic diagram
Volume	compact, small size	multi-layer, huge volume
Elements	component, logic diagram	fault event, logic gate
Description	reflect original system structure	reflect the failure cause and effect
Notation	more operators with rich expression	less operators with poor expression

## V. CONCLUSION

This paper presents a new structural GO-Bayes method. It is a comprehensive system safety analysis and evaluation modeling methodology. Using a system diagram model, we can obtain the system's normal work probability output, which is essential for fault backward reasoning. The paper discusses basic components or units and their related analysis results. The application of the proposed method to a metro vehicle braking system shows its contribution to safety analysis and assessment. The results can be used to trace, maintain and improve system components and eventually ensure the entire system's safe operation.

## ACKNOWLEDGMENT

This work was supported by China High Technologies Research Program (2015BAG19B02) (KIK15007531).

## REFERENCES

- [1] Yang, J.; Wu, Z.Z.; Chen, T.Z. Review of the urban road traffic safety evaluation methods. Proceedings of the Fourth International Conference on Transportation Engineering, December 2013, pp.2503-2508.
- [2] Pugel, A.E.; Simianu, V.V.; Flum, D.R. Use of the surgical safety checklist to improve communication and reduce complications. Journal of infection and public health, 2015, 8(3), pp.219-225.
- [3] Long, J.B.; Birmingham, P.K.; De Oliveira, Jr.G.S. Transversus Abdominis Plane Block in Children: A Multicenter Safety Analysis of 1994 Cases From the PRAN (Pediatric Regional Anesthesia Network) Database. Survey of Anesthesiology, 2015, 59(3), pp.139-140.
- [4] Saunders, R.P.; Wilcox, S.; Baruth, M. Process evaluation methods, implementation fidelity results and relationship to physical activity and healthy eating in the Faith, Activity, and Nutrition (FAN) study. Evaluation and program planning, 2014, pp.43: 93-102.
- [5] Ibáñez, L.; Hortal, J.; Queral, C. Application of the Integrated Safety Assessment Methodology to Safety Margins. Dynamic Event Trees,



- Damage Domains and Risk Assessment. Reliability Engineering & System Safety, 2015.
- [6] Purba, J.H.; Lu, J.; Zhang, G. A fuzzy reliability assessment of basic events of fault trees through qualitative data processing. Fuzzy Sets and Systems, 2014, 243, pp.50-69.
  - [7] Appelbaum, P.S.; Robbins, P.C.; Monahan, J. Violence and delusions: Data from the MacArthur violence risk assessment study. American Journal of Psychiatry, 2015.
  - [8] Zhou, L.M.; Cai, G.Q.; Yang, J.W.; Jia, L.M. Monte-Carlo Simulation Based on FTA in Reliability Analysis of Gate System. The 2nd International Conference on Computer and Automation Engineering. February 26-28, 2010, pp.713-717.
  - [9] Xu, W.Z. Comparative study of the GO method and fault tree reliability modeling. Journal of Beijing institute of light industry, 1999, 17(2).
  - [10] Shen, Z.P.; Huang, R.X. Z. Shen, R. Huang. Principle and application of GO - a method for reliability analysis of system. Beijing: Tsinghua University Press, 2008.
  - [11] Takkishi, M.; Michiyuki, K. GO-FLOW: A New Reliability Analysis Methodology. Nuclear science and engineering, January 1988, 98(1), pp.64-78.
  - [12] Xie, B.L.; Liu, Q. A New Inference Method on Fuzzy Control and Simulation. 2011 International Conference on Mechanical, Industrial, and Manufacturing Engineering (MIME 2011). January 2011, pp.159-161.
  - [13] Cooper, G.F.; Herskovite, E. A bayesian method for the induction of probabilistic networks from data[J]. Machine learning, 1992, 9(4): pp.309-347.
  - [14] Holmes, C.C.; Mallick, B.K. Bayesian regression with multivariate linear splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2001, 63(1), pp.3-17.
  - [15] Wang, H.B.; Wang, Z.W. Study on the method and procedure of logistics system modeling and simulation. 2nd International Conference on Science and Social Research (ICSSR 2013), June 2013, pp.776-780.
  - [16] Qin, T.H.; Wang, Y.F. Application oriented simulation modeling and analysis with ExtendSim. Beijing Tsinghua University Press, 2009.

# Testing a Storytelling Tool for Digital Humanities

Fabio Pittarello

Università Ca' Foscari Venezia  
Via Torino 155 - Venezia, Italia  
pitt@unive.it

**Abstract**—This work presents an evaluation of ToBoA-3D, a social web platform for the annotation of 3D models and the creation of stories. ToBoA-3D was evaluated through a pilot study realized in a real educational context, the 2014 edition of the Fall School in Digital Humanities, held at the Università Ca' Foscari Venezia in collaboration with the École Polytechnique Fédérale de Lausanne (EPFL). The tool was presented during the School's lectures and then used by the students for creating stories during the labs' activities. The results of the study were collected through direct observation and a questionnaire. The answers evidenced positive feedbacks for the core features of the platform and led to define an initial set of guidelines for its educational use.

*Keywords:* annotation; cultural heritage; digital humanities; education; pilot study; storytelling; web3d

## I. INTRODUCTION

The availability of 3D representations for scholars is a great opportunity to support the processes of teaching and learning, especially for those disciplines that are deeply involved in the study of objects that have a 3D shape. A further exploitation of 3D objects can be obtained by their annotation that permits, for example, to search annotated content across a set of different 3D worlds, overcoming the navigation mechanisms provided by the authors of each model. ToBoA-3D [1] is a social web platform that permits to exploit the educational potential of annotated 3D worlds. It can be used for collaborative annotation, navigation and search of 3D worlds; the users can even add new 3D environments to the web repository and share them with the other users. The latest evolution of the platform [2] introduces the possibility to create educational stories on the top of the annotated worlds and to share them on the web. While ToBoA-3D can be personalized for different knowledge domains, so far the development has been focused on art and architecture. In this work we describe how ToBoA-3D was tested in a real educational context, a necessary step for evidencing its points of strength and weaknesses and designing more extensive educational experiences. The occasion for the pilot study came from the Fall School in Digital Humanities, held in Venice in October 2014, in the context of the collaboration between the Università Ca' Foscari Venezia and the EPFL (École Polytechnique Fédérale de Lausanne). Digital Humanities are an area of research and teaching at the intersection of computing and humanistic disciplines. This area combines the methodologies of the traditional humanities with tools and methods provided by computer science. The faculty of the Venice Fall School was compliant with this intersection, being composed by art and architecture historians but also computer scientists, which transmitted to the students

complementary knowledge about the School's educational themes. The School was focused on the Venetian Renaissance and included class lessons, visits to historical sites and labs. About 20 students, mainly PhD students in Humanities, were selected for participating to the School, programmed for a full week. During this week a group of students had the opportunity to learn how to use ToBoA-3D and to apply this knowledge to the School's themes. The results of the experience were collected through direct observation and a questionnaire and led to define a set of guidelines for the future use of ToBoA-3D.

## II. RELATED WORKS

The related literature encompasses different fields, from annotation of 3D worlds to storytelling. Because of space limits we'll give only some hints, addressing to [1] [2] for further references. As far as annotation is concerned, there are different proposals for adding high-level semantics to the components of 3D environments, based on specifications such as MPEG-7, Collada or X3D. Most proposals use annotations referred to taxonomies or ontologies, but there are also systems that permit to use free tags [1]. The latter ones offer a more expressive and informal approach suited also to common people. Their benefits for cultural heritage have already been underlined in [3]. Annotation is however a first step towards a more advanced use of 3D worlds. Scopigno et al. [4] underline that the greater challenge for digital technologies is the creation of tools that use 3D models for supporting cultural heritage research. We claim that the introduction of storytelling for annotated 3D worlds can bring further advantages for researchers and pupils. Computer-enhanced storytelling represents an evolution of traditional storytelling. It takes advantage of technology for creating and delivering stories, but also for designing and managing new narrative models and proposing new relations between the narration, the reader and the context. The analysis of narratology, the science that studies the structure of the story, has greatly helped the building of models and architectures for interactive storytelling [5]. Several studies have demonstrated the usefulness of storytelling for educational experiences [6]. In the cultural heritage domain storytelling is used for engaging students during the learning process, associating the narration to multimedia components [7], real scenes [8], augmented [9] and virtual reality [10]. While other tools offer repositories of assets for speeding the creation of stories, ToBoA-3D fully exploits the potential of collaboration and knowledge sharing.

### III. THE ToBoA-3D PLATFORM

Each user can contribute to the ToBoA-3D platform in different fashions, uploading 3D models, annotating the components that define the 3D environments or building narrations. All the 3D environments belonging to the platform can be explored both using a first-person navigation style or querying the system for retrieving interesting objects and places contained inside of them. Although users are not required to perform all the types of activities, a typical session with ToBoA-3D includes a mix of them. As far as the creation of narrations is concerned, the author creates a story starting from a list of personal bookmarks, corresponding to visited 3D viewpoints related to annotated objects and locations, and selects an ordered set of them for defining the stages of a linear story. Each stage is then associated to multimedia content that will be automatically played during the narration. Additional content, such as textual descriptions, images and links to web resources, can be included. The interface for listening to stories is displayed in Fig.1. The story is automatically played synchronizing the delivery of the information associated to each stage with the automatic navigation to the associated 3D locations. If the story is narrated across different environments belonging to the platform, the system takes care of downloading automatically the required environment. The story can be stopped at any time, for allowing the listener to focus on details, exploring the 3D view or accessing the associated hypermedia. Stories can be navigated also selecting the single stage from the list available on the lower panel of the interface. Two key features of ToBoA-3D are that the results of all the activities are shared and that their authors are easily identifiable. The first feature enables all the users to take advantage of the work done by the other members of the community (e.g., upload of 3D models, annotations), avoiding to start from scratch for authoring. The second feature offers interesting scenarios for research and teaching. For example ToBoA-3D permits students to annotate a set of 3D architecture components after a lesson focused on classical orders and then the teacher to check the annotations made by each student, marked with the student's ID. Further details about the ToBoA-3D functionalities are available in [2].

### IV. ToBoA-3D AT THE FALL SCHOOL

The goal of the Fall School was to investigate, through a set of coordinated lectures and visits to Venetian palaces and collections, new ways of visualizing the evolution of architecture and artwork display. The School was focused in particular on investigating the evolution of the Grimani Palace in Venice and of the artwork collection contained in its main room, the so-called Tribuna. During the preparatory work we discussed with the other organizers how to present and use the annotation platform during the School's activities. A skilled 3D modeler created a simplified model of the Grimani Palace and its Tribuna that was used as the scenario for an educational narration built with ToBoA-3D. The narration was created with the contribution of Cristiano Gueneri, one of the architecture historians involved in the School. It was organized

as a self-paced linear story, guiding the students with a virtual camera through the locations of the palace. The first part of the School included lectures held by art historians and focused on the main School's theme, but also talks related to the use of new technologies for cultural heritage, among which ToBoA-3D. The last two days of the school were dedicated to technological labs, held in parallel for small groups of students. For this reason, only four students out of twenty had the chance to attend the ToBoA-3D lab. In spite of the low number of students, the results were interesting. The initial phase of the ToBoA-3D lab was dedicated to a tutorial illustrating the different features of the platform. The students were then invited to try the techniques acquired on some test 3D environments prepared for the School. Finally the students listened to the introductory story by the art historian that ended with the presentation of two tasks to accomplish:

- the creation of a narration describing a tour through the rooms of the palace, enriched with the snapshots taken during the visit to the real building;
- the creation of a story related to an hypothesis of reallocation of the artworks of the Tribuna Grimani, which are currently placed in a different site.

The students interpreted both the themes proposed, although with some simplifications due to time constraints. The first story was an humorous interpretation where one of the students played the part of a Venetian nobleman and guided the listeners through the rooms of the palace (see Fig.1 on the left). The second story was a more serious narration, focused on the hypothesis of reallocation of the artworks in the Tribuna Grimani (see Fig.1 on the right; the red dots in the 3D scene represent the original position of the artworks).

### V. RESULTS OF THE PILOT STUDY

The four students were PhD candidates representatives of different research domains: visual arts and architecture (two students), computer science and interaction design, literature. While we should not consider this study as exhaustive, their answers have a great value for identifying the points of weakness and strength of the platform from different facets. Only two of them (the computer scientist and one of the art historians) had a fair knowledge of modeling techniques and interactive 3D environments. None of them had previous experience related to the annotation of 3D worlds. The results of the pilot study were collected through discussions with students and a final questionnaire, composed of closed and open questions, and articulated in 6 sections focused on annotation, search, storytelling, sharing, workflow and usability/engagement. We obtained positive results for most of the features of ToBoA-3D. While we don't have the space to analyze the single results, we underline how the pilot study led to define more precisely the profile of the students interested in using this platform and the set of tasks that should define a complete experience. These results will be useful for designing a more advanced study or more extensive and complete educational experiences based ToBoA-3D:

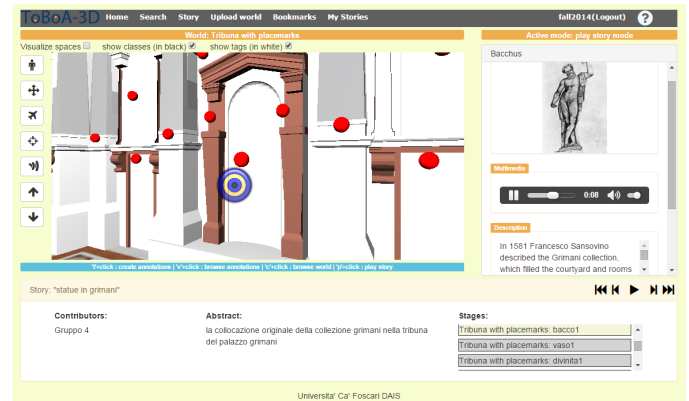
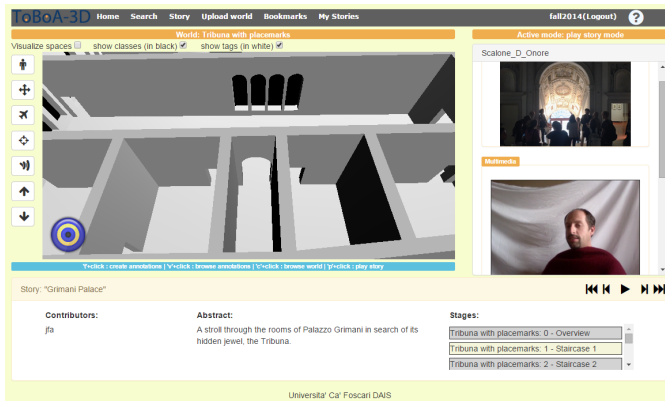


Fig. 1. The two stories created by the students at the Digital Humanities Venice Fall School

- *design educational experiences for students interested in visual representation*; the study revealed that ToBoA-3D resulted more appealing to students with a background focused on visual representation rather than on literature; we suggest, as a complementary guideline, that for longer term experiences such as a master course, the introduction of a basic 3D modeling course would be useful for giving a basic knowledge of 3D representation and attracting students with different cultural backgrounds;
- *include search and information sharing among the activities of the educational experience*; in the pilot study students were shown the functionalities related to search and information sharing and declared their interest for them; however, starting also from the observations of the students, we estimate that it would be interesting to include specific tasks focused on these activities in the structure of the educational experience, proposing for example to search an annotated set of 3D worlds in relation to a given goal or to ask a group students to annotate individually the same set of objects and then check the annotations of their fellows for identifying different point of view; probably these activities would augment the awareness of the potential of these techniques and would stimulate their application to individual research;
- *include knowledge checking as part of the educational experience*; while the educational experience proposed in this experience included a set of goals defined by a teacher, for time constraints it was not possible to fulfill all the teacher's requests and to check the results; a full educational experience should include a final check by the teacher of the individual work and a feedback to students.

Other suggestions came from the feedback related to the ToBoA-3D functionalities. While we obtained positive judgments for the core features of ToBoA-3D, most of the students complained about the quality of the 3D models, realized under heavy time constraints. This is an issue that we'll take into account for the future educational experiences. The students suggested improvements to the platform as well. While the students with an humanistic background focused more on content and structure, suggesting for example the possibility to

create stories with branching structures, the computer science student focused on interaction design issues, suggesting ways to refine the interface or adding additional functionalities. While the positive findings encourage us to propose the use of the platform in real educational contexts, the future development will take care of all the points of weaknesses underlined by the users and improve further its features.

#### ACKNOWLEDGMENTS

Thanks to Ivano Gatto for all the contributions given to the development of the ToBoA-3D platform. I acknowledge Frédéric Kaplan and Isabella Di Lenardo (EPFL), which co-organized the Fall School in Digital Humanities, and Cristiano Guarneri (Università IUAV), which gave a great contribution for the educational story for the students.

#### REFERENCES

- [1] 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.
- [2] I. Gatto and F. Pittarello, "Creating web3d educational stories from crowdsourced annotations," *Journal of Visual Languages & Computing*, vol. 25, no. 6, pp. 808–817, 2014.
- [3] 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.
- [4] 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.
- [5] M. Cavazza and D. Pizzi, "Narratology for interactive storytelling: A critical introduction," in *Proc. of TIDSE'06*. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 72–83.
- [6] J. Ohler, *Digital Storytelling in the Classroom: New Media Pathways to Literacy, Learning, and Creativity*. Corwin Press, 2008.
- [7] 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.
- [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*, 2006, pp. 17–24.
- [9] Z. Zhou, A. D. Cheok, J. Pan, and Y. Li, "An interactive 3d exploration narrative interface for storytelling," in *Proc. of IDC '04*. New York, NY, USA: ACM, 2004, pp. 155–156.
- [10] S. Mystakidis, N. Lambropoulos, H. M. Fardoun, and D. M. Alghazzawi, "Playful blended digital storytelling in 3d immersive learning environments: A cost effective early literacy motivation method," in *Proc. of IDEE '14*. New York, NY, USA: ACM, 2014, pp. 97–101.

# A Quick Survey on Sentiment Analysis Techniques: a lexical based perspective

Flora Amato<sup>1</sup>, Francesco Colace<sup>2</sup>, Luca Greco<sup>2</sup>, Vincenzo Moscato<sup>1</sup>, Antonio Picariello<sup>1</sup>

<sup>1</sup>DIETI- Department of Electrical Engineering and Information Technology

*Università degli Studi di Napoli "Federico II", Napoli - Italy*

{flora.amato, vmoscato, picus}@unina.it

<sup>2</sup>DIEM - Department of Information Engineering, Electrical Engineering and Applied Mathematics

*Università degli Studi di Salerno, Fisciano (Salerno) - Italy*

{fcolace,lgreco}@unisa.it

**Abstract—** With the advent of *World Wide Web* and the widespread of on-line collaborative tools, there is a increasing interest towards automatic tools for *Sentiment Analysis* to provide a quantitative measure of “positivity” or “negativity” about opinions or social comments. In this paper, we provide an overview of the most diffused techniques for sentiment analysis based on the lexical-based approaches as a quick reference guide in the choice of the most suitable methods for solving a specific problem in the sentiment analysis field.

**Index Terms—** Sentiment analysis, Computational linguistics, Text Classification .

## I. INTRODUCTION

People’s opinion has always driven human choices and behaviors, even before the diffusion of Information and Communication Technologies. With the advent of *World Wide Web* and the widespread of on-line collaborative tools such as blogs, focus groups, review web sites, forums, social networks (e.g *Facebook*, *Twitter*, *MySpace*, etc.), users more and more use to make available to everyone their tastes and liking, and in general, their opinions and sentiments about an event, a topic, a public person, a political faction, a TV program, etc.

In such a context, there is an increasing need to have available automatic tools for *Sentiment Analysis* (or *Opinion Mining*) and *Tracking* in order to provide a quantitative measure of “positivity” or “negativity about opinions (*polarity*) or comments related to a certain topic of interest and to track along the time such information.

More in details, sentiment analysis aims at finding the opinions of authors (thought leaders and ordinary people) about specific entities, by analyzing a large number of documents (in any format such as PDF, HTML, XML, etc.).

It can be considered as a sub-discipline of Computational Linguistics, indeed it is a *Natural Language Processing* and *Information Extraction* task [14], or challenged by the use of classical *Machine Learning* based approaches.

The most studied languages in the opinion mining field are English and Chinese, but there are several researches on other languages like Italian, Thai and Arabic [12].

Opinion mining allows to identify problems by listening, rather than by asking, ensuring an accurate reflection of reality [14].

The analyzed textual information can be of two types: *facts* and *opinions*. The facts are objective expressions that describe entities, conversely the opinions deal with people’s emotions, sentiments and feelings and so they are subjective.

Generally, we can see an opinion (or a sentiment) as a quintuple:  $\langle o, f, s, h, t \rangle$ , where  $o$  is the object evaluated by the opinion holder  $h$ ,  $f$  is a feature of the object  $o$ ,  $t$  is the time when the opinion has been expressed and  $s$  is the value of the opinion (for example positive or negative) [1][14].

Sentiment analysis techniques have as main goal the automatic extraction of the polarity measure “attached” to an object and can adopt several methods and techniques derived both from Computational Linguistics and Machine Learning theory. Here, we focus our attention on *lexical-based* techniques belonging to the branch of Computational Linguistics approaches.

The paper is organized as follows. Section II contains a review of the most diffused lexical-based approaches. Finally, Section III reports some conclusions and final considerations about our study.

## II. AN OVERVIEW OF LEXICAL BASED SENTIMENT ANALYSIS TECHNIQUES

In lexical-based approach a predefined list of words is used to determine a specific sentiment. A relevant problem regards ambiguity of natural language: sentiment value for a given word depends on the specific context.

There are several approaches to sentiment lexicons’ creation. A manual construction is often difficult and very time consuming. In the literature, the most used methods can be classified as Corpus-based and Dictionary-based.

### i. Corpus-based Approach

In this approach, a set of seed words grows by using a corpus of documents of a specific domain. Therefore a specific domain lexicon is constructed on the basis of a labeled corpus of documents.

One of the first works in this field is [6] where, given some seed adjectives, a corpus is used to identify additional sentiment adjectives. A key point regards the presence of conjunctions: for example the conjunction ‘and’ between two adjectives can refer to the same sentimental polarity. A graph with same or different orientation links between adjectives is created. These adjectives are then separate with a clustering algorithm into two subsets.

Another example is [8] where a corpus of 10000 blog posts from LiveJournal.com is used; the posts are labeled “happy” or “sad”. A happiness factor is assigned to words by calculating their frequency: the ratio between the number of occurrences of a word in the happy blogposts and its frequency in the entire corpus.

Among the most recent studies there is the work in [4]. The key of this approach is searching the connotative polarity between a conative predicate and its semantic argument. It is done by using a graph-based algorithm that use PageRank [9] and HITS [7] that collectively learn connotation lexicon together with connotative predicates.

### ii. Dictionary-Based Approach

In this approach a small set of seed words is first manually collected and then is expanded with words synonyms and antonyms. This is done by using online resources (dictionaries). The most well-known example is *Wordnet* that is an online lexical database for English language.

A great disadvantage of this approach is that the lexicon acquired is independent from a specific domain.

#### ➤ *WordNet-Affect*

WordNet-Affect [11] is a linguistic resource, composed by 2,874 synsets and 4,787 words, developed considering WordNet Domains, that is a multilingual extension of Wordnet.

It aims at providing correlations between affective concepts and affective words by using a synset model.

A subset of synsets, which are able to represent affective concepts, is derived from WordNet. Then, these synsets are labeled with one or more affective categories.

The Core of WordNet Affect is created by considering a lexical database, called Affect, composed by 1,903 words that are mostly adjectives and nouns.

Lexical and affective information are associated to each term; they includes parts of speech, definitions, synonyms and antonyms.

In order to assign an affective category to terms, an attribute called Ortony is used. Terms can be classified in emotional terms, non-emotional affective terms, non-affective mental state terms, personality traits, behaviors, attitudes etc.

Ortony information is projected on the subset selected from Wordnet but doesn't cover all Affect items and for this reason

some labels are manually assigned. When the subset is completely labeled, WordNet-Affect Core is defined and can be extended exploiting WordNet relations.

#### ➤ *SentiWordNet*

SentiWordNet is a lexical resource proposed in [2].

SentiWordNet is built with a ternary classification, indeed each synset (set of synonyms) is labeled as positive, negative or objective by using a set of ternary classifiers. If all of them will give to the synset the same label, therefore that label for that synset will have the maximum score; otherwise this score will be proportional.

Each classifier follows a semi-supervised approach that is a learning process where the training set  $Tr = L \cup U$  so that:  $L$  is a small subset of manually labeled training data, and  $U$  is a subset of training data labeled by the process by using  $L$ , and other available resource, as input.

In [2]  $L$  is divided into:  $L_p$ ,  $L_n$ , that are two small synsets respectively for positive and negative training data, and  $L_o$  for the objective ones.

$L_p$  and  $L_n$  are expanded with  $K$  iterations obtaining the following result for the  $i$ -th iteration:

$Tr_p^i$  (resp  $Tr_n^i$ ) will contain, in addition to  $Tr_p^{i-1}$  (resp  $Tr_n^{i-1}$ ), all the synsets that are related to synsets in  $Tr_p^{i-1}$  (resp  $Tr_n^{i-1}$ ) by WordNet lexical relations and have the same Positive(resp. Negative) polarity, and the synsets that are related to synsets in  $Tr_n^{i-1}$  (resp  $Tr_p^{i-1}$ ) and have the opposite polarity.

$Tr_o^K$  coincides with  $L_o$  and it consists of 17,530 synsets that doesn't belong either to  $Tr_p^K$  or to  $Tr_n^K$ . To each synset is associated a vectorial representation by applying a cosine-normalized tf\*idf to its gloss, that is a textual representation of its semantic.

Hence now the training synset, for a class  $c_i$ , can be given to a standard supervised learner that generates two binary classifiers. One of these will distinguish *positive* and *not\_positive* terms, and takes  $Tr_p^K \cup Tr_o^K$  in the training phase, the other one will classify terms as *negative* or *not\_negative*, and takes  $Tr_n^K \cup Tr_o^K$  in the training phase.

It produces a resulting ternary classifier that will classify the entire WordNet.

SentiWordNet has been developed in several versions, but the most significant is SentiWordNet 3.0 that, in the automatic annotation of WordNet, adds to the semi-supervised learning step a random-walk step for refining the scores. This version is compared with the previous one, and an improvement in accuracy of about 20% is found.

#### ➤ *Context Dependent Opinion Observer (CDOO)*

CDOO is a system implemented in C++ and it is based on a method that tries to infer the semantic orientation of opinion sentences by associating contextual information to opinion words obtained from WordNet.

This approach goes through four steps.

In the first step, after a preprocessing phase, opinion sentences are extracted from the inputs by using feature keywords directly.

In the second step Context independent opinions that don't



require any contextual information are analyzed to determine the semantic orientation. In this step opinion words from Wordnet are simply considered and in particular are utilized adjective synonym set and antonym set.

In the third step distinct-dependent opinions are analyzed: adjacent sentences are needed to define the semantic orientation by using Linguistic rules, especially conjunction rule.

In the fourth and final step Context indistinct opinions that need contextual information from other reviews are analyzed. In order to collect contextual segments sets for given features, a large number of online reviews are considered. Subsequently, contextual information is extracted from the segment sets by using Emotional-ATFxPDF to compute weight of terms in text segment set. Then the orientation of the opinion is calculated using semantic similarity.

#### ➤ *SenticNet*

SenticNet is inspired by SentiWordNet but it assigns to each concept  $c$  only one value  $p_c$  belonging to  $[-1,1]$ .

The polarity of a concept  $c$  is defined in the following way:

$$p_c = \frac{Plsn(c) + |Attn(c)| - |Snst(c)| + Aptt(c)}{9}$$

where *Plsn* is *Pleasantness*, *Attn* is *Attention*, *Snst* is *Sensitivity*, *Aptt* is *Aptitude*.

They start from Hourglass model and for example, in order to find positive concepts correlated with Pleasantness, they begin to search concepts semantically correlated to words like "joy", "serenity" and uncorrelated to words like "sadness".

Two different techniques are used: Blending and Spectral Assumption. When polarity is assigned, SenticNet is encoded in RDF triples using a XML syntax.

The current version of SenticNet contains almost 15,000 concepts.

In recent studies SenticNet is often associated to WordNet-Affect. For example in [10] researchers assign to SenticNet concepts, which are not present in WordNet-Affect, emotion labels. It is actually an expansion of WordNet-Affect based on SenticNet. By analyzing several features and utilizing a SVM framework for classification, they obtain an accuracy of 85.12% in their best result.

#### ➤ *Panas-t*

The original PANAS is created by Watson and Clark and they analyzed 10 moods on a 5-point scale [13].

They also expanded it in PANAS-x where eleven specific affects are considered: Fear, Sadness, Guilt, Hostility, Shyness, Fatigue, Surprise, Joviality, Self-Assurance, Attentiveness, and Serenity. To each affect a list of adjectives is associated.

In [5] it is expanded in Panas-t which is an adaptation that analyzes short text from Online Social Media and in particular from Twitter.

They consider a dataset composed by tweets from all the public accounts registered before August 2009. First tweets that explicitly contain feelings (and hence tweets that contain

words like "I am", "feelings", "myself") are identified.

Then a preprocessing phase is performed where individual terms are isolated, using white-space boundaries, and punctuation and other non-alphanumeric characters are removed.

It is assumed that a tweet can be mapped to the first sentiment  $s$  that appears in the tweet. This can be done by verifying the position of the adjectives.

### III. CONCLUSIONS

The paper provided an overview of the most diffused techniques for sentiment analysis based on the lexical-based approaches and the related systems.

The paper wants to be a quick reference guide in the choice of the most suitable lexical-based approaches for a specific problem of sentiment analysis.

### REFERENCES

- [1] F. Colace, L. Casaburi, M. De Santo, L. Greco, "Sentiment detection in social networks and in collaborative learning environments", *Computers in Human Behavior*, Available online 27 December 2014, ISSN 0747-5632, <http://dx.doi.org/10.1016/j.chb.2014.11.090>.
- [2] A. Esuli, and F. Sebastiani – "Sentiwordnet: A publicly available lexical resource for opinion mining", *Proceedings of LREC*. Vol. 6. 2006.
- [3] R. Feldman – "Techniques and Applications for Sentiment Analysis", *Magazine - Communication of the ACM* (April, 2013).
- [4] S. Feng, B. Ritwik, and C. Yejin – "Learning general connotation of words using graph-based algorithms", *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011.
- [5] P. Gonçalves, F. Benevenuto, and M. Cha - "Panas-t: A psychometric scale for measuring sentiments on twitter", *arXiv preprint arXiv:1308.1857* (2013). 14
- [6] V. Hatzivassiloglou and K. R. McKeown – "Predicting the semantic orientation of adjectives", *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics*. Association for Computational Linguistics, 1997.
- [7] J. M. Kleinberg – "Authoritative sources in a hyperlinked environment", *Journal of the ACM (JACM)* 46.5 (1999): 604-632. 1999.
- [8] R. Mihalcea, and H. Liu – "A Corpus-based Approach to Finding Happiness", *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*. 2006.
- [9] L. Page, S. Brin, R. Motwani and T. Winograd – "The PageRank citation ranking: Bringing order to the web" 1999.
- [10] S. Poria, A. Gelbukh, E. Cambria, P. Yang, A. Hussain and T. Durrani - "Merging SenticNet and WordNet-Affect emotion lists for sentiment analysis", *Signal Processing (ICSP)*, 2012 IEEE 11th International Conference on, vol. 2, no., pp. 1251, 1255, 21-25 Oct. 2012.
- [11] C. Strapparava, and A. Valitutti - "WordNet Affect: an Affective Extension of WordNet", *LREC*. Vol. 4. 2004.
- [12] G. Vinodhini, R. M. Chandrasekaran – "Sentiment Analysis and Opinion Mining: A Survey", *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 2, Issue 6 (June 2012).
- [13] D. Watson, L. A. Clark, and A. Tellegen - "Development and validation of brief measures of positive and negative affect: the PANAS scales", *Journal of personality and social psychology* 54.6 (1988).
- [14] M. Shelke, S. Deshpande, V. Thakre – "Survey of Techniques for Opinion Mining", *International Journal of Computer Applications* (0975-8887) Volume 57-No.13 (November 2012).

**Proceedings of**  
**Distributed Multimedia Systems**



# Effective Removal of Artifacts from Views Synthesized using Depth Image Based Rendering

Jiangbin Zheng, Danyang Zhao

Dept. of Computer Science and Engineering, School of  
Computer, Northwestern Polytechnical University  
Xi'an, China  
zhengjb0163@163.com

JinChang Ren

Dept. of Electronic and Electrical Engineering, University  
of Strathclyde Glasgow  
G1 1XW, United Kingdom  
jinchang.ren@strath.ac.uk

**Abstract**—Depth Image Based Rendering (DIBR), as a free-viewpoint synthesis technique, enables interactive selection of the view for watching. However, many rendering methods based on DIBR usually bring contour, crack and disocclusion artifacts. To address these problems, we propose effective methods to remove these artifacts. Firstly, before warping, the reference depth maps and color images are analyzed to find the regions causing contour artifacts. A combination of depth map edge detection and color consistency correction is applied to the analysis. By omitting warping the found regions, the synthesized views contain no contour artifacts and edges of foreground objects are well preserved. Secondly, cracks are filled using surrounding non-crack pixels rather than filtering for local consistency and smoothness. Thirdly, we apply a method based on texture extrapolation with depth information to inpaint the disocclusions. For two well-known sequences, ‘Ballet’ and Breakdancers’, we obtain large Peak Signal to Noise Ratio (PSNR) gains in comparison to state-of-the-art techniques. In addition, the proposed method also obtains good results in Structural Similarity Index Measurement (SSIM) and visual quality.

**Keywords**- Depth Image Based Rendering (DIBR); depth map edge detection; color consistency correction; contour artifacts; crack artifacts; disocclusion artifacts

## I. INTRODUCTION

With the development of computer techniques, transmission and display of digital images and videos become increasingly popular in a number of applications such as 3D video display [1] and 3D reconstruction. To satisfy the new requirements in terms of free-view and immersive experiencing, multi-view imaging (MVI) has attracted much more attention. As one of the most important applications of MVI, free-viewpoint TV (FTV) [2] brings a new visual experience where users can interactively select the view whilst watching videos.

Recently, Depth Image Based Rendering (DIBR) is emphasized, which involves the projection of a viewpoint into another. In general, two surrounding reference images are used in DIBR for view synthesis[3-12]. But there are different kinds of artifacts in the synthesized images based on DIBR. Three major artifacts need to be solved in such cases, which include contours caused by pixels at boundary of high depth discontinuities, cracks due to sampling rate of the reference image and disocclusions remained after blending projected images. Fig.1 shows the artifacts before processed.

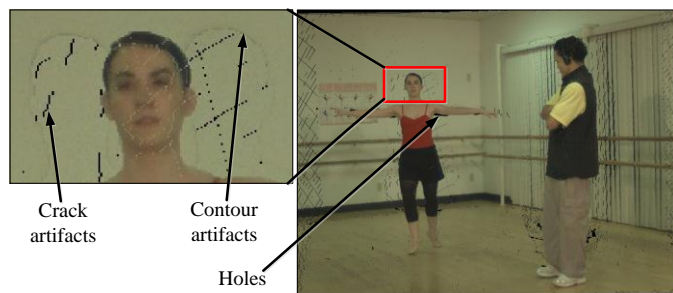


Figure 1. Artifacts in the synthesized view before processing

In this paper, we propose efficient algorithms to solve these artifacts. Firstly, depth map edge detection and color consistency correction are used in labeling pixels from contours, and only unlabeled pixels are used in 3D warping and blending images. Secondly, the depth differences are used in identifying crack regions and non-crack regions and cracks are filled using surrounding non-crack pixels. Thirdly, the hole artifacts are filled using texture extrapolation with depth information. Experimental results have demonstrated the efficacy of the proposed approaches. Detailed description of the proposed approaches and relevant results are reported in the next sections.

This paper is organized as follows. The related work is summarized in section 2. An overview of proposed algorithm is presented in section 3. The main removal approaches of artifacts are described in detail in section 4. In section 5, we show the experimental results and have a discussion. Finally, the conclusion is made in section 6.

## II. RELATED WORK

Depth Image Based Rendering (DIBR) techniques use depth-plus-video data for free-viewpoint rendering. The 3D sense representation based on depth-plus-video combine the advantages of geometry based representation and image based representation. Comparing to geometry and image based representation, depth-plus-video 3D sense representation does not use explicit 3D models and a large number of images and only depth map and fewer images need to be processed [8]. Previous research on DIBR from one reference image has two inherent limitations, which are viewpoint dependency of textures and disocclusions. To overcome these limitations, most recent methods employ warping from two surrounding

reference images to a virtual viewpoint. Disocclusions from one reference are compensated by the other view. Zitnick et al. [9] pointed out that three main artifacts need to be removed in rendering a high quality virtual viewpoint. First of all, pixels at high discontinuities tend to cause contour artifacts which need to be fixed. Secondly, empty pixels and holes due to insufficient sampling of the reference images need to be filled. Thirdly, the remaining disocclusions after blending the warped images need to be generated.

To overcome these artifacts, recent research involved different methods. We classify the techniques into three categories according to different kind artifacts. We will review some typical techniques as follows.

### 2.1 Techniques of contour artifacts removing

In the reference image, because of borders staying in high discontinuities, the blended images always contain contour artifacts. Muller et al. [13] provided a free-viewpoint rendering algorithm based on a layered depth map presentation. They defined three layers: foreground boundary layer, background boundary layer and main layer. They first project the main layer and get the blended image. Secondly, they projected foreground and background boundary layer and they used a simple depth test to avoid contour artifact existing. The quality of this algorithm is not measured. It also requires amount of pre-process and post-process.

Mori et al. [5] used boundary matting method to remove contour artifacts. After 3D warping, they expanded the boundary to background direction and successfully removed contour artifacts in background, but this method is not efficient to the contour artifacts in foreground.

Luat et al. [6] first detected the foreground boundary in depth map and labeled the unreliable regions which corresponding to Fig. 3. When they did 3D warping they only used the unlabeled regions. This method may remove the contour artifacts, but boundary shrinks to the foreground direction which damages the foreground objects.

### 2.2 Techniques of crack artifacts removing

Crack artifacts often occur in the virtual image after forward 3D warping. Each point from an original image is projected separately into the virtual view, and falls in general onto a floating point coordinate. This position is quantized to the nearest neighbor position of the integer sample raster. Unfortunately, quantization may leave some samples unfilled being visible as thin black lines in Fig. 5(a). In some case, such cracks in foreground regions are filled by background information.

Mori et al. applied a bilateral filter. This method is efficient to the cracks in background. But this method didn't consider depth information and was not efficient to cracks filled by background information in foreground regions. Based on the algorithm of Mori et al., Luat et al. proposed to process the virtual depth image with a median filter. Afterwards they compared the input and output of the median filter and performed an inverse warping when pixels have changed. This

method can fill cracks both in foreground and background, but its inverse warping may increase computational errors.

### 2.3 Techniques of disocclusion inpainting

The blended virtual images often contain some disocclusion regions which are visible in virtual view but occluded in both reference views.

Mori et al. inpainted disocclusion regions with the method proposed by Telea [14]. This method used only texture information and no depth information. Using this method, the inpainted regions between background and foreground may contain blurs. Luat et al. introduced inpainting method with depth information on the basis of Telea's method. In [15], each virtual view image featuring disocclusions is compensated using image information from a causal picture neighborhood via a background sprite. Residual uncovered areas are initially coarsely estimated and then refined using texture synthesis. Chen [16] proposed edge dependent Gaussian depth filter and interpolation to fill holes.

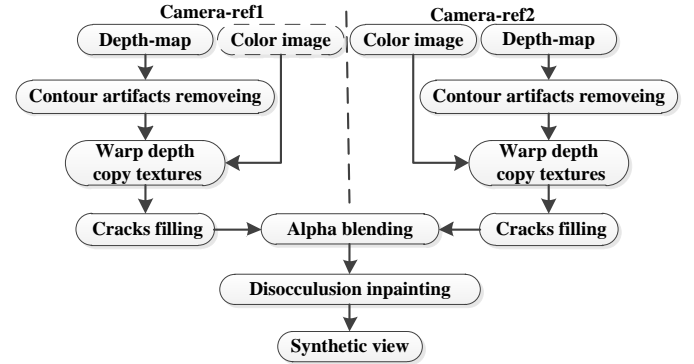


Figure 2. Block diagram of proposed rendering algorithm

## III. PROPOSED FREE-VIEWPOINT RENDERING ALGORITHM

An overview of proposed algorithm is given in Fig. 2. The synthesis approach can be treated as a mapping process with some processing steps which are used to remove the artifacts during the map process. In our algorithm, the inputs are two reference views with depth and color images.

The first step is depth reliability analysis, which is used to analyze the misalignment of sharp depth map edges as detailed in Section 3.1. The misalignment of sharp depth map edges is the major cause of the contours in the synthesized images, and our solution to solve this problem is presented in Section 3.2. The second step is warping. In order to reduce the computational errors, when the depth map of the reference plane is projected to virtual view, the textures corresponding to the depth map are directly copied.

The third step is crack-filling. The insufficient sampling of the reference image introduces cracks in the synthesized images. To preserve the original warped information and achieve good filling quality, we propose a crack filling method which need not filter and re-projection. Instead, we use the surrounding depth and color information of non-crack pixels for crack-filling, and relevant method is discussed in Section 3.3. The next step is alpha blending which used to synthesize the warped images together.

The last step is disocclusion inpainting. After get synthesized depth maps and textures, the images still contain disocclusions, due to the fact that some parts of the scene can be seen at the synthesized view but invisible in the reference views. The surrounding non-hole pixels with depth information are used to fill these disocclusions.

#### IV. ARTIFACTS ANALYSIS AND SOLUTIONS

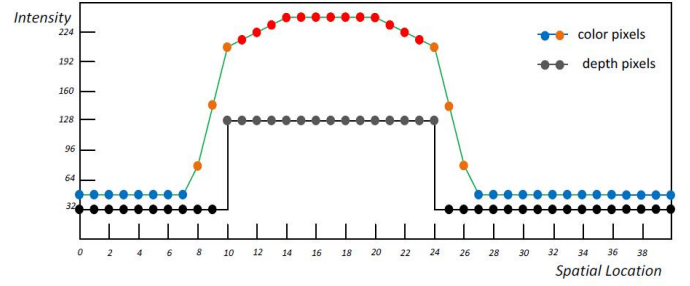
In this section, we will give detailed description of proposed approaches for solving the major artifacts in synthesized images.

##### 4.1 Depth map reliability analysis

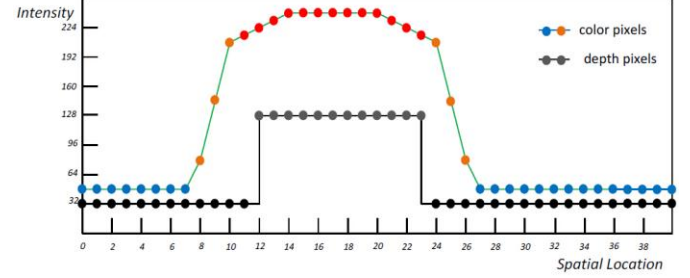
Generally, the depth map captured by depth camera or estimated from video frames may not align with color view correctly. Depth map is a piece-wise image that has large homogeneous regions within scene object and contains sharp changes at object boundaries. However, the edges in color image usually have intensity changing smoothly over transition regions where the object boundaries contain combination of foreground and background. There are detailed descriptions in [17]. Generally, if the depth map aligns with the color image as showed in Fig. 3(a), the object's sharp boundary edge will be at the transition region of color image. However, if the depth map misaligns with the color image, there are two kinds of misalignments between depth map and color image. The first kind misalignment is showed in Fig. 3(b) where the object's sharp boundary edges of depth map align with the foreground of color image. The second kind misalignment is showed in Fig. 3(c) where the object's sharp boundary edges of depth map align with the background of color image.

The synthesized image will contain contour artifacts if these misalignments are not processed. Mori et al. applied a boundary matting method. After 3D warping of depth maps, they expanded the occlusion region to background direction, this erased the mixture of foreground and background color in the background and the occlusion region were filled by the other warped depth map. As this can only remove contours in the background, and fail to deal with contours in the foreground in Fig. 4(b).

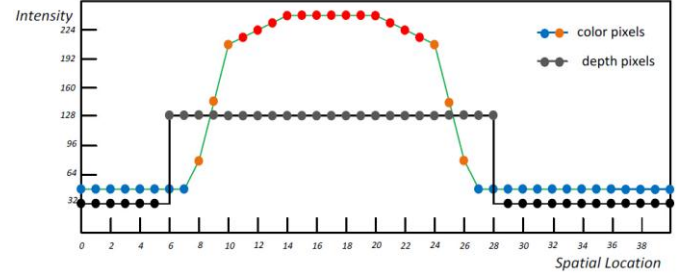
Luat et al. proposed a depth map preprocess method which detected the foreground boundary, and labeled the boundary as unreliable regions. When they did the 3D warping, they only warped the pixels that are not labeled. This method can get a good result for the alignment in Fig. 3(a). For another two kinds of misalignment showed in Fig. 3(b) and Fig. 3(c), this method still can't yield satisfactory results. For the first kind misalignment when the edge of depth aligns with foreground of color image, this method can remove the contour, but it also erased the foreground of color image. This becomes more apparent especially when the foreground object is thin, such as fingers showed in Fig. 4(a). For the second kind misalignment, this method will bring contours in the foreground region as shown in Fig. 4(b). To deal with all three cases of misalignment and remove the contours we propose a new method based on depth map edge detection and color consistency correction. This is discussed in detail the next subsection.



(a) Depth map boundaries aligned with transition region.



(b) Depth map boundaries misaligned with foreground region.



(c) Depth map boundaries misaligned with background region.

Figure 3. Color pixel intensity values and depth values for a horizontal line in video-plus-depth image format[17]

##### 4.2 Contour artifacts removing

According to the above analysis, the depth map can be refined in the preprocessing stage in order to remove the contour artifacts. For this artifact, we need process all the three kinds of alignment or misalignments. To this end, combination of depth map edge detection and color consistency correction is employed.

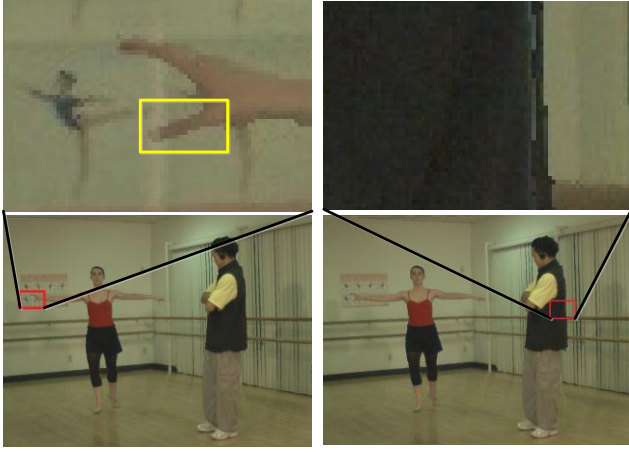
In the proposed method, firstly we analyze the depth maps and find out the unreliable regions. If pixels in depth map are satisfied (1) or (2), they are labeled as unreliable pixels. Pixels satisfied (1) correspond to foreground and pixels satisfied (2) correspond to background, respectively.

$$\forall_{u,v} \in S, \left( \sum_{i=-2}^2 \sum_{j=-2}^2 D(u+i, v+j) \right) - (5*5) * D(u, v) > T_d \quad (1)$$

$$\forall_{u,v} \in S, \left( \sum_{i=-2}^2 \sum_{j=-2}^2 D(u+i, v+j) \right) - (5*5) * D(u, v) < -T_d \quad (2)$$

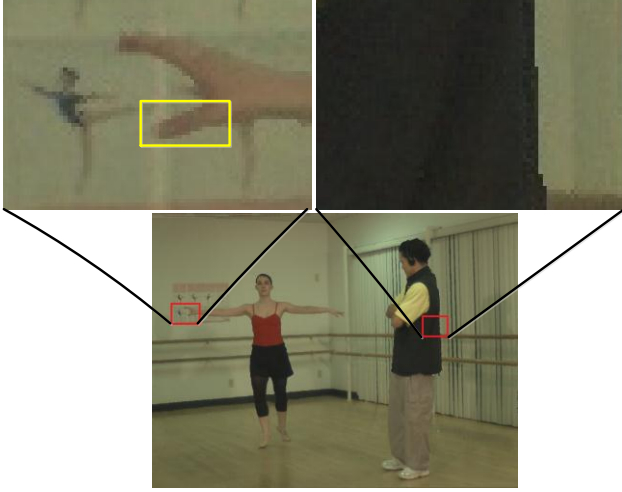
Where  $S$  is the image space,  $D$  denotes the depth map of the reference camera and  $T_d$  is predefined threshold. Generally the mixture region of foreground and background has a width of two to three pixels. To detect such regions, the image block to be examined should be larger than this size, and this is why we use a 5x5 block above.





(a) Result of Luat et al.

(b) Result of Mori et al. and Luat et al.



(c) Result of proposed method

Fig.4: Results comparison. In (a), object boundary shrinks to the foreground direction and finger become thinner. In (b), contours still exist in foreground. In (c), no contour exists in images

We only warp the unlabeled pixels in the depth map and thus there will be no contour artifacts in background of the synthesized view. This process is effective for contours in the background. We needn't worry about excessively erasing the background of depth map as the erased region will be filled by the other reference view. Although it may also remove contours in the foreground, the object boundary shrinks to the foreground direction which damages the foreground objects, especially those small or thin ones such as fingers showed in Fig. 4(a).

To preserve the foreground boundary information and remove contours in foreground at the same time, further work based on color consistency correction are used as follows. First, for every pixel that satisfied (1), we search its neighbor pixels in eight directions and find the nearest unlabeled pixels in each direction. Second, we calculate the differences of color values between current pixel and the unlabeled pixels that in foreground. We re-label the pixels to reliable if (3) is satisfied.



(a) Image with cracks

(b) Image after cracks filled

Figure 5. Processing image crack

$$\forall_{u,v} \in R, \left| \frac{\sum_{i=1}^N c_i * r_i}{\sum_{i=1}^N r_i} - c(u,v) \right| < T_c \quad (3)$$

$$\text{with } r_i = \begin{cases} 1 & , \text{if reliable \& foreground} \\ 0 & , \text{others} \end{cases}$$

In (3),  $R$  is the unreliable region which satisfy (2); parameter  $N=8$ , denotes the eight directions;  $c_i$  denotes the color value of the nearest unlabeled pixel in direction  $i$ .  $T_c$  is a predefined threshold. As we can see, if the unreliable pixels in depth map satisfy (3) which belong to the foreground of color image, they should be relabeled as reliable regions. This process preserves foreground boundary and removes contours in foreground at the same time. The results are given in Fig. 4 (c). As can be seen, our method has the advantage of removing contours from both foreground and background whilst preserving the foreground boundary information.

#### 4.3 Cracks filling

Small cracks often occur in the entire image area and are introduced by the forward mapping nature of image-based 3D warping. In this section, we will present our method for crack filling. To fill the cracks efficiently and preserve the original warped depth information of non-crack regions, we propose a method that only fills cracks and does nothing to non-crack regions. Usually the width of crack is one to two pixels. First, we used (4) to detect the cracks.

$$\forall_{u,v} \in S, crack_{u,v} = \begin{cases} 1 & , \text{if } \sum_{i=-2}^2 \sum_{j=-2}^2 f_{ij} \geq T_n \\ 0 & , \text{else} \end{cases} \quad (4)$$

$$\text{where } f_{ij} = \begin{cases} 1 & , \text{if } |D(u+i, v+j) - D(u, v)| > T_d \\ 0 & , \text{else} \end{cases}$$

Where  $S$  is the image space,  $D$  denotes the depth map of the synthesized view,  $T_d$  and  $T_n$  are predefined thresholds,  $f_{ij}$  denotes the difference between the current pixel and each surrounding pixel.  $crack_{u,v}$  is a flag, indicating whether current pixel is in crack or not. We check the depth difference between the current pixel and each surrounding pixel in a  $5 \times 5$  block. If the current pixel is in crack regions, most of the surrounding pixels are not in crack regions, we will get a large summation. On the contrary, if the current pixel is not in crack regions, summation will be small.

After detecting cracks, we use the surrounding pixels which are not in crack regions to calculate average values to fill pixels in crack regions, specified by (5).

$$\forall_{u,v} \in S, c(u,v) = \frac{\sum_{i=1}^N c_i * (1 - crack_i)}{\sum_{i=1}^N (1 - crack_i)} \quad (5)$$

Where  $S$  is the image space, again  $c_i$  denotes the color value of the nearest non-crack pixel in direction  $i$ , parameter  $N = 8$ , denotes the pixels in eight directions,  $crack_i$  defined in (4) is the crack label of  $c_i$ . After this step, as shown in Fig. 5, cracks in foreground and background are successfully filled and the original information of non-crack regions are preserved.

#### 4.4 Alpha blending

After the warped images were removed contour artifacts and filled cracks, we blend the two images to one using the (6)

$$C(u,v) = \begin{cases} (1-\alpha)C_L(u_L, v_L) + \alpha C_R(u_R, v_R) & (occ_L(u,v) = 0, occ_R(u,v) = 0) \\ C_L(u_L, v_L) & (occ_L(u,v) = 0, occ_R(u,v) = 1) \\ C_R(u_R, v_R) & (occ_L(u,v) = 1, occ_R(u,v) = 0) \\ 0 & (occ_L(u,v) = 1, occ_R(u,v) = 1) \end{cases} \quad (6)$$

$$occ_L(u,v) = \begin{cases} 1 & (D_L(u,v) = 0) \\ 0 & (D_L(u,v) > 0) \end{cases} \quad (7)$$

$$occ_R(u,v) = \begin{cases} 1 & (D_R(u,v) = 0) \\ 0 & (D_R(u,v) > 0) \end{cases}$$

$$\alpha = \frac{|t - t_L|}{|t - t_L| + |t - t_R|} \quad (8)$$

In (6),  $C(u,v)$  means the pixel value at  $(u,v)$  virtual plane.  $C_L$  and  $C_R$  mean the images generated by left and right reference views.  $occ$  is the occlusion map defined in (7).  $D_L(u,v)$  and  $D_R(u,v)$  mean the left and right pixel depth values at  $(u,v)$  depth image. While depth value equals to zero, it means this pixel stay in disocclusion region which corresponding to the black blank regions in Fig. 6. These pixels can't be used for blending.  $\alpha$  is a weighting coefficient defined in (8).  $t_L, t_R$  and  $t$  are the translation vectors of left camera, the right camera and the virtual camera, respectively. The blending result is shown in Fig. 6.

#### 4.5 Disocclusion inpainting

As we can see from Fig. 6(c), after blending two warped images into virtual image, the synthesized image still contains disocclusions. These regions either occur due to erroneous depth values, or are areas that become visible in the virtual view, while being occluded in both original views. Most disocclusions inpainting methods are based on texture extrapolation. These methods may get good results when disocclusions only surrounded by foreground or background. If the disocclusions are surrounded by both foreground and background, the inpainted regions often contain a certain

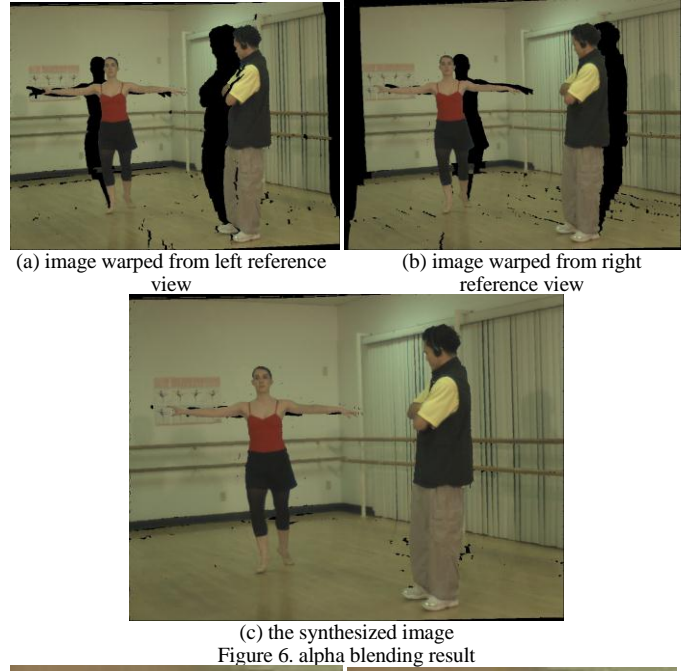


Figure 6. alpha blending result



Figure 7. the inpainting result

amount of blurs. In the experiment, we found that the disocclusions are always background and not part of foreground. On the basis of this fact, we use texture extrapolation based on depth information to inpaint the disocclusions, specified by

$$\forall_{u,v} \in O, \quad C(u,v) = \frac{\sum_{i=1}^N d_i^{-2} * f_i * C_i}{\sum_{i=1}^N d_i^{-2} * f_i} \quad (9)$$

$$\text{with } f_i = \begin{cases} 0 & , \text{if } \max D_j - D_i > T \quad (j = 1, \dots, N) \\ 1 & , \text{else} \end{cases}$$

Here,  $O$  is the disoccluded region, parameter  $N = 8$ ,  $d$  is the distance of current pixel to the edge of the disoccluded region and  $C_i$  is the texture value of edge pixel in direction  $i$ .  $f_i$  reflects  $C_i$  staying in foreground or background.  $D_i$  denotes the depth value.  $T$  is predefined threshold. In this summation, only those texture values belonging to the background are used. The advantage of this method is that no blur exists between foreground and background. The drawback is that the inpainted region may become a low-frequency patch, when the disocclusion region is large. Fig. 7 shows the comparison between traditional texture extrapolation and our method.

## V. RESULTS

### 5.1 Image dataset

To evaluate the performance of proposed method, the well-known multi-views sequence of “Ballet” and “Breakdancer” are used. The camera setup of the test sequences consists of eight reference cameras positioned along an arc, spanning about  $30^\circ$  from one end to the other. Both of the two sequences include 100 images captured from every camera. The depth maps were computed from stereo and also included for each camera along with the calibration parameters. The captured images have a resolution  $1024 \times 768$ . We select camera 4 as the virtual view and compare the generated images with the reference images.

### 5.2 Experimental results

Both subjective and objective evaluations are applied to evaluate the performance of proposed method. For objective evaluation, the synthesized images are compared with the reference images in the virtual view based on the Peak signal-to-noise ratio (PSNR) and structural similarity metric (SSIM). For subjective evaluation, we present some synthesized images and compare with the reference images.

To compare the object performance of our proposed method with those in Mori et al. [5] and Luat et al. [6], average PSNR results calculated over 100 images are illustrated in Fig. 8 with changing distance between the left and right reference cameras. The distances are calculated by  $|t - t_L| + |t - t_R|$  which defined in (8). To compare the performance in detail, Fig. 9 (a) and (b) show detailed PSNR results of 100 images synthesized by camera 3 and camera 5. The average PSNR of these images are corresponding to the first points of proposed method in Fig. 8(a) and Fig. 8(b), respectively. In order to evaluate our method more comprehensively, we also present our SSIM results in Table 1. Because Mori et al. and Luat et al. didn’t measure their results on SSIM, we only present our results.

In order to evaluate subjective performance of our method, we present some synthesized images in Fig. 10. Compared with the reference images, our results obtain good visual quality.

### 5.3 Discussion

As can be seen, for the two sequences, our approach has significantly outperformed the other two and generates much higher PSNR values than them. The large gains are caused by the fact that we successfully remove the contours in foreground and background and we preserve the foreground boundary information at the same time. Our crack-filling method only focuses on the crack regions instead of using filtering which may break the original warped information of non-crack regions. Our disocclusion inpainting method is based on depth information and reduces blurs in the synthesized images.

From experimental results, we can see that the quality of synthesized images have large difference between the two sequence. Through the study of dataset, we find that the scene of “Breakdancer” is more complex than “Ballet” and the depth maps are calculated using stereo vision algorithm, so the depth maps of “Breakdancer” have lower accuracy. The rendering

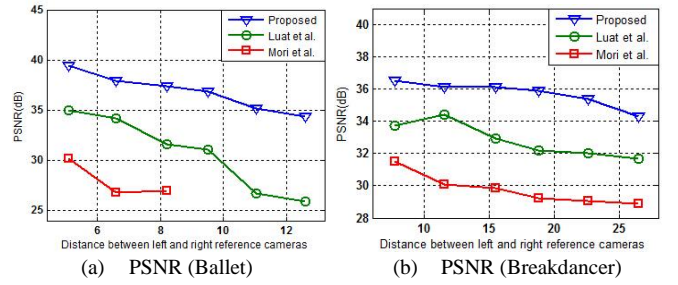


Figure 8. PSNR results with different distances (mm) between left and right reference cameras. Each point of proposed method is an average result of 100 synthesized images.

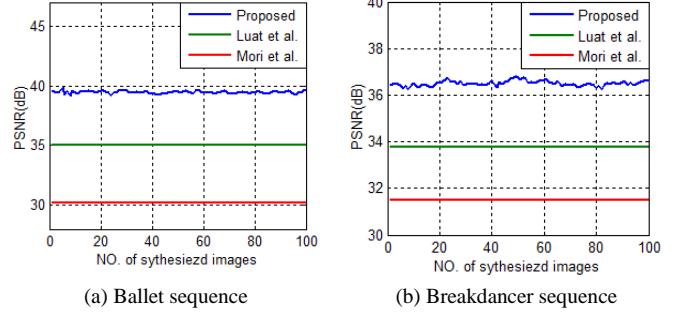


Figure 9. PSNR results of 100 images synthesized by camera 3 and camera 5. Lines from top to bottom are results of proposed, Luat et al. and Mori et al. respectively.

TABLE I. SSIM RESULTS WITH DIFFERENT DISTANCES BETWEEN LEFT AND RIGHT REFERENCE CAMERAS. EACH RESULT IS AN AVERAGE OF 100 SYNTHESIZED IMAGES.

Sequence	Distance between left and right view					
	7.7	11.5	15.5	18.8	22.6	26.5
Ballet	0.95	0.94	0.93	0.91	0.87	0.84
Breakdancer	0.92	0.92	0.91	0.89	0.88	0.85

method based on DIBR is sensitive to the accuracy of depth maps.

We also notice that with the increase of distance between left and right reference cameras, the quality of synthesized images become worse. When the baseline between reference cameras becomes larger, there will be more computing errors and larger occluded areas.

Through the above analysis, the qualities of synthesized images are sensitive to the depth map accuracy and length of baseline between reference cameras.

## VI. CONCLUSION AND FUTURE WORK

The view synthesized using methods based on DIBR always contains various artifacts, where contour, crack and disocclusion artifacts are typical cases which have veritably degraded the quality of the synthesized images. In this paper, we propose effective methods to remove such artifacts. Based on depth map edge detection and color consistency correction, contour artifacts are successfully removed from both foreground and background whilst preserving the foreground boundary information. For filling cracks, depth difference is applied to detect crack and non-crack pixels. The proposed





Figure 10. Comparison synthesized images (left) with reference images (right) in the view of camera4.

method for filling cracks is proved to be effective. With the disocclusion regions, we applied a method of texture extrapolation with depth information. Experimental results on two well-known sequences have demonstrated the promising results of the proposed approaches. According to quantitative assessment using PSNR, the proposed algorithm outperforms two state-of-the-art approaches. Good results also are obtained in SSIM and visual quality. To get good synthesized result in

large baseline and faster rendering speed, future work will focus on large baseline and real-time free-viewpoint system.

## VII. ACKNOWLEDGMENT

This work was supported by Importance Industry Chain Projects of Science and Technology Coordination Innovation Engineering, Shaanxi, China (2015KTZDGY04-01).

## REFERENCES

- [1] M. Domański, A. Dziembowski, A. Kuehn, 2014. Experiments on acquisition and processing of video for Free-viewpoint television, IEEE 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), Budapest, Hungary, pp.1-4.
- [2] M. Tanimoto, M. P. Tehrani, and T. Fujii, 2011. Free-viewpoint TV, IEEE Signal Processing Magazine, vol. 28, no. 1, pp. 67-76.
- [3] Y. Cai, R. Wang, and T. Cui, 2013. Intermediate view syn-thesis based on edge detecting, IEEE International Conference on Image Processing, Paris, France, pp. 3172-3175.
- [4] V. Paradiso, M. Lucenteforte, M. Grangetto, 2012. A novel interpolation method for 3D view synthesis, IEEE 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), Zurich, Switzerland, pp. 1-4.
- [5] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, 2009. View generation with 3D warping using depth information for FTV, Signal Processing: Image Communication, vol. 24, no. 1, pp. 65-72.
- [6] L. Do, G. Bravo, S. Zinger, 2012. GPU-accelerated real-time free-viewpoint DIBR for 3DTV. IEEE Transactions on Consumer Electronics, vol. 58, no. 2, pp. 633-640.
- [7] H. C. Shin, G. Lee, N. Hur, 2014. View interpolation using a simple block matching and guided image filtering, IEEE 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), Budapest, Hungary, pp. 1-4.
- [8] A. Smolic. 2011. 3D video and free viewpoint video—From capture to display. Pattern recognition, vol. 44, no. 9, pp. 1958-1968.
- [9] C. L. Zitnik, S. B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, 2004. High-quality video view interpolation using a layered representation, ACM SIGGRAPH, Los Angeles, USA, pp. 600-608.
- [10] K. Oh, S. Yea, Y. Ho, 2009. Hole-Filling Method Using Depth Based Inpainting For View Synthesis in Free Viewpoint Television (FTV) and 3D Video, IEEE Picture Coding Symposium, pp. 1-4.
- [11] A. Smolic, K. Müller, K. Dix, P. Merkle, P. Kauff, T. Wiegand, et al. 2008. Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems, IEEE Image Processing, pp. 2448-2451.
- [12] S. Zinger, L. Do, and P. H. N. de With, 2010. Free-viewpoint depth image based rendering, Journal of Visual Communication & Image Representation, pp. 533-541.
- [13] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, T. Wiegand, et al. 2008. View Synthesis for Advanced 3D Video Systems, EURASIP Journal on Image and Video Processing, pp. 1-12.
- [14] A. Telea, 2004. An image inpainting technique based on the fast marching method, Journal of Graphics Tools, vol. 9, no. 1, pp. 25-36.
- [15] P. Ndjiki-Nya, M. Koppel, D. Doshkov, 2011. Depth image-based rendering with advanced texture synthesis for 3-D video. IEEE Transactions on Multimedia, vol. 13, no. 3, pp. 453-465.
- [16] W. Chen, Y. Chang, S. Lin, 2005. Efficient depth image based rendering with edge dependent depth filter and interpolation. IEEE International Conference on Multimedia and Expo. Amsterdam, The Netherlands, pp. 1314-1317.
- [17] X. Xu, L. M. Po, K. W. Cheung, 2013. Watershed based depth map misalignment correction and foreground bi-ased dilation for DIBR view synthesis, IEEE International Conference on Image Processing, Paris, France, pp. 3152-3156.

# Cross-Covariance-Based Features for Speech Classification in Film Audio

Matt Benatan and Kia Ng

University of Leeds,  
School of Computing,  
Leeds LS2 9JT, United Kingdom  
mattbenatan@gmail.com, k.c.ng@leeds.ac.uk

**Abstract**— As multimedia becomes the dominant form of entertainment through an ever increasing range of digital formats, there has been a growing interest in obtaining information from entertainment media. Speech is one of the core resources in multimedia, providing a foundation for the extraction of semantic information. Thus, detecting speech is a critical first step for speech-based information retrieval systems. This work focuses on speech detection in one of the dominant forms of entertainment media: feature films. A novel approach for voice activity detection (VAD) in film audio is proposed. The approach uses correlation to analyze associations of Mel Frequency Cepstral Coefficient (MFCC) pairs in speech and non-speech data. This information then drives feature selection for the creation of MFCC cross-covariance feature vectors (MFCC-CCs) which are used to train a random forest classifier to solve a binary speech/non-speech classification problem on audio data from entertainment media. The classifier performance is evaluated on a number of test sets and achieves a classification accuracy of up to 94%. The approach is also compared with state of the art and contemporary VAD algorithms, and demonstrates competitive results.

**Keywords**- *voice activity detection; speech detection; binary classification; film audio; entertainment media*

## I. INTRODUCTION

Consumption of multimedia has become ubiquitous, with TV, films, games, and digital music now providing the majority of our entertainment in a range of easily accessible formats. With this rise in multimedia, there is a continually increasing interest in obtaining information from media - using it to understand human interaction and behavior [1], and to extract semantic information that can be used in the creation of metadata [2]. Speech classification plays a key role in data extraction through detecting speech regions in audio or video data. These regions can then be used for further feature extraction, e.g. speech recognition. While many speech detection techniques exist, few have been developed specifically for use with one of our most challenging and popular forms of media: film. Unlike radio and news broadcasts, films contain an extremely diverse range of speech and other audio content. Film introduces challenges that

are not present in most natural scenarios, such as speech in the presence of highly dynamic background noise and sound effects, or heavily manipulated speech, where sound design has been used to create unnatural voice characteristics through the addition of harmonics and distortion.

We present a novel approach for speech detection developed specifically for classification of speech within film audio. This approach aims to account for unusual voice characteristics by analyzing the relationships between pairs of spectral features within speech and non-speech data. We use the process to identify Mel Frequency Cepstral Coefficient (MFCC) pairs which are then processed to create cross-covariance-based feature vectors (MFCC-CCs). MFCC covariance statistics have been used previously for audio classification tasks, such as in [3] and [4], where covariance is used alongside other statistical representations of MFCC data, resulting in as many as 60 dimensions per frame (as described in [4]). In this work, cross-correlation is used to select specific MFCC pairs which demonstrate the greatest difference in correlation between speech and non-speech data. Cross-covariance vectors for the five highest scoring MFCC pairs are then created, providing a single vector which represents the covariance relationship for each pair. The resulting feature vector is comprised of five speech-sensitive MFCC-CC features per frame, thus reducing dimensionality from 13 MFCCs to five MFCC-CC features. Through using this feature vector with a random forest classifier, we have achieved a classification accuracy of 94% on challenging audio data.

## II. BACKGROUND

Recent developments in mixed-audio speech detection have demonstrated high accuracy [5], however, while using mixed audio signals, the datasets used in much of the work to date is still fairly constrained. These include radio broadcasts [6], news broadcasts [7], and speech detection in the presence of background noise [8]. Speech detection in these scenarios is likely to be a simpler task than speech detection within film audio. This is due to the dynamic nature of film audio: not only does it contain various types of background noise, but the acoustic environments change frequently (simulated or

otherwise, e.g. via reverb effects [9]) and the format makes use of many synthetic sound effects [10], which can obscure speech information in the audio. As well as this, voice synthesis or distortion is now also common in feature films [10], all of which make speech detection more challenging when using typical spectral features. To address this, we have developed an approach for speech detection that uses cross-covariance to represent the relationship between pairs of MFCCs [11]. This reduces feature dimensionality, resulting in a feature set designed to improve speech/non-speech discrimination. The resulting feature vector is used to train a learning machine to perform binary classification (speech/non-speech) on an annotated ground-truth dataset. Results demonstrate an accuracy of between 86.15% and 87.26%, which are promising performance statistics when considering the challenging nature of the dataset.

An approach discussed in [6], for classifying speech/non-speech in radio broadcasts, exploits spectro-temporal variations of speech signals via short time Fourier transforms (STFTs) to discriminate between speech and non-speech signals. This has demonstrated good performance on their data set, however, this approach applies a median filter or approximately 10 seconds duration to the classifier output. Thus, it is primarily useful for broadly classifying sections of audio, rather than for higher resolution speech activity detection. Furthermore, the data used is sourced exclusively from radio broadcasts, and is thus not reflective of film audio content, likely being less dynamic and thus simplifying the classification problem.

Another recent approach described in [5] uses a voice activity detector based on Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN). This demonstrates good performance on a synthetic test validation set, with an average equal error rate (EER) of 10.4%, outperforming the state-of-the-art SOHN algorithm. However, it is less effective on film audio, with an average EER of 33.2%.

One film-centered approach [12] utilizes bilingual audio streams for speech detection. This identifies speech segments through correlating spectral coefficients between two different language tracks, and demonstrates an accuracy of between 84% and 87% in classifying clean and noisy speech in film audio. While this approach demonstrates good performance on film data, it requires bilingual audio tracks to perform classification, and thus would not work with single language audio data.

Another approach discussed in [13] uses a dataset comprised entirely of television material (thus similar to film) and looks to differentiate between speech and music data. This uses discrete wavelet transforms (DWT) as the audio feature and performs classification via a support vector machine. While this performs with an accuracy of up to 94.5%, the approach is focused on discriminating between speech or music data, and thus does not consider environmental noise, silence, sound effects and other sonic components common to film audio.

Several other reviewed approaches have demonstrated an accuracy of >90%, however, these either have limited data, such as [14], which has only 9 main speakers in its dataset, or make use of non-film audio, such as [7], whose data includes radio and news broadcasts (which typically do not have the same sonic variance as film data).

### III. PROPOSED APPROACH

#### A. Process Overview

The speech classification process consists of three core stages. The first of these is feature selection, which analyzes the audio data using cross-correlation to determine which features yield the most useful information to discriminate between speech and non-speech data. The second stage consists of processing this information to create the MFCC-CC feature vectors, and in the third stage a classifier is fit to a training set of ground-truth labelled data.

#### B. Feature Selection

Numerous approaches for spectral feature parameterization exist [15], however MFCCs are one of the most frequently used spectral features in both automatic speech recognition (ASR) [16] and voice activity detection (VAD) [17]. Given their wide adoption in speech processing, MFCCs have been chosen as the method of representing spectral features in this work. Within this application we replace the zeroth MFCC with the log of the total frame energy, as this has proven to be useful in speech processing applications [18] [19].

Feature selection is achieved using cross-correlation to analyze the difference in cross-MFCC similarities in speech and non-speech data from the training set. A correlation coefficient is obtained for each MFCC with respect to each of the other MFCCs. This is done separately for speech and non-speech data. The speech/non-speech difference in the resulting correlation coefficients for each MFCC feature pair is obtained. This is used to determine which feature pairs demonstrate the greatest change in correlation between speech and non-speech data. The Pearson product-moment correlation coefficient,  $\rho$ , is obtained from the covariance matrix ( $C$ ) of a pair of MFCC feature vectors via the coefficient matrix  $P_{ij}$ :

$$P_{ij} = \frac{c_{ij}}{\sqrt{c_{ii} * c_{jj}}} \quad (1)$$

The correlation coefficient has a value between -1 and 1, where 1 denotes total positive correlation, and -1 denotes total negative correlation.

The MFCC pairs are chosen based on the difference between their speech and non-speech correlation coefficients. Figure 1 shows the resulting correlation coefficient differences. Higher values indicate a greater variance in the MFCC pair relationships between speech and non-speech data. This in turn indicates that the pairs are more likely to provide information relating to the presence/absence of speech spectral data, thus facilitating more effective speech/non-speech discrimination.



MFCC	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0	0.04	0.1	0.21	0.18	0.06	0.34	0	0.06	0.23	0.12	0.21	0.1
1	0.04	0	0.1	0.01	0.31	0.06	0.11	0.02	0.2	0.02	0.19	0.08	0.15
2	0.1	0.1	0	0.06	0.13	0.08	0.08	0	0.24	0.12	0.15	0.13	0.04
3	0.21	0.01	0.06	0	0.12	0.15	0.04	0.13	0.04	0.03	0.1	0.07	0.09
4	0.18	0.31	0.13	0.12	0	0.04	0.07	0.04	0.04	0.01	0.03	0.01	0.11
5	0.06	0.06	0.08	0.15	0.04	0	0.1	0.12	0.03	0.04	0.08	0.06	0.02
6	0.34	0.11	0.08	0.04	0.07	0.1	0	0.2	0.03	0.1	0.07	0.1	0.2
7	0	0.02	0	0.13	0.04	0.12	0.2	0	0.07	0.13	0.07	0.17	0.17
8	0.06	0.2	0.24	0.04	0.04	0.03	0.03	0.07	0	0.1	0.19	0.05	0.18
9	0.23	0.02	0.12	0.03	0.01	0.04	0.1	0.13	0.1	0	0.07	0.15	0.03
10	0.12	0.19	0.15	0.1	0.03	0.08	0.07	0.07	0.19	0.07	0	0.12	0.2
11	0.21	0.08	0.13	0.07	0.01	0.06	0.1	0.17	0.05	0.15	0.12	0	0.17
12	0.1	0.15	0.04	0.09	0.11	0.02	0.2	0.17	0.18	0.03	0.2	0.17	0

Figure 1. Matrix of MFCC pair correlation coefficient differences between speech and non-speech data. Darker squares indicate greater values.

### C. Feature Vector Processing

The final MFCC-cross-covariance feature vectors are attained by computing the cross-covariance of the MFCC pairs corresponding to the top  $n$  correlation coefficient differences. For each pair of MFCCs, the cross-covariance vector is obtained through computing the cross-covariance of segments of the two signals along their length via a rectangular sliding window:

$$(f * g)_i \stackrel{\text{def}}{=} \sum_j f_j * g_{i+j} \quad (2)$$

$$f = v1_{k:k+w}, g = v2_{k:k+w} \text{ for all } k \text{ in } v1, v2$$

where  $v1$  is the first MFCC vector,  $v2$  is the second MFCC vector,  $k$  is the index and  $w$  is the size of the sliding window.

As temporal information has proven to be useful in speech classification problems [5, 6], a window size of 450ms has been used for  $w$ . This was determined based on average phoneme duration being around 176ms [20]. As such, a frame size of 450ms is therefore long enough to account for multiple phonemes, thus avoiding false classification of brief speech-like phenomena, but still allowing for the detection of finer resolution (sub-1s duration) speech features.

### D. Classification and Tuning

Classification is achieved through the use of a learning machine trained on the MFCC-cross-covariance (MFCC-CC) features from the annotated training data set. In this case, random forests were chosen as the classifier based on their strong performance in speech classification applications [6, 21]. The random forest classifier was investigated using varying numbers of MFCC-CC features and a range of estimators (trees per forest) in order to determine optimal parameters for classification.

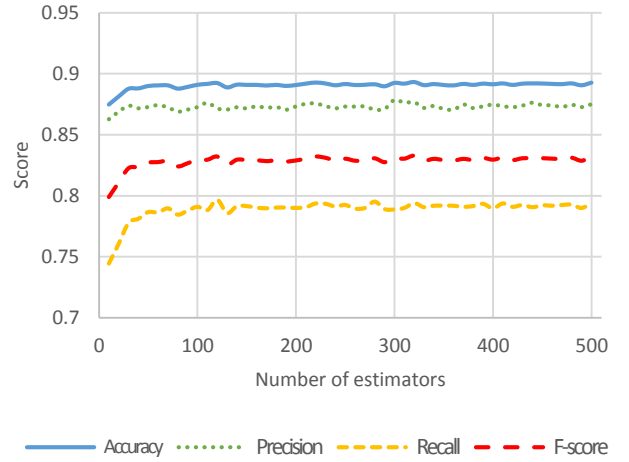


Figure 2. Random forest classification results using a range of estimators.

While testing numbers of estimators, it was found that the performance metrics stabilize after approximately 150 estimators (Figure 2), with little-to-no performance advantage achieved after this. Furthermore, previous work on random forest-based speech classifiers has demonstrated that optimal performance is achieved with the use of 200 estimators [6]. As such, the number of estimators for the random forest classifier was set at 200.

To test the impact of the number of MFCC-CC vectors used, vectors were added in order of significance, with the most significant relating to the MFCC pair with the greatest correlation coefficient difference across speech and non-speech data. Results from this (Figure 3) demonstrate that classification performance improves dramatically up to three features, and stabilizes at around five features. Therefore, five features were chosen as the optimal setting, as there was negligible gain in performance after this point.

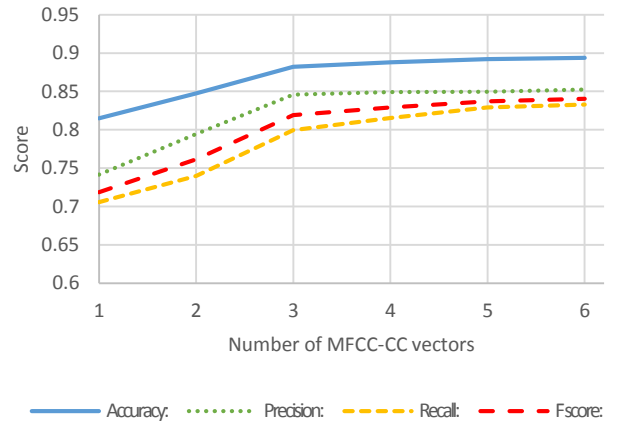


Figure 3. Random forest classification results with escalating numbers of MFCC-CC vectors.

#### IV. CASE STUDY DESIGN

Unlike other forms of data used within voice activity detection tasks, such as speech recordings in various acoustic environments [8] or synthesized acoustic environments, film is unique in that it is intentionally mixed [9]. While it may be intuitive to assume that this would make speech detection a simpler task (as the speech is mixed to be intelligible), this has proven not to be the case when testing a number of state-of-the-art voice activity detection algorithms on data from feature films [5]. This suggests that the intentional mixing of film audio separates it from audio data used in other typical VAD scenarios. As such, we have focused solely on the use of audio data from film – ensuring that both the training and test sets use intentionally mixed audio.

Two test scenarios have been used. The first uses a data set consisting of 120 minutes of data taken from four 30 minute segments of four feature films. To maximize usefulness, a cross-validation approach is used, whereby the data is reconfigured four times for each test. Each iteration uses 90 minutes of data for the training set (from three films), and 30 minutes of data for the test set (from the remaining film). This ensures that the classifier is naïve to the test data and maximizes testing cycles for the validation test set. The second test scenario uses all 120 minutes of data from the cross-validation set for training, and uses the films detailed in [5] as test data. This has been done to provide a direct comparison between the MFCC-CC approach, the approach from [6] and the results described in [5] (which includes results from testing the VAD described in [22] on feature film data).

The data has been manually annotated to provide a human-defined ground-truth, whereby sections are labeled as either speech or non-speech. The non-speech content consists of various audio mixtures including: silence, traffic noise, crowd noise, gunfire, engine noise, music (with and without singing), and other synthetic sound effects and sound design components. The speech content contains a number of speech varieties, including: speaking (various volumes), whispering, and shouting. Speech content is also mixed with the range of background audio (similar to that described for the non-speech content). The degree of variation in both speech and non-speech samples is pseudo-random according to individual film content.

#### V. RESULTS

##### A. Initial Testing Results

Initial testing indicated strong performance of the MFCC-CC classifier, with an average accuracy of 89.2% (see Table I). Strong performance was also observed when testing on an animated feature film using training data from non-animated content. This indicates that the approach is capable of handling atypical speech characteristics, as the animated content contains a significant amount of extreme/accuated voice characteristics, for example the voice of the *Gingerbread Man* character in *Shrek 3*.

TABLE I. CLASSIFICATION RESULTS FROM RANDOM FOREST CLASSIFIER TRAINED ON MFCC-CC FEATURES

Test set / Genre	Accuracy	Precision	Recall	Fscore
<i>Constantine</i> / action/horror	0.903	0.902	0.794	0.844
<i>Shrek 3</i> / animated/fantasy	0.861	0.792	0.789	0.790
<i>Knocked Up</i> / romantic comedy	0.881	0.783	0.889	0.833
<i>Blood Diamond</i> / drama/thriller	0.924	0.920	0.845	0.881
<i>Mean</i>	0.892	0.849	0.829	0.837

The MFCC-CC classifier was also evaluated using receiver operating characteristics (ROC), a common method of assessing binary classifier performance. The ROC curves in Figure 4 indicate strong performance, with an average area under curve (AUC) of 0.955 (see Table II), indicating that the classifier exhibits strong discrimination between the two classes. The equal error rate (EER) observed here further indicates strong system accuracy, with an average EER of 11.1% achieved across the four test scenarios. This suggests better performance than the VAD in [5], which achieved an average EER of 33.2% on film audio data.

To assess performance with respect to [6], an implementation of the classifier used by Sonnleitner *et al.* was trained and tested using the cross-validation approach described in section IV. In [6] a median filter is used on the classification output. To assess equivalent performance, the median filter is not applied here, as a median filter has not been used on the MFCC-CC classifier output. Thus, only the raw classifier output is considered.

TABLE II. AUC AND EER FROM RECEIVER OPERATING CHARACTERISTIC PLOT

	<i>Const.</i>	<i>Shrek 3</i>	<i>Kno. Up</i>	<i>Bl. D.</i>	<i>Mean</i>
<i>AUC</i>	0.969	0.925	0.954	0.973	0.955
<i>EER [%]</i>	9.5	15.0	11.6	8.1	11.1

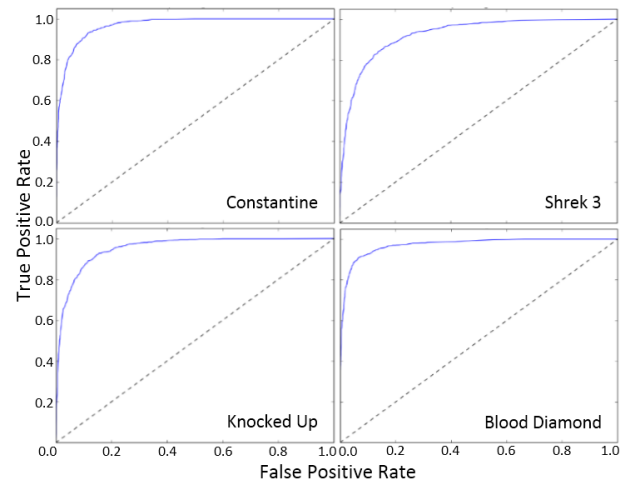


Figure 4. Receiver operating characteristic curves for MFCC-CC classification results from initial testing.

TABLE III. CLASSIFICATION RESULTS FROM RANDOM FOREST CLASSIFIER TRAINED ON FEATURES DESCRIBED IN [6]

Test set / genre	Accuracy	Precision	Recall	Fscore
Constantine / action/horror	0.714	0.642	0.315	0.423
Shrek 3 / animated/fantasy	0.701	0.642	0.228	0.337
Knocked Up / romantic comedy	0.678	0.539	0.224	0.317
Blood Diamond / drama/thriller	0.701	0.637	0.236	0.344
Mean	0.699	0.615	0.251	0.355

As demonstrated when comparing Table I and Table III, the MFCC-CC approach achieves greater results across all performance statistics used for evaluation, thus early investigations indicated that the proposed MFCC-CC features are more effective for speech classification when compared to the feature proposed in [6].

### B. Further Testing Results

Further investigations applied the MFCC-CC approach to whole feature films in order to provide a more comprehensive evaluation of its performance with respect to existing methods. The methods used for comparison were a long-standing state of the art VAD approach used to provide baseline performance statistics [22], as well as approaches that have demonstrated strong performance on entertainment media [5][6].

Results in Table IV indicate that the approach from [6] demonstrated competitive performance against both [5] and [22], however the MFCC-CC approach exceeds the performance of all methods investigated, with greater AUC values for all test sets and lower EER.

TABLE IV. COMPARISON OF VAD APPROACHES

Test set	AUC			
	[5]	[22]	[6]	MFCC-CC
<i>I Am Legend</i>	0.704	0.567	0.718	0.921
<i>Kill Bill Vol. 1</i>	0.627	0.554	0.800	0.893
<i>Saving Private Ryan</i>	0.743	0.577	0.717	0.946
<i>The Bourne Identity</i>	0.685	0.603	0.730	0.977
Mean	0.690	0.575	0.741	0.934
[%]	EER			
ALL	33.18	45.73	31.41	13.49

TABLE V. PERFORMANCE STATISTICS OF MFCC-CC APPROACH AND CLASSIFIER FROM [6] WHEN APPLIED TO WHOLE-FEATURE-FILM DATA SET

Test set	Accuracy		Precision		Recall		Fscore	
IAL	<b>0.88</b>	0.81	<b>0.62</b>	0.47	<b>0.81</b>	0.17	<b>0.70</b>	0.25
KB.I	<b>0.84</b>	0.79	<b>0.64</b>	0.62	<b>0.72</b>	0.26	<b>0.68</b>	0.37
SPR	<b>0.87</b>	0.77	<b>0.91</b>	0.45	<b>0.66</b>	0.29	<b>0.77</b>	0.35
TBI	<b>0.94</b>	0.76	<b>0.88</b>	0.45	<b>0.88</b>	0.25	<b>0.87</b>	0.32
Mean	<b>0.88</b>	0.78	<b>0.76</b>	0.50	<b>0.77</b>	0.24	<b>0.75</b>	0.32

Left columns (bold): MFCC-CC results. Right columns: results from approach described in [6]

Table V provides a more detailed performance comparison of the MFCC-CC approach and [6] (as this demonstrated the most competitive results in Table IV). The MFCC-CC approach demonstrates some reduced performance when compared to the initial testing results in Table I, however, this was anticipated given the limited training set and larger test set. Despite this, the approach continues to exhibit competitive results, outperforming [6] across all performance metrics. In particular, it can be seen that while the approach from [6] demonstrates relatively strong accuracy scores, significantly greater F-score values for our approach can be observed, indicating more robust performance.

## VI. CONCLUSIONS AND FUTURE DIRECTION

The results presented here demonstrate strong performance of the proposed MFCC-CC speech detection approach, yielding performance metrics which exceed those of state of the art and other contemporary VAD approaches applied to feature film audio data. While these results are encouraging, more comprehensive testing is underway in order to gain further insight into the performance of the proposed approach on a larger data set. Given the small size of the training set used here (120 minutes), it will be particularly interesting to investigate the effects of more training data on classification performance, and to explore VAD performance on a greater variety of film genres and across multiple languages.

Further investigations will also explore the use of MFCC-CC features with other classifiers, such as support vector machines, and will examine the possibility of expanding the feature selection method to explore whether genre-specific MFCC feature pairs can be utilized to enhance classifier performance. The long-term goal of this work is to apply audio speech detection in combination with visual features to gain a better understanding of their associations and to develop automated solutions for film post-production workflows.

## REFERENCES

- [1] M. Benatan and K. Ng, "Multimodal Feature Matching for Event Synchronization," in Proceedings of the 19<sup>th</sup> International Conference on Distributed Multimedia Systems, Brighton, UK, 2013.
- [2] Z. Rasheed, Y. Sheikh and M. Shah, "On the Use of Computable Features for Film Classification," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, 2005, pp. 52-64.
- [3] J. Bergstra, M. Mandel and D. Eck, "Scalable Genre and Tag Prediction with Spectral Covariance," in Proceedings of the 11<sup>th</sup> International Society for Music Information Retrieval Conference, Utrecht, Netherlands, 2010.

- [4] D. P. W. Ellis, X. Zeng and J. H. McDermott, "Classifying Soundtracks with Audio Texture Features," in the 36<sup>th</sup> International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic, 2011.
- [5] F. Eyben, F. Weninger, S. Squartini and B. Schuller, "Real-life Voice Activity Detection with LSTM Recurrent Neural Networks and an Application to Hollywood Movies," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, USA, May 2013, pp. 483-487.
- [6] R. Sonnleitner, B. Niedermayer, G. Widmer and J. Schlüter, "A Simple and Effective Spectral Feature for Speech Detection in Mixed Audio Signals," in Proceedings of the 15<sup>th</sup> International Conference on Digital Audio Effects, York, UK, 2012.
- [7] L. Lu, H-J Zhang and H. Jiang, "Content Analysis for Audio Classification and Segmentation," in IEEE Transactions on Speech and Audio Processing, vol. 10, October, 2002, pp. 504-516.
- [8] J. Bach, J. Anemüller and B. Kollmeier, "Robust Speech Detection in Real Acoustic Backgrounds with Perceptually Motivated Features," in Speech Communication, vol. III, May, 2011, pp. 690-706.
- [9] T. Holman, Sound for Film and Television. Kidlington, Oxford: Elsevier, 2010.
- [10] R. Viers, The Sound Effects Bible. Studio City, CA: Michael Wiese Productions, 2008.
- [11] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," in IEEE Transactions on Acoustic, Speech and Signal Processing, vol. 28, 1980, pp. 357-366.
- [12] A. Tsiartas, P. Ghosh, P.G. Georgiou and S. Narayanan, "Bilingual Audio-subtitle Extraction Using Automatic Segmentation of Movie Audio," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Prague, Czech Republic, 2011.
- [13] T. Ramalingam and P. Dhanalakshmi, "Speech/Music Classification Using Wavelet Based Feature Extraction Techniques," Journal of Computer Science, vol. 10, 2014, pp. 34-44.
- [14] J. Pinquier and C. Senac, "Speech and Music Classification in Audio Documents," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, USA, 2002.
- [15] T. Drugman and Y. Stylianou, "Fast Inter-Harmonic Reconstruction for Spectral Envelope Estimation in High-Pitched Voices," in IEEE Signal Processing Letters, vol. 21, July, 2014, pp. 1418-1422.
- [16] W. Ghai and N. Singh, "Literature Review on Automatic Speech Recognition," in International Journal of Computer Applications, vol. 41, 2012, pp. 42-50.
- [17] Y. X. Zou, W. Q. Zheng, W. Shi and H. Liu, "Improved Voice Activity Detection Based on Support Vector Machine With High Separable Speech Feature Vectors," in Proceedings of 19<sup>th</sup> International Conference on Digital Signal Processing, Hong Kong, China, 2014.
- [18] F. Sheng, G. Zhang and Z. Song, "Comparison of Different Implementations of MFCC," in the Journal of Computer Science and Technology, vol. 16, Sept, 2001, pp. 582-589.
- [19] W. Li and H. Bourlard, "Sub-Band Based Log-Energy and its Dynamic Range Stretching for Robust In-Car Speech Recognition," in Proceedings of the 13<sup>th</sup> Annual Conference of the International Speech Communication Association, Portland, OR, USA, September, 2012.
- [20] E. M. Mugler, J. L. Patton, R. D. Flint, Z. A. Wright, S. U. Schuele, J. Rosenow, J. J. Shih, D. J. Krusienski and M. W. Slutzky, "Direct Classification of All American English Phonemes Using Signals from Functional Speech Motor Cortex," in Journal of Neural Engineering, vol. 11, June, 2014.
- [21] Y. Su, F. Jelinek and S. Khudanpur, "Large-Scale Random Forest Language Models for Speech Recognition," in Proceedings of Interspeech, Antwerp, Belgium, 2007.
- [22] J. Sohn and N. Kim, "A Statistical Model-based Voice Activity Detection," IEEE Signal Processing Letters, vol. 6, January, 1999, pp. 1-3.

# A Symbolic Representation of Motion Capture Data for Behavioral Segmentation

Ruxiang Wei, Weibin Liu  
Institute of Information Science  
Beijing Jiaotong University  
Beijing 100044, China  
e-mail: wblu@bjtu.edu.cn

Weiwei Xing  
School of Software Engineering  
Beijing Jiaotong University  
Beijing 100044, China

**Abstract**—For building and understanding computational models of human motion, behavioral segmentation of human motion into actions is a crucial step, which plays an important part in many domains such as motion compression, motion classification and motion analysis. In this paper, we present a novel symbolic representation of human motion capture data, called the Behavior String (BS). Based on the BS, a further motion segmentation algorithm for human motion capture data is proposed. The human motion capture data is treated as a high-dimensional discrete data points, which are clustered by an alternative algorithm based on density, and each cluster is divided into a character. Then, the BS is produced for the motion data by temporal reverting. By analyzing the BS, the human motion capture data is segmented into distinct behavior segments and the cycles of motion are found. Experiments show that our method not only has a good performance in behavioral segmentation for motion capture data, but also finds cycles of motion and the motion clips of the same behaviors from long original motion sequence.

**Keywords**—motion analysis; behavioral segmentation; clustering; motion capture data; cycle

## I. INTRODUCTION

Computer animation technology has been extensively employed in film and television production, entertainment and games, virtual reality, and many other fields [1]. Human animation is the most challenging topics in computer animation. Motion capture technique has become a major means of human body animation production which has been applied extensively in recent years. For the benefit of motion compression, motion classification and motion analysis in computer graphics, behavioral segmentation is part of the hot research topics, the motivation of which is to get the independently semantic subsequences from long original motion sequence [2], as shown in Fig. 1. It is very tedious to segment long sequences by hand. Therefore, the research of motion capture data segmentation algorithm is of major importance. This paper focuses on efficient and quite robust technique which is able to automatically create

a segmentation even when the behaviors have not been seen before.

A new symbolic representation of human motion capture data for behavioral segmentation is proposed from this paper, which provides two key abilities: the ability to gets a higher rate of recall and precision, the ability to finds the cycles of motion and extracts the motion clips which represent the same behaviors from long original motion sequence. Our method can be simply described as follows: Each frame in the motion capture data is considered as a point. Thus, human motion capture data is treated as a high-dimensional discrete data points. These points are clustered by an alternative algorithm based on density [3]. Then, the points reordered according to the order of the original frames, and different clusters are indicated by different letters. Thus, a symbolic representation of motion capture data is produced which is called the BS. The human motion capture data is segmented into distinct behavior segments and the cycles of motion are discovered by analyzing the BS. From experimental results, we have found that this algorithm has a very good performance in behavioral segmentation for motion capture data.

The paper is organized as follows. In Section II, we discuss related work. In Section III, we detail how to create the BS including computing the distance, clustering and the parameters setting. In Section IV, we describe the analyzing of BS, so that the segmentation points and the cycles will be discovered. In Section V, we analyze experimental results and compare with other state-of-the-art motion segmentation methods. Finally, in Section VI we provide the conclusions and provide directions for future work.

## II. RELATED WORK

Finding the segmentation points of different behaviors for virtual humans with motion capture data is a challenging task that has seen much work over the last decade. Numerous researches have made efforts and a series of achievements, but still have some problems.

A few years ago, for the method based on classifier, Arian et al. [4] constructed Support Vector Machine (SVM) to realize human motion segmentation with the manually annotation training database. But this method depended on large training

---

This research is partially supported by National Natural Science Foundation of China (No. 61370127, No.61100143, No.61473031, No. 61472030), 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). The opinions expressed are solely those of the authors and not the sponsors.  
{ Corresponding author: Weibin Liu, wblu@bjtu.edu.cn }

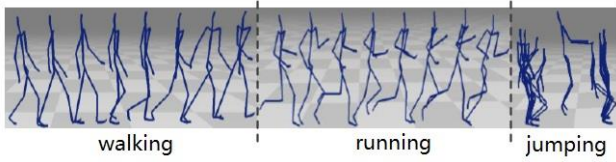


Figure 1. Segmenting motion capture data into different behaviors.

datasets and empirical value. For the method based on dimension reduction, Barbic et al. [5] assumed that different human motion behaviors can be represented with different intrinsic dimensionality, and realized the behavioral segmentation by subspace analysis of PCA theory. For the longer human motion data, however, the recall and precision through this method were lower [5]. For the method based on dimension model, Lu et al. [6] presented a two-threshold, multidimensional segmentation algorithm to automatically decompose a complex motion into a sequence of simple linear dynamic models. However, the scope of application was restricted. This method could only process human motion data with repetition period [6]. Barbic et al. [5] proposed PPCA segmentation method, which assumed that different motion classes belonged to different Gaussian distributions and could assign behavioral cuts with Mahalanobis distance. In addition, Gaussian Mixture Model (GMM) is utilized to cluster motion sequences with distinct motion classes. Nevertheless, users needed to set the number of motion classes beforehand.

In recent years, for the method based on clustering, Zhou et al. [7] used Aligned Cluster Analysis (ACA) method to segment human motion data, measured the similarities among diverse human motion sequences through dynamic time warping kernels, and accomplished motion segmentation with kernel k-means classification. However, users required to determine the cluster number with respect to temporal constraint [7]. In the past two years, Zhou et al. [8] utilized HACA to combine kernel k-means with the generalized dynamic time alignment kernel to cluster time series data. Moreover, HACA provides a natural framework to locate a low-dimensional embedding for the time series. HACA is efficiently optimized with a coordinate descent strategy and dynamic programming. While this method depended on the choice of the kernel parameters and the functional form of the kernel [8]. Yang et al. [9] presented a

symbolic representation of motion capture data which called the Motion String (MS). MS was used to segment human motion data. But users needed to set the number of motion classes beforehand. On the other hand, this algorithm could only segment of the human motion data which have small changes between each behaviors, and could not segment the data which have complex behaviors [9]. Beaudoin et al. [10] developed a string-based motif-finding algorithm which allows for user-controlled compromise between motif length and the number of motions in a motif. Motion motifs represent clusters of similar motions and together with their encompassing motion graph they lend understandable structure to the contents and connectivity of substantial motion datasets.

To solve the shortcomings and limitations of existing motion segmentation methods, inspired by the Motion String [9] and Motion-Motif Graph [10]. This paper proposes a novel symbolic representation of human motion capture data, which called the Behavior String (BS). Based on the BS, a further motion segmentation algorithm for human motion capture data is proposed. This algorithm is able to automatically find the number of cluster centers, while the Motion String algorithm needs the user to set the number of motion classes beforehand which is impractical. Our method can also find the cycles of motion and the motion clips of the same behaviors from long original motion sequence which the Motion String cannot do. Distinct from the Motion String which treats the string to periodic substrings, static substrings and transitional substrings [9]. In this paper, a string matching method is presented to find cut points and the cycles of motion, which regards each letter as a part of a meaningful behavior rather than a transition between different behaviors.

### III. CREATING THE SYMBOLIC REPRESENTATION

In this section, the procedure of creating the BS which is a symbolic representation of motion capture data is presented, as shown in Fig.2. First, compute the similarity between every two points which are regarded as the frames in the motion capture data sequences. Second, calculate the local density and the distance for the clustering method. Third, find cluster centers which have a large local density and distance at the same time. Finally, cluster each remaining point which is not the cluster centers and recorded all points with the time sequence.

#### A. Computing similarity between points

To build the behavior strings, we need to cluster each frame from one of the human motion capture data sequences which also called the long original motion sequence. We regard each frame as a high-dimensional point, and compute the distance between every two points. Lee et al. [1] presented an algorithm to calculate the distance between two frames from the motion sequence and Wang et al. [2] improved this algorithm. But these algorithms only compute the principal joints and ignore other joints. So that these algorithm cannot distinguish the relatively complex behaviors. Therefore, we use the algorithm [11] which compute not only the difference of pose but also the velocity between the two frames.

This paper employs the following human skeleton model which has 31 joints shown in Fig.3. There are 62-dimensional in one frame, including root position vector, root orientation vector

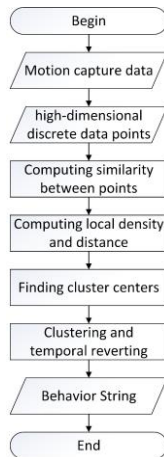


Figure 2. The procedure of creating the BS.



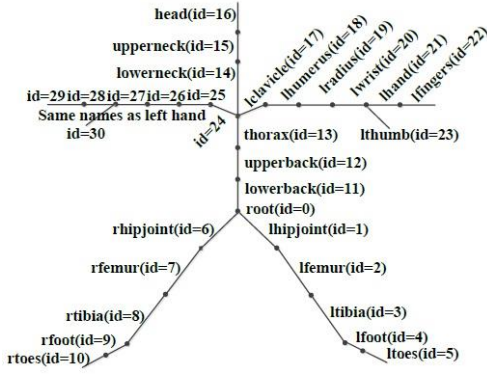


Figure 3. The human skeleton model.

and other joints' direction vector. The  $i$ th frame's pose consists of all joints' rotation angle in the  $i$ th frame except the root position vector and the root orientation vector which including 6-dimensional. Each pose  $p_i = \{a_{i,1}, a_{i,2}, a_{i,3}, \dots, a_{i,56}\}$  is represented as a point in 56-dimensional space, where  $a_{i,m}$  is one of an Euler angle. The  $i$ th frame's velocity  $v_i$  is computed by the Euclidean distance between  $p_{i+1}$  and  $p_i$ . Especially the last frame's velocity  $v_n$  is equal to the last but one frame's velocity  $v_{n-1}$ .

$$v_i = \begin{cases} \sqrt{(a_{i+1,1} - a_{i,1})^2 + (a_{i+1,2} - a_{i,2})^2 + \dots + (a_{i+1,56} - a_{i,56})^2}, & i \neq n \\ v_{i-1}, & i = n \end{cases} \quad (1)$$

We calculate the distance by:

$$d_{ij} = \alpha p_{ij} + \beta v_{ij} \quad (2)$$

Where the  $p_{ij}$  is the Euclidean distance of pose between the  $i$ th frame and the  $j$ th frame. The  $v_{ij}$  is the difference of velocity between the  $i$ th frame and the  $j$ th frame. The  $\alpha$  and the  $\beta$  are the weights. In our experiments, the  $\alpha$  and the  $\beta$  is equal to 1. With this method, we compute the all distance between any two frames. Then we get the distance matrix  $D_{n \times n}$ , where  $n$  is the length of the long original motion sequence. Obviously,  $d_{ij} = d_{ji}$  ( $i \neq j$ ), and  $d_{ij} = 0$  ( $i = j$ ).

### B. Local density and distance for clustering

This clustering approach based on the idea that cluster centers are characterized by a higher density than their neighbors

and by a relatively large distance from points with higher densities [3].

For each point  $i$ , we compute two quantities: its local density  $\rho_i$  and its distance  $\delta_i$  from points of higher density. Both of the quantities only depend on the distances  $d_{ij}$  between every two points. We use the Gaussian kernel function [12] [13] to compute the local density  $\rho_i$ :

$$\rho_i = \sum_j e^{-\left(\frac{d_{ij}}{d_c}\right)^2} \quad (3)$$

Where  $d_c$  is a distance which is the parameter we need to set in this clustering algorithm. The  $\rho_i$  is larger when there are more points closer than  $d_c$  to point  $i$ .

We reorder the local density  $\{\rho_i\}_{i=1}^n$  in descending order, and use  $\{q_i\}_{i=1}^n$  to indicate the local density's subscript. Thus  $\rho_{q_1} \geq \rho_{q_2} \geq \dots \geq \rho_{q_n}$ . The distance  $\delta_i$  can be defined as:

$$\delta_{q_i} = \begin{cases} \min_{j < i} \{d_{q_i q_j}\}, & i \geq 2 \\ \max_{j \geq 2} \{\delta_{q_j}\}, & i = 1 \end{cases} \quad (4)$$

Obviously if the point  $i$  has the biggest local density, the distance  $\delta$  is the maximum distance of other points. Otherwise, the distance  $\delta$  is the minimum distance from points with higher densities.

### C. Finding cluster centers

According to our assumption in the beginning of 3.2. Cluster centers have a large  $\rho$  and  $\delta$  at the same time. Considering the  $\rho$  and  $\delta$  may have a different order of magnitude. We multiply  $\rho$  by  $\delta$  after normalization, and let the results denoted by  $\gamma_i = \rho_i \times \delta_i$ . Then we reorder the  $\{\gamma_i\}_{i=1}^n$  in descending order, and let  $\{\gamma_j\}_{j=1}^n$  indicate this sequence. Thus, cluster centers are recognized as points for which the value of  $\gamma$  is anomalously large. The first  $j$  points which can be found by the maximum  $j$  satisfying  $\gamma_j - \gamma_{j+1} > \theta$ . The  $\theta$  is a threshold which is set to 0.05 in our Experiments.

We select a long original motion sequence from the CMU database [14], which has 2500 frames including 'running' and 'walking'. Fig. 4(a) shows the  $\rho$  and  $\delta$  computed from these

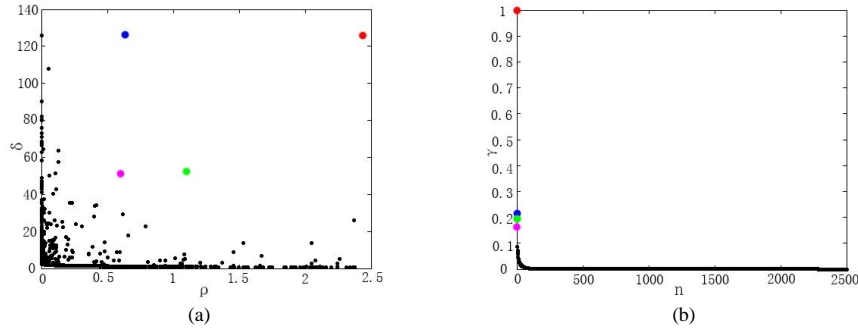


Figure 4. Cluster analysis of the human motion capture data from CMU Database. (a) The  $\rho$  and  $\delta$  computed from 2500 frames that selected from the CMU database. (b) The value of  $\gamma_i = \rho_i \times \delta_i$  in decreasing order for the data in (a). Colored points correspond to cluster centers.

$q_i$ (local density in descending order)	1	2	3	4	5	6	7	8	...
$b_{q_i}$ (larger local density and minimum distance)	0	1	1	1	3	3	4	6	...
cluster center? (Y/N)	Y	N	Y	N	N	Y	N	N	...
cluster's label	A	A	B	A	B	C	A	C	...

Figure 5. Clustering other points except the cluster centers. Each remaining point is assigned to the same cluster as its nearest neighbor of higher density.

frames. Fig. 4(b) shows the  $\{\gamma_j\}_{j=1}^{2500}$  in descending order. Colored points correspond to cluster centers. Notice that there are 4 cluster centers, each cluster center is not indicate each behavior, but represent the feature of behaviors.

#### D. Clustering and temporal reverting

After the cluster centers have been found, each remaining point is assigned to the same cluster as its nearest neighbor of higher density, shown in Fig. 5. A detailed description of this procedure is as follows. We define the  $b_i$  as:

$$b_{q_i} = \begin{cases} \arg \min_{j < i} \{d_{q_i q_j}\}, & i \geq 2 \\ 0, & i = 1 \end{cases} \quad (5)$$

The  $\{b_i\}_{i=1}^n$  indicate a sequence number of the point which has a larger local density than  $i^{\text{th}}$  point and a minimum distance. The  $\{q_i\}_{i=1}^n$  is defined in 3.2 which indicate the subscript of local density. First, we classify the cluster centers. Then, each remaining point will be classified in the cluster that belongs to  $b_{q_i}$ . For example, if the  $q_m^{\text{th}}$  point is not the cluster center, the cluster of  $q_m^{\text{th}}$  point is assigned to the cluster of  $b_{q_m}^{\text{th}}$  point. Finding the cluster according to  $\{q_i\}_{i=1}^n$  which indicate the subscript of local density assure the  $b_{q_i}^{\text{th}}$  point has a definite cluster, because the point which has a largest global density is certainly a cluster center.

After the clustering, each frame which is regarded as a high-dimensional point has a cluster's label, like A, B or C. These letters are recorded according to the order of the original frames, and this step is called the temporal reverting. The motions can

now be converted into strings using the letters associated with each frame. This string can be further simplified by removing consecutive repetitions of the same letter, and the length of sequential repetitions saved by the subscript of each letter. For instance, the string  $\{AAAABBCCCC\}$  can be simplified to  $\{A_4B_2C_3\}$ . This symbolic representation which called the Behavior String (BS) represents human motion capture data. For example, the sequence which have 1200 frames including 'running' and 'walking' can be shown as:

$\{A_{100}B_{100}A_{100}B_{100}A_{100}B_{100}C_{150}D_{150}C_{150}D_{150}\}$ . By analyzing this Behavior String,  $\{A_{100}B_{100}\}$  is a circular sequence meaning a motion 'x', while  $\{C_{150}D_{150}\}$  is a circular sequence meaning other motion 'y'. There are three 'x' and two 'y' in this sequence. The cycle of 'x' is 200 frames, while the cycle of 'y' is 300 frames.

#### IV. ANALYZING BS FOR BEHAVIORAL SEGMENTATION

How to get the cut points and cycles of motion from BS are presented in this section. Extract the 'Key Word' which represent the behavior and use the string matching method to find the cut points and the cycles of motion.

##### A. Extracting 'Key Word' representing the behavior

We use the Moving Window to get the statistics of 'words'. In our experiments, a behavior can be converted to the 'word' which has at most three letters, like 'AB', 'BC' or 'ADC'. As shown in Algorithm 1. Setting the step size to one, from the first frame to the last frame, the 'word' will be counted with the window size of 2. Then we do the same work with the window size of 3. Specially, we ignore the 'word' which has a same letter in it, like 'ABA' or 'CDC'. Note that the 'word' which is constituted by the letters in the window does not consider the order of the letters. For example, 'AB' is equal to 'BA', and 'ABC' is equal to 'BCA'. If the number of the 'word' is less than 3, which as a rule of thumb in our experiments, the 'word' will be also removed, because the 'word' is a transition between two different behaviors. Moreover, if letters from one 'word' are

Algorithm 1. The 'Key Word' discovery algorithm.

---

**Input:**  
N is the length of the BS,  
Vector BS.

**Output:**  
Vector KeyWord.

**Begin:**  
Window size = 2 or 3;  
**for** ( i = first frame to last frame ) **do**  
  Compare the Word in the neighbour Window;  
  Count the Word with 2 or 3 character;  
  Save the Word and the number of Word;  
**end if**  
**for** ( i = 1 to the number of Word ) **do**  
  Delete the smaller number of the similar Word;  
**end if**  
**for** ( i = 1 to the number of the simplified Word ) **do**  
  **if** ( the number of Word[i]  $\geq 3$  )  
    Keyword[i] = Word[i];  
  **end if**  
**end for**

---

**End**

---

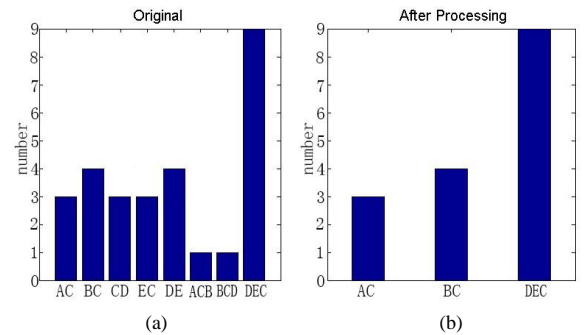


Figure 6. Finding 'Key Word'. (a) The statistics of 'word'. (b) The 'word' after processing is called the 'Key Word'.

included in another ‘word’, save the ‘word’ which has largest counts and abandon others. For example, to facilitate the expression, let the  $N_{AB}$  to indicate the number of ‘AB’. If the Behavior String is {ABCABCBCDEDEDE} which hid the subscripts, we only keep the ‘ABC’ and ‘DE’, because the  $N_{AB} = 3$ ,  $N_{BC} = 3$  and  $N_{AC} = 2$  are less than  $N_{ABC} = 7$ , while  $N_{BCD} = N_{CDE} = N_{CD} = 1$  are less than 3 meaning the transition between two different behaviors. The remaining ‘words’ after the aforementioned processing are called the ‘Key Words’.

Fig. 6 shows the ‘Key Word’ found from the statistics of the ‘words’ in Behavior String:

{ $A_{500}C_{99}A_{398}C_{195}B_{127}C_{217}B_{118}C_{286}D_{34}E_{60}C_{51}D_{41}E_{68}C_{42}D_{37}E_{70}C_{49}D_{52}E_{56}$ }

Which have 2500 frames from the CMU database [14] including ‘walking’, ‘stretching’ and ‘swinging arms’, represented by ‘AC’, ‘BC’ and ‘DEC’. In Fig. 6(a), there are 8 ‘words’ computed by the statistic of the ‘words’ in Behavior String. As we can see, the number of ‘ACB’ and ‘BDC’ is less than 3, so these ‘words’ are regarded as the transition of two behaviors, in other words, these two ‘words’ are removed. The ‘CD’, ‘EC’ and ‘DE’ are included in ‘DEC’ which not consider the order of the letters. According to our algorithm, the largest number of ‘word’ is reserved, while the smaller number of ‘words’ are removed. The number of ‘DEC’ is 9 which is larger than ‘CD’, ‘EC’, and ‘DE’. So ‘DEC’ is the only one ‘word’ which is reserved. In general, there are 3 ‘words’ which is regarded as the ‘Key Words’ as shown in Fig. 6(b).

### B. Finding cut points and cycles

We use the string matching method to find the cut points and the cycles of motion. Use every ‘Key Word’ to match with the original Behavior String, and save the last letter’s frame number to the Cut Array if the next ‘word’ is not this ‘Key Word’. Considering the Behavior String may have a letter which is not included in all of the ‘Key Words’. If the subscript of the letter is more than 600, as a rule of thumb in our experiments, adding the last letter’s frame number to the Cut Array. This letter can be regarded as an independent behavior, which means each frame in this behavior is clustered in one cluster, but we cannot find its cycle. Else, we will do nothing because the letter is regarded as a noise. Each element in Cut Array adds one. Then reorder the number after the repeated numbers are removed. Cut points between two behaviors are given as a frame number in Cut Array. Next, for each ‘Key Word’, find the average length of the matched ‘word’ in Behavior String. The length is the cycle of motion represented by the ‘Key Word’.

## V. ANALYZING BS FOR BEHAVIORAL SEGMENTATION

We test our techniques on the largest freely available motion capture databases, the CMU motion databases [14], proving that our approach is not only theoretically applicable but also solves the problem of segmenting motion capture data into distinct behaviors in practice.  $d_c$  is a parameter deciding the results of the clustering which has been presented in Section III.B. In our experiments, we set the  $d_c$  so that the average number of neighbors is around 1% of the total number of points in the data set [3]. We choose 34 motion sequences with different behaviors such as walking, running, jumping, punching, stretching, and swinging arms. We use our method to segment this 34 long original motion sequences into distinct behaviors. As

TABLE I. SOME OF THE HUMAN MOTION CAPTURE DATA FOR BEHAVIORAL SEGMENTATION IN OUR EXPERIMENTS.

Serial number	Name	Number of frames	Number of behaviors
A	CMU_86_01	4579	4
B	CMU_86_03	8401	6
C	CMU_86_04	10078	6
D	CMU_86_05	8340	8
E	CMU_86_06	9939	9
F	CMU_86_07	8702	6
G	CMU_86_08	9206	9
H	CMU_86_11	5674	4

comparison, we use manual labeling method, PCA method, PPCA method [5], ACA method [7] and HACA method [8] to accomplish motion segmentation. There are 8 representative sequences as shown in Table I and the results of the behavioral segmentation using these sequences are shown in Fig. 7. The exact frame of the transition is particularly difficult to determine when the motion capture subject makes a smooth transition from one behavior to another. Consequently, we allow a range of frames to be specified as the ground truth by the human subjects. In Fig. 7, the black stripes in the sequences indicate the cut points assigned by the algorithms. Specially, for the human observer, instead of a single frame, the range in which the transition occurred is given, as all the frames in the range are acceptable positions for a motion cut. Notice that the ‘ACA’, ‘HACA’, ‘BS’ and ‘Human’ sequences in Fig. 7 have many colors, while the ‘PCA’ and ‘PPCA’ sequences only have a single color. Because the former methods can find not only the cut points but also the motion clips which represent the same behaviors in the original motion sequence, while the latter methods could only find the cut points. For one original motion sequence, the different colors correspond to different behaviors while the same colors correspond to the motion clips which represent the same behaviors in this sequence.

We compared five algorithms using the standard precision/recall framework [5]. Precision is defined as the ratio of reported correct cuts versus the total number of reported cuts. Recall is defined as the ratio of reported correct cuts versus the total number of correct cuts. The closer precision and recall are to 1, the more accurate the algorithm is. Table II gives precision and recall scores for the five algorithms.

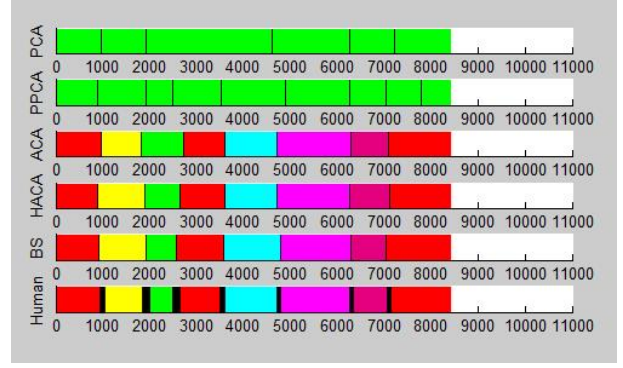
From the experimental results in Table II, we can see that the precision and recall of our method are obviously higher than the PCA method, a little higher than the PPCA method, and similar as the ACA and HACA methods. Because the clustering algorithm in this paper is able to cluster the motion capture data which have a high-dimensional and nonspherical clusters with

TABLE II. PRECISION AND RECALL SCORES FOR THE PCA, PPCA, ACA, HACA AND OUR ALGORITHMS.

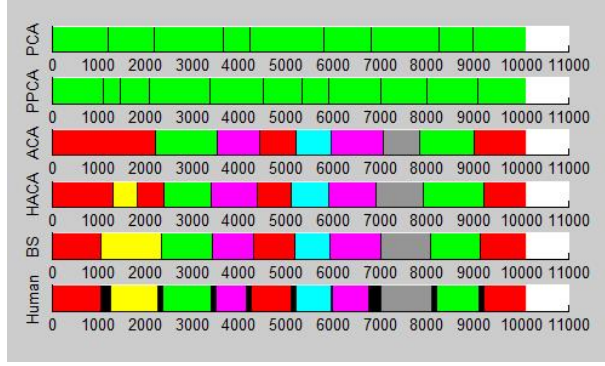
Algorithm	Precision	Recall
PCA	74.11%	80.72%
PPCA	88.30%	90.36%
ACA	90.89%	92.37%
HACA	91.12%	92.53%
Our algorithm	91.25%	92.77%



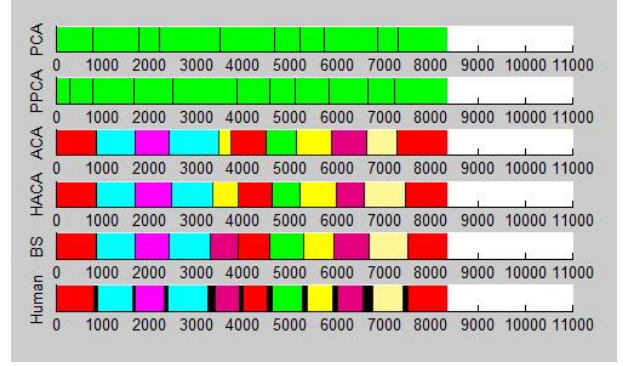
(A)



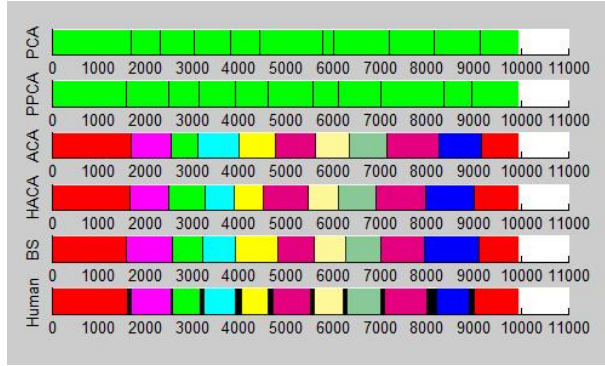
(B)



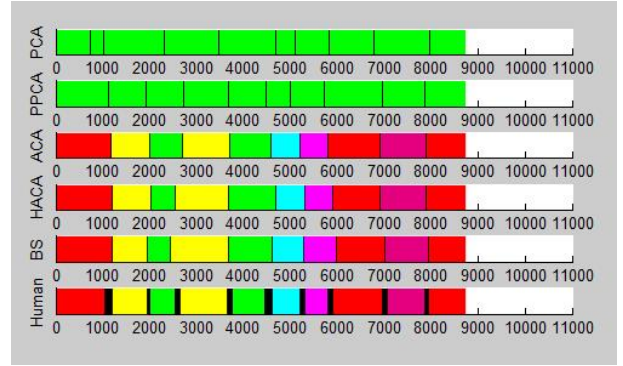
(C)



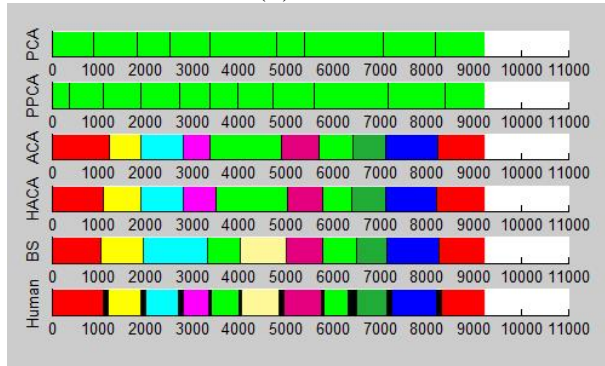
(D)



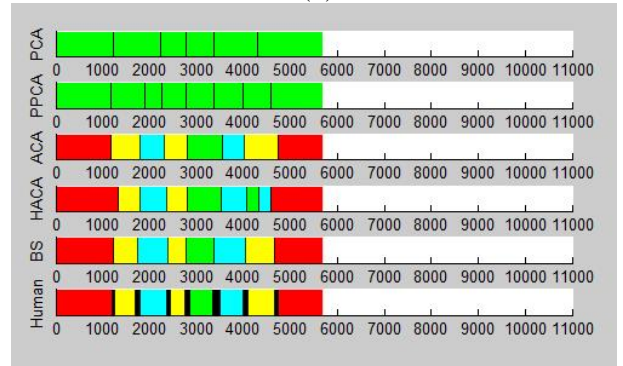
(E)



(F)



(G)



(H)

Figure 7. Motion segmentation points. Each chart corresponds to one motion from the Table I, the x-axis corresponds to the frame number, and the vertical bars specify the cut points assigned by the algorithms. For the human observer, instead of a single frame, the range in which the transition occurred is given, as all the frames in the range are acceptable positions for a motion cut. Different colors correspond to different behaviors while the same colors correspond to the motion clips which represent the same behaviors in the original motion sequence.



TABLE III. THE CYCLE OF MOTION IN THE ‘A’ SEQUENCE.

Behaviors	Cycle by BS	Cycle by human observer
walking	186 frames	170 to 190 frames
jumping	271 frames	250 to 270 frames
punching	164 frames	150 to 160 frames
kicking	287 frames	290 to 320 frames

the non-visual number of clusters. Thus, this clustering algorithm can accurately cluster the each frame into each cluster so that the cut points can be found exactly than PCA and PPCA methods compared with the human observer, shown in Fig. 7. Moreover, our segmentation method can find the motion clips representing the same behaviors which the PCA and PPCA methods cannot, because the clustering algorithm is able to cluster similar frames into one cluster. Notice that in Fig. 7(D), although ACA and HACA methods find the right cut points, they could cluster the clips which represent different behaviors into one cluster that leads to the wrong recognition of behaviors. However, in the experiments with our method, this situation not happened due to the clustering algorithm has a better performance for motion capture data.

Our method can find not only the cut points but also the cycles of motion which the other four methods cannot. Table III gives the cycle of motion computed by our method and human observer from the sequence ‘A’ in Table I. Considering the difference of each motion, like the manual labeling method to find the cut points, we also allow a range of frames to be specified as the ground truth by the human subjects.

From the experimental results in Table III, we can see that the cycle of motion computed by our method accord with the human observer. While the cycle of motion may not be found by our method when all the frames in the behavior are clustered in one cluster. For example, if an original motion sequence has one behavior which has an enormous difference to other behaviors, and each frame in this behavior has a little different to other frames, the behavior may be clustered in one independent letter in Behavior String. Essentially, the parameter  $d_c$  deciding the result of the clustering which has been set to a constant value in our experiments rather than automatically changed with each motion sequence.

## VI. CONCLUSIONS AND FUTURE WORK

This paper introduces a new symbolic representation of motion capture data called the Behavior String. The BS can be created by three steps: First, each frame in the motion capture data is regarded as a point. Thus, human motion capture data is treated as a high-dimensional discrete data points. The distances between every two points are computed by considering the pose and the velocity. Second, these points are clustered by an alternative algorithm based on the idea that cluster centers are characterized by a higher density than their neighbors and by a relatively large distance from points with higher densities, which is able to cluster the motion capture data having a high-dimensional and nonspherical clusters with the non-visual number of clusters. Third, the points reordered according to the order of the original frames, and different clusters are indicated

by different letters. Then the BS is produced by the further simplified.

Based on the BS, in the behavioral segmentation area, the behavioral cut points and the motion clips which represent the same behaviors are found with the string analysis method. This method is able to automatically create a segmentation even when the behaviors have not been seen before, and the user need not to know the number of behaviors in the original motion sequence beforehand, which make it closer to the real. Experimental results show that our method has a good performance than other state-of-the-art methods in behavioral segmentation for motion capture data. Furthermore, the cycles of motion are also found with this method which plays an important role in many domains such as motion compression, motion classification and motion analysis.

Although experimental results have shown our method provided a good performance. Our method could only segment the original motion sequence which has simple behaviors like walking, running, jumping and so on. For complicated behaviors, such as dances, this method could not distinguish them.

## REFERENCES

- [1] J. Lee, J. Chai, and P. Reitsma, “Interactive control of avatars animated with human motion data,” *ACM Transactions on Graphics (TOG)*, Vol. 21, No. 3, pp. 491-500, 2002.
- [2] J. Wang and B. Bodenheimer, “An evaluation of a cost metric for selecting transitions between motion segments,” *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*. Eurographics Association, pp. 232-238, 2003.
- [3] R. Alex and A. Laio, “Clustering by fast search and find of density peaks,” *Science*, Vol. 344, No. 6196, pp. 1492-1496, 2014.
- [4] O. Arikan, D. A. Forsyth, and J. F. O’Brien, “Motion synthesis from annotations,” *ACM Transactions on Graphics (TOG)*, Vol. 22, No. 3, pp. 402-408, 2003.
- [5] J. Barbič, et al., “Segmenting motion capture data into distinct behaviors,” *Proceedings of the 2004 Graphics Interface Conference*, London, May, 2004.
- [6] C. Lu and N. J. Ferrier, “Repetitive motion analysis: segmentation and event classification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No.2, pp. 258-263, 2004.
- [7] F. Zhou, F. Torre, and J. K. Hodgins, “Aligned cluster analysis for temporal segmentation of human motion,” *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 1-7, 2008.
- [8] F. Zhou, F. Torre, and J. K. Hodgins, “Hierarchical aligned cluster analysis for temporal clustering of human motion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 3, pp. 582-596, 2013.
- [9] Y. Yang, L. Wang, and A. H. Aimin, “Motion String: A Motion Capture Data Representation for Behavior Segmentation,” *Journal of Computer Research and Development*, Vol. 45, No.3, pp. 527-534, 2008.
- [10] P. Beaudoin, S. Coros, M. Panne, and P. Pierre, “Motion-motif graphs,” *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. Eurographics Association, pp. 117-126, 2008.
- [11] Y. Lee, K. Wampler, G. Bernstein, J. Popovic, and Z. Popovic, “Motion fields for interactive character locomotion,” *ACM Transactions on Graphics (TOG)*, Vol. 29, No. 6, pp. 138-146, 2010.
- [12] A. Shrivastava, V. Patel, and R. Chellappa, “Multiple Kernel Learning for Sparse Representation-based Classification,” *IEEE Transactions on Image Processing*, Vol. 23, No. 7, pp. 3013-3024, 2014.
- [13] S. Gao, I. Tsang, and L. Chia, “Sparse representation with kernels,” *IEEE Transactions on Image Processing*, Vol. 22, No. 2, pp. 423-434, 2013.
- [14] CMU: <http://mocap.cs.cmu.edu>, 2003.

# Shadow Detection in Complex Environments via An Innovative Information Fusion Approach

Alfredo Cuzzocrea

DIA Department, University of Trieste and ICAR-CNR, Italy  
alfredo.cuzzocrea@dia.units.it

Enzo Mumolo

DIA Department, University of Trieste, Italy  
mumolo@units.it

Alessandro Moro

Chuo University, Tokyo, Japan  
moro@sensor.mech.chuo-u.ac.jp

Kazunori Umeda

Chuo University, Tokyo, Japan  
umeda@mech.chuo-u.ac.jp

Gianni Vercelli

DIBRIS Department, University of Genova, Italy  
gianni.vercelli@unige.it

## Abstract

*In this paper a novel moving shadows detecting algorithm is proposed. The algorithm can be used in indoor and outdoor environments. The algorithm we propose fuses together color and stereo disparity information using the Dempster-Shafer combination rule. Some considerations on the nature of the shadow improves the algorithm's ability to candidate the pixels as shadow or foreground. The candidate of both color and disparity information are then weighted by analyzing the effectiveness in the scene.*

## 1. Introduction

Shadow detection plays an important role in many machine vision applications. Correct shadow detection may lead to important performance improvements in scene understanding, object segmentation, tracking, recognition.

It is not difficult for human eyes to distinguish shadows from objects. However, identifying shadows by computer is still a challenging research problem. Shadows occur when objects totally or partially occlude direct light from a light source. Generally speaking, shadows are composed by two parts: self-shadow and cast shadow. The former is the part of the object which is not illuminated by the light source. The last one is the area projected on the scene by the object and is further classified in umbra and penumbra. The umbra corresponds to the area where the direct light is totally blocked by the object, whereas in the penumbra area it is partially blocked. The cast shadow is more properly called moving cast shadow if the object is moving. Moving

shadows cause the erroneous segmentation of objects in the scene. To solve this problem, moving shadows have to be detected explicitly to prevent them being misunderstood as moving objects or their parts.

In order to systematically develop and evaluate various shadow detector, the quality measures to minimize include the following: (i) *Detection rate*, which is the error probability to detect correct shadow pixels; (ii) *Discrimination rate*, which is the probability to identify wrong points as shadow, i.e. the false alarms rate; (iii) *Localization error*, i.e. the average distance between the pixels marked as shadows and the real position of the shadow pixels.

Algorithms to solve the shadow detection problem can be coarsely divided in two groups: *property-based* algorithms and *model-based* algorithms. The most common and flexible are the property-based approaches which use features like geometry, brightness, or color to identify shadowed regions. These techniques do not use any a-priori knowledge as scene geometry, objects disposition and types, or light condition. Instead model based approaches are well suited to particular situations, as car tracking in highway, but have shown less robustness than property-based algorithms when used in a different scene and illumination conditions.

Shadow detection algorithms may be also defined in base of the main property analyzed: geometrical, luminance, color space and difference, texture analysis and edges. These basic components can be combined together to overcome the limitations offered by the methods separately. Moreover the detection of shadow can be used to reconstruct the image, not only to fastener the process. Dark shadows and soft shadows however do not change the phys-



ical dimension of an object. In this paper we propose an approach that takes in consideration the mentioned aspects, together the depth for the shadow.

## 2. Related Work

Considering the different solutions the authors proposed, we can group the works as described in the introduction.

A familiar direction is to conjecture that shadows reduce the luminance of an image, meanwhile the chrominance stays almost unchanged [20, 14]. However these approach are not valid in outdoor scenes.

Shadow analysis based on color spaces is pretty popular. Cucchiara *et al.* [21] hypothesized that shadows reduce brightness and saturation maintaining hue properties in HSV color space. More recently, Moghimi *et al.* [15], use combination of orthogonal transformation and HSV color space to detect shadow pixels. Yang *et al.* [19] use three components in YUV color space to identify shadow pixels from the candidate foreground. Dong *et al.* [25] analyze the differences between the pixels of object/shadow and that of background according to a RGB color model followed by edge ratios analysis. In [22], the authors compare shadow suppression using RGB and HSV color space and show that HSV color space should be preferred over RGB color space.

Also geometrical information had space on the shadow detection field. Many of the methods in literature normally requires shadows to be on a flat plane. The use of a disparity model has been proposed by Ivanov, *et al.*, [26]. The method described is defined as invariant to an arbitrarily rapid changes in illumination, for modeling background. The negative aspect is that, to overcome rapid changes in illumination, at least three cameras are required. Onoguchi [18] proposed a method based on two cameras to eliminate the shadows of pedestrian based on object height. In order to detect shadow, both object and shadow must be visible on the camera. Salvador, *et al.*, [9] adopt the fact that a shadow darkens the surface, to identify an initial set of shadowed pixels. This particular set is reduced by using color invariance and geometric properties of shadow.

The shadow is not only detected but it is also possible to remove. Finlayson, *et al.*, [12] utilizes shadow edges along with illumination invariant images to recover full color images. Despite that, a part of the color information is lost in removing the effect of the scene illumination. Weiss [24] uses the reflectance edges of the scene to obtain an intrinsic image without shadows. The approach proposed requires significant changes, and as result the scene illumination is contained in the reflectance image. Matsushita, *et al.*, [27] extend the previous concept. However their method does not consider dynamic cast shadows but only static.

Huerta, *et al.*, [13] apply a multi-stage approach combining color, gradient and textural information with known

shadow properties. This method improve previous models but partial loss of foreground borders due to edge and has weakness to texture-less background and objects.

## 3. Algorithm Overview

The algorithm is outlined in Fig.1. First, a segmentation of moving objects is realized using background subtraction. On the moving object extracted, a pixel-wise analysis of color consistency is performed. Difference between foreground and background distance is computed with the stereo camera. Finally, a shadow measure is computed using the Dempster-Shafer combination rule.

## 4. Shadow Properties and Models

In this work both color and distance information obtained by a stereo camera are considered. We consider fundamental radiometric model of the radiance of points in a scene illuminated by a combination of sunlight and, if present, colored direct or diffuse light like sky light (combination of multiple light sources). The model and the assumptions that are made are described in this section. We assume that the camera has a linear relationship between the radiance of a surface and the pixel value assigned to the image point of the surface. That type of camera is defined as linear camera.

For this work it is assumed that the images represent a well illuminated environment in both cases, indoor and outdoor. The material of the objects in the scene are essentially diffuse, which exhibit Lambertian reflectance, constant over time. Instead albedo's (diffuse reflectance) of the surfaces is not necessary to be constant over time. The images is supposed to be properly exposed, i.e. the important area of the image are neither over-exposed (color channel values near 255) not severely under-exposed (value near 0).

### 4.1. Color Model

It is supposed that the color information  $\rho$  at a given pixel  $p$  obtained from a recording camera depends on four components: the spectral power distribution (SPD) of the illuminant denoted  $E(\lambda)$ , the surface reflectance  $R(\lambda)$ , the sensor spectral sensitivity  $Q(\lambda)$  evaluated at each pixel  $p$  and a shading factor  $\sigma$ . This assumption is valid for Lambertian surfaces.

$$\rho_p = \sigma \int R(\lambda)E(\lambda)Q_p(\lambda)d(\lambda) \quad (1)$$

The surface reflectance  $R(\lambda)$  depends on the material which can have different albedo.

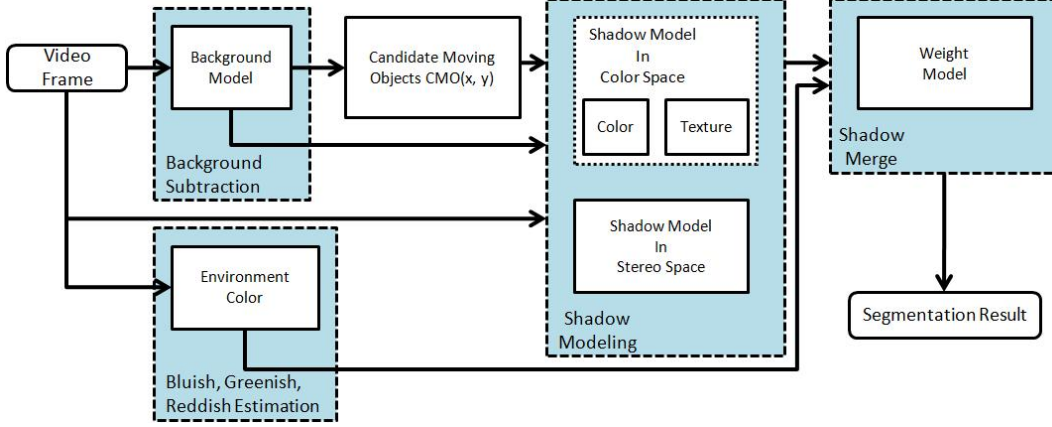


Figure 1. Flow diagram of the proposed method.

In a setting as described above it is possible to formulate the value,  $\rho$ , of a pixel as follows, using subscript  $r$  to indicate elements related to the red channel (green and blue being similar):

$$\rho_r = \frac{c_r \cdot \varphi_r \cdot E_r}{\pi} \quad (2)$$

where  $\varphi_r$  is the diffuse albedo of the surface point being imaged (ratio of outgoing radiosity to incoming radiance), and  $E_r$  is the incoming radiance in the red channel. Thus,  $\varphi_r \cdot E_r$  is the reflected radiosity. Dividing this by  $\pi$  [sr] yields the reflected radiance of the surface (since the radiosity from a diffuse surface is  $\pi$  times the radiance of the surface). Finally,  $c_r$  is the (typically unknown) scaling factor translating the measured radiance into pixel value (0 to 255 range for an 8 bit camera) for a linear camera. This scaling value depends on the aperture of the lens, the shutter speed, the gain, the white-balancing etc. of the camera.

In the kind of outdoor daylight setting we are addressing in this paper the total incoming radiance at a point is a sum of two contributions,  $E_r = E_r^{SUN} + E_r^{SKY}$ , again using subscript  $r$  for red color channel as example. The amount of radiance received from the sun,  $E_r^{SUN}$ , depends on several factors: the radiance of the sun, how large a fraction of the sun's disk is visible from the point in interest (if the sun's disk is completely occluded the point is full shadow, also called umbra), and on the angle between the surface normal at the point and the direction vector to the sun from the point. If the sun's disk is only partially occluded the point is in the penumbra (soft shadow).

Two kinds of shadows can appear in an image: the penumbra and the umbra. The difference between them can be modelled by the following equation

$$\rho(x, y) = E(x, y)\varphi(x, y) \quad (3)$$

## 4.2. Reddish, Greenish and Bluish

Many works took in consideration the diffuse sources. In particular in outdoor scenes, the diffuse source it is obviously the sky. Diffuse source has different values of total incoming radiance at a point  $E(a)$ . A no white diffuse source can have effect to the cast shadow. If we consider an outdoor environment, beside a reduction in the intensity, an outdoor cast shadow will result in a change of chrominance. Considering again the outdoor scene, the illumination of the sky has higher power components in the lower wavelengths  $\lambda$  (450-495nm) of the visible spectrum, and it is therefore assumed bluish as argued in [17]. When the direct illumination of the sun is blocked and a region is only illuminated by the diffuse ambient light of the sky, materials appears to be more bluish. This "bluish effect" and the chrominance distortion can be exploited for shadow detection and grouping of potential shadow pixel. Colour balance, or in the specific case, white balance can apply an adjustment of the intensities of colours. Even with this compensation, the effect of coloured diffuse sources remain on the objects, and it is possible to take an advantage on it, when possible. Because, in order to detect the shadow, the variation between a background image and current image is estimated, coloured effect can be detected. Bluish however is just one large but particular condition. In fact it is possible that, due to the particular scenario, the diffuse light has a different colouration. For instance, lights in coloured environment and also coloured lights which cause a chrominance distortion. If it exists, then it is possible to consider a general chrominance distortion in one of the three channels. Thus, objects which suffer more environment color will have an intensity variation bigger on the component which does not represent the color. For example, in outdoor scenario, the shadow will suffer more the effect of the sky and the intensity changes will be bigger in red and green channels than in blue chan-

nel.

## 5. An Innovative Information Fusion Approach for Supporting Shadow Detection in Complex Environments

As described in Section III we initially segment each frame into background, foreground, and shadow. This is performed by combining the results from background subtraction process and shadow detection. The methods are described below. The advantage of these methods is that they do not require a training phase. Nevertheless they give generally good performances as it will be shown in the experimental Section.

### 5.1. Background Subtraction

In this work we use a subtraction stereo method described in [11]. This method consist in a threshold subtraction .

For each pixel  $p$ , a background model is learned, from which the foreground probability can be estimated. Potential moving objects can be extracted by simply thresholding this density distribution, and within the segmentation the cast shadows can be evaluate over both the color and stereo domain.

Another drawback of using background subtraction is that for long sequence (day s, week s, months), it can be difficult to maintain the background model due to high varying illumination, precipitation, season changes, etc.

### 5.2. Color Shadow Detection

The method we are going to describe is a modification of the works described in [16] and [10]. It offers interesting performance in shadow detection without any training phase.

Considering that the radiance influences the cast shadow not linearly than the intensity ratios between neighboring shadow pixels depends on the source direction. The variations in background and segmented image will be used to calculate the error score within a small region, used for discriminating a pixel as shadow. The error score is computed as reported in eq.(4).

$$\Psi(x, y) = \sum_{c \in R, G, B} \sum_{i, j \in \omega(x, y)} |d_c(i, j) - d'_c(i, j)| \quad (4)$$

Even if colors cannot be used singularly to extract shadows, they represent an important source of information. The color difference which is used to estimate if a pixel belongs

to a cone shadow is calculated comparing the color information between the background and detected foreground image. The color space is modeled as follows

$$\begin{cases} C_1(x, y) = \arctan\left(\frac{I_r(x, y)}{I_b(x, y)}\right) \\ C_2(x, y) = \arctan\left(\frac{I_g(x, y)}{I_b(x, y)}\right) \end{cases} \quad (5)$$

The color score error due to the variation of color is computed as

$$\Lambda(x, y) = |C_1(x, y) - C'_1(x, y) + C_2(x, y) - C'_2(x, y)| \quad (6)$$

### 5.3. Stereo Shadow Detection

As previously mentioned, in ideal condition, it would be possible to separate an object from the shadow by the estimated distance. Once obtained the distance value of each point of the image from the background, a pixel is considered object if the difference is higher than zero (ideal condition), or  $\epsilon$  (real condition).

Because the calculation of distance generally is problematic in saturation regions, only the points with an intensity less than a threshold are used.

Stereo cameras have a range within the error measures calculated which is lower than the measured distance. The working range changes from camera to camera and it depends on several factors, software (matching algorithms) and hardware (sensors, camera displacement and numbers, focal length).

The estimated errors in Fig. 2 shows a quadratic relation between distance and error. According to our tests, accurate values can be obtained within 5 meters.

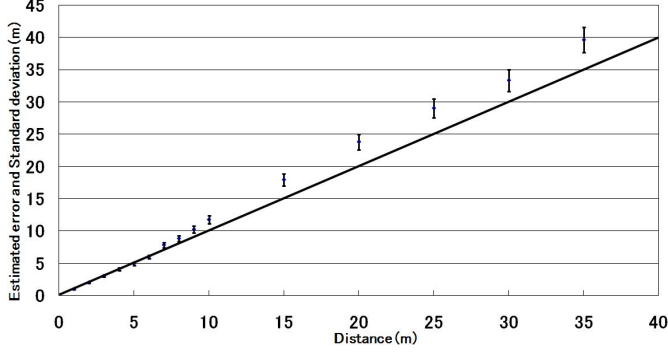
We took in consideration the Bumblebee2 stereo camera and estimated the errors due to the range measuring the real distance and estimated distance. The error is shown in Fig. 2 as vertical segments.

If the measure is taken after the best distance, the error shows a quadratic behaviour. Otherwise it has a quasi-linear behaviour.

If the distance information has been calculated both for background and foreground and it is within the best range, then if  $d_{x,y} \neq \emptyset \wedge d'_{x,y} \neq \emptyset \wedge d_{x,y} \leq \epsilon$

$$S_{x,y} = \begin{cases} 0 & \text{if } |d_{x,y} - d'_{x,y}| \leq m(d_{x,y} - \epsilon) + q \\ 1 & \text{otherwise.} \end{cases} \quad (7)$$

If the distance information has been calculated both for background and foreground but it is over the best range, then if  $d_{x,y} \neq \emptyset \wedge d'_{x,y} \neq \emptyset \wedge d_{x,y} > \epsilon$



**Figure 2. Error calculated for the stereo camera used. The error increases with the distance.**

$$S_{x,y} = \begin{cases} 0 & \text{if } |d_{x,y} - d'_{x,y}| \leq \frac{1}{m} (d_{x,y} - \epsilon)^2 + q \\ 1 & \text{otherwise.} \end{cases} \quad (8)$$

From the hypothesis that the camera is able to estimate a correct distance within a certain range  $r$ , we propose a method to automatically adjust  $m$  and  $q$ . Given the distance difference between background and current images, it is reasonable to suppose that within a certain confidence range  $r'$ , the pixels have to be labeled as shadow. Instead out of  $r'$ , pixels are expected to belong to a moving object. A set of points measured around  $r$  are collected and labeled based on the previous consideration. If a pixel is recognized as *shadow* and the distance difference is lower than  $r'$  then the point is marked as *success* otherwise as *failure*. If a pixel is recognized as *no shadow* and the distance difference is equal or greater than  $r'$ , then the point is marked as *success* otherwise as *failure*. An example can be seen in fig.3.

Once obtained the graph, we modify the parameters only if the percentage of success inside and outside the range  $r'$  is lower then an accuracy value.

At each iteration, the parameters are adjusted using a random function as follow:

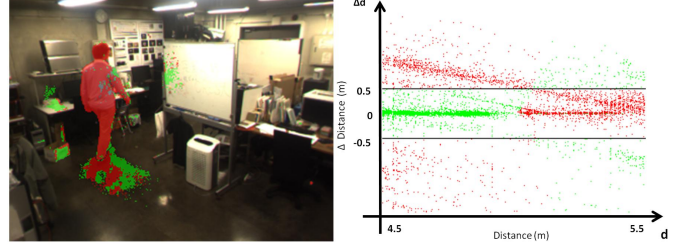
$$m = m + \text{sign} \cdot \text{rand} \quad (9)$$

$$q = q + \text{sign} \cdot \frac{\text{rand}}{100} \quad (10)$$

where *sign* is positive if the percentage of failures inside the range is higher or equal to the percentage of failures outside. Negative in opposite case.

#### 5.4. Dempster-Shafer Fusion

The Basic Belief Assignment can be viewed as a generalization of a probability density function. More precisely,



**Figure 3. Example of success and failure in stereo shadow detection. In the left picture, the pixels are depicted in red or green to put in evidence if it is shadow (green) or foreground (red). On the right, the graph shows the estimation success. Each colored pixel represent a case of success or failure. In red, a pixel is labeled as estimated. In green, a failure.**

a Basic Belief Assignment  $m(\cdot)$  is a function that assigns a value in  $[0, 1]$  to every subset  $\mathcal{A}$  of  $\theta$  that satisfies the following:

$$\sum_{\mathcal{A} \subseteq \theta} m(\mathcal{A}) = 1, \quad m(\emptyset) = 0$$

It is worth noting that  $m(\mathcal{A})$  is the belief that supports the subset  $\mathcal{A}$  of  $\theta$ , not the elements of  $\mathcal{A}$ . This reflects some ignorance because this means that we can assign belief only to subsets of  $\theta$ , not to the individual hypothesis as in classical probability theory.

Consider two Basic Belief Assignments  $m_1(\cdot)$  and  $m_2(\cdot)$  and the corresponding belief functions  $bel_1(\cdot)$  and  $bel_2(\cdot)$ . Let  $\mathcal{A}_j$  and  $\mathcal{B}_k$  be subsets of  $\theta$ . Then  $m_1(\cdot)$  and  $m_2(\cdot)$  can be combined to obtain the belief mass assigned to  $\mathcal{C} \subset \theta$  according to the following formula [23]:

$$m(\mathcal{C}) = m_1 \oplus m_2 = \frac{\sum_{j,k, \mathcal{A}_j \cap \mathcal{B}_k = \mathcal{C}} m_1(\mathcal{A}_j) m_2(\mathcal{B}_k)}{1 - \sum_{j,k, \mathcal{A}_j \cap \mathcal{B}_k = \emptyset} m_1(\mathcal{A}_j) m_2(\mathcal{B}_k)} \quad (11)$$

The denominator is a normalizing factor, which measures how much  $m_1(\cdot)$  and  $m_2(\cdot)$  are conflicting.

#### 5.5. Basic Belief Assignment for Shadow Estimation

The quantities above computed, i.e.  $\Psi(x, y)$ ,  $\Lambda(x, y)$  and  $S(x, y)$  can be considered as the outputs of three experts that, from different knowledge, represent the possibility that the pixel  $(x, y)$  is a shadow. More precisely, the more each of these parameters is close to zero, the more likely the relative pixel represents a shadow. From such quantities, three other quantities are computed as follows:  $\Psi'(x, y) = \max((1 - \Psi(x, y)), 0)$ ,  $\Lambda(x, y) = \max((1 -$

$\Lambda(x, y), 0), S'(x, y) = (1 - S(x, y))$ . These three quantities are then normalized such that their sum is equal to one. The next step we performed is to divide, with empirical thresholds, each quantity in three sections, corresponding to Umbra, Penumbra and Luminance respectively. Since the  $S'(x, y)$  parameter given that the  $S'(x, y)$  parameter distinguishes only the shadow by the light, the probability that the pixel represents umbra or penumbra is the same. Thus, we considered the following set of possible hypotheses,  $\theta = \{U, P, L\}$  that is Umbra, Penumbra, Light. This give rise to the following power set:  $\{U\}, \{P\}, \{L\}, \{UP\}, \{UL\}, \{PL\}, \{UPL\}$ . Each subset is assigned a belief  $m$  on the basis of the three knowledge sources. The final shadow index  $\Theta$  is obtained by combination:

$$\Theta = \bigoplus_{i=1}^n m_i = ((m_1 \oplus m_2) \oplus m_3)$$

## 6. Confidence Estimation

In order to estimate the diffuse chrominance of the scene, we propose a method which analyzes the variation in a sequence of images. If  $I_c$  is the intensity of the current frame and  $I'_c$  is the intensity of a given background image, the diffuse chrominance is considered the highest variation in a sequence.  $c$  can assume red, green, or blue value, and *gray* is equal to gray value. We consider that each channel has 8 bit resolution. Even if it is obvious that colored objects influence the estimation of the color, considering all the points of the image in a sequence will reduce that effect.

Given the background image and the current image, the histogram of the differences is computed as

$$\forall_p, H_{c, |I_p - I'_p|} = H_{c, |I_p - I'_p|} + 1 \quad (12)$$

The color value is then

$$cv_c = \frac{\sum_{i=0}^{256} H_{c,i} \cdot i}{s} \quad (13)$$

The difference in intensity between the gray scale image in each channel is computed as reported in eq.(14).

$$DI_c = \frac{1 - |cv_c - cv_{gray}|}{256} \quad (14)$$

and the proportional variation of each channel respect the gray scale image is computed as

$$PVI_c = \begin{cases} - \left( 1 - \min \left( \frac{cv_c}{cv_{gray}}, \frac{cv_{gray}}{cv_c} \right) \right) & \text{if } cv_c > cv_{gray} \\ 1 - \min \left( \frac{cv_c}{cv_{gray}}, \frac{cv_{gray}}{cv_c} \right) & \text{otherwise.} \end{cases} \quad (15)$$

Differences and proportional variations are gathered for the number of frames necessary to estimate the average and variance of the sequence analyzed. Empirically we estimated that 100 frames are sufficient for that analysis.

First the  $DI$  are normalize respect the maximum and minimum value of all the color  $DI$ .

$$DI_c = \frac{DI_c - \min(DI)}{\max(DI) - \min(DI)} \quad (16)$$

We compute the average and variance of the differences and proportional variations. The sequence average difference and variation are estimated as shown in eq.(17) and eq.(18).

$$SAD_c = \ln \left( \frac{\mu DI_c}{1 - \delta^2 DI_c} \right) \quad (17)$$

$$SAPV_c = \left| \ln \left( \frac{\mu PVI_c}{1 - \delta^2 PVI_c} \right) \right| \quad (18)$$

For each combination of colors the difference of sequence average difference and variation is computed in order to estimate the prevailing color.

$$\begin{aligned} \Delta_{rg} &= SAD_r - SAD_g \\ \Delta_{rb} &= SAD_r - SAD_b \\ \Delta_{gb} &= SAD_g - SAD_b \end{aligned} \quad (19)$$

and finally the color strenght is estimated as described in eq.(20).

$$\begin{aligned} CS_r &= (\Delta_{rg} \geq 0) \cdot |\Delta_{rg}| + (\Delta_{rb} \geq 0) \cdot |\Delta_{rb}| \\ CS_g &= (\Delta_{rg} < 0) \cdot |\Delta_{rg}| + (\Delta_{gb} \geq 0) \cdot |\Delta_{gb}| \\ CS_b &= (\Delta_{rb} < 0) \cdot |\Delta_{rb}| + (\Delta_{gb} < 0) \cdot |\Delta_{gb}| \end{aligned} \quad (20)$$

The suggested best chrominance will be the highest value.

Obviously, one color will prevails, unless the image is not completely gray scale. Because we want to avoid errors due to noisy or particular configurations, we consider colored diffuse light only the best color value which satisfies the following equation.

$$color = \begin{cases} white & \text{if } CS_{best} < \min(SAPV) \\ best & \text{otherwise.} \end{cases} \quad (21)$$

We consider that a pixel segmented as foreground cannot be a shadowed pixel if its intensity is higher than background. Thus, a pixel is candidate as shadow if

$$sp_a = (I_a^R < \mu^R) \wedge (I_a^G < \mu^G) \wedge (I_a^B < \mu^G) \quad (22)$$

Moreover, in the case of bluish effect (similar to greenish and reddish), the changes on the intensity component of the red and blue channels are bigger than the blue channel. To

be more flexible, the increment will be proportional. This fact can be used to reduce the shadow region as follows:

If bluish

$$cs_a = (k(I_a^R - \mu^R) > (I_a^B - \mu^B) \wedge k(I_a^G - \mu^G) > (I_a^B - \mu^B)) \wedge sp_a \quad (23)$$

If reddish

$$cs_a = (k(I_a^G - \mu^G) > (I_a^R - \mu^R) \wedge k(I_a^B - \mu^B) > (I_a^R - \mu^R)) \wedge sp_a \quad (24)$$

If greenish

$$cs_a = (k(I_a^R - \mu^R) > (I_a^G - \mu^G) \wedge k(I_a^B - \mu^B) > (I_a^G - \mu^G)) \wedge sp_a \quad (25)$$

The mask so obtained is then used to mark the pixels that should be analyzed.

If the color of the environment is not white the pixels which are not candidate to be shadow are labeled as moving objects.

### 6.1. Merge Methods

Once the shadow values are obtained with the previously described methods, the difficult task is to find a relationship between result obtained from color information and shadow information. Because the measures are not directly relationable, we estimate the strength of each detection method.

The shadow detection method described in section V.A return a value we called shadow parameter. It is possible to combine the shadow parameter with the distance in order to obtain a *confidence* value. Likewise, it is possible to obtain a confidence value from the stereo shadow detection. The confidence values are higher for the stereo information if the distance of a point from the camera, is near the focal point. On the opposite side, confidence value is higher for the points which have a distance lower or higher than the focal point.

The curve that define the probability to be shadow is not centered around zero, but differs if a pixel is detected as shadow or no shadow. The equations can be so resumed. If the distance from the camera is less or equal to the focal point:

$$W_{c_{x,y}} = \begin{cases} pShadow \cdot \left( \frac{d}{m} + q \right) & \text{if shadow} \\ \frac{Fp}{pShadow} \cdot \left( \frac{d}{m} + q \right) & \text{otherwise.} \end{cases} \quad (26)$$

$$W_{s_{x,y}} = \begin{cases} \frac{1}{\Delta_d} \cdot \frac{1}{\frac{d}{m} + q} & \text{if shadow} \\ 2\Delta_d \cdot \frac{1}{\frac{d}{m} + q} & \text{otherwise.} \end{cases} \quad (27)$$

However, if the distance from the camera increases:

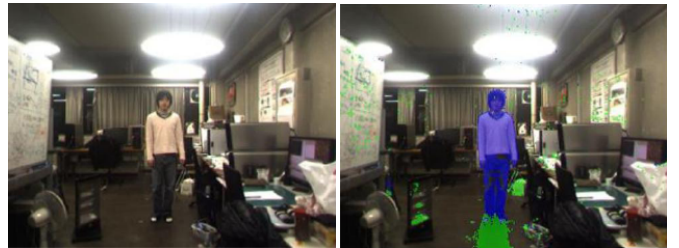
$$W_{c_{x,y}} = \begin{cases} pShadow \cdot \left( \frac{1}{m} \cdot (d - Fp)^2 + q \right) & \text{if shadow} \\ \frac{Fp}{pShadow} \cdot \left( \frac{1}{m} \cdot (d - Fp)^2 + q \right) & \text{otherwise.} \end{cases} \quad (28)$$

$$W_{s_{x,y}} = \begin{cases} \frac{1}{\Delta_d} \cdot \frac{1}{\frac{1}{m} \cdot (d - Fp)^2 + q} & \text{if shadow} \\ 2\Delta_d \cdot \frac{1}{\frac{1}{m} \cdot (d - Fp)^2 + q} & \text{otherwise.} \end{cases} \quad (29)$$

Stereo information may be not always available in all the points of the image. This is due to several factors: light conditions, distance of the object from the camera, visibility of an object from both the cameras. If the stereo information is not available, we choose to use only the color information. In the case a pixel is labeled as shadow (or no shadow) with both the methods, the pixel will be labeled with the detected value. If the detection value is different, the pixel will be labeled with the value of the method which have higher weight.

## 7. Qualitative Evaluation

The proposed algorithm has been implemented using a Bumblebee2 stereo camera from Point Grey Research on an Intel Quad CPU at 2.83 GHz and 4GB Ram. Unfortunately we did not find a 3D data set suitable for us because we need results obtained with that stereo camera. Thus we acquired several video sequences in different conditions of illumination and camera orientation. Both in-door and out-door environments have been considered. The results presented hereafter are obtained from our captured sequences. The results presented here are only qualitative; in fact we did not performed actual comparisons with ground truths yet but we want only to evaluate the results from a qualitative point of view, for now. The results appear to us quite good, as shown in Fig.4) and Fig.5).



**Figure 4.** In this figure an example of shadow detection performed with the proposed algorithm for an in-door environment is reported. Shadow pixels are represented in green.





**Figure 5.** In this figure an example of shadow detection performed with the proposed algorithm for an out-door environment is reported. Shadow pixels are represented in green.

## 8. Conclusions and Future Work

In this paper, a new moving cast shadow detection algorithm that requires a stereo camera is proposed. The algorithm exploits the color, texture, temporal and depth information. Although many shadow detections have been performed, in this paper only the theory and some examples are reported. Accurate ground-truth based performances and comparisons with state of the art algorithms will be presented in future papers. Also a GPU implementation of the algorithm will be explored. Future work is mainly focused in extending our proposed framework by means of several characteristics, as to enhance it significantly. For instance, some interesting properties to be investigated in future are: *fragmentation* (e.g., [1, 3]), *approximation* (e.g., [4, 5]), *privacy preservation* (e.g., [6, 7]), *big data* (e.g., [8, 2]).

## References

- [1] A. Bonifati and A. Cuzzocrea. Efficient fragmentation of large XML documents. In *DEXA 2007, Regensburg, Germany, September 3-7, 2007*, pages 539–550, 2007.
- [2] A. Cuzzocrea, L. Bellatreche, and I. Song. Data warehousing and OLAP over big data: current challenges and future research directions. In *DOLAP 2013, San Francisco, CA, USA, October 28, 2013*, pages 67–70, 2013.
- [3] A. Cuzzocrea, J. Darmont, and H. Mahboubi. Fragmenting very large XML data warehouses via k-means clustering algorithm. *IJBIDM*, 4(3/4):301–328, 2009.
- [4] A. Cuzzocrea, F. Furfaro, S. Greco, E. Masciari, G. M. Mazzeo, and D. Saccà. A distributed system for answering range queries on sensor network data. In *3rd IEEE (PerCom 2005 Workshops), 8-12 March 2005, Kauai Island, HI, USA*, pages 369–373, 2005.
- [5] A. Cuzzocrea and U. Matrangolo. Analytical synopses for approximate query answering in OLAP environments. In *DEXA 2004 Zaragoza, Spain, August 30-September 3, 2004*, pages 359–370, 2004.
- [6] A. Cuzzocrea, V. Russo, and D. Saccà. A robust sampling-based framework for privacy preserving OLAP. In *DaWaK 2008, Turin, Italy, September 2-5, 2008*, pages 97–114, 2008.
- [7] A. Cuzzocrea and D. Saccà. Balancing accuracy and privacy of OLAP aggregations on data cubes. In *DOLAP 2010, Toronto, Ontario, Canada, October 30, 2010*, pages 93–98, 2010.
- [8] A. Cuzzocrea, D. Saccà, and J. D. Ullman. Big data: a research agenda. In *IDEAS '13, Barcelona, Spain - October 09 - 11, 2013*, pages 198–203, 2013.
- [9] A. C. E. Salvador and T. Ebrahimi. Cast shadow segmentation using invariant color features. *CVIU*, 95(2):238–259, 2004.
- [10] A. M. et al. Auto-adaptive threshold and shadow detection approaches for pedestrian detection. *AWSVCI*, pages 9–12, 2009.
- [11] K. U. et al. Subtraction stereo. a stereo camera system that focuses on moving regions. *Proc. Of SPIE-IS T Electronic Imaging*, 7239 Three Dimensional Imaging Metrology, 2009.
- [12] C. L. G. Finlayson, S. Hordley and M. Drew. On the removal of shadows from images. *IEEE TPAMI*, 28(1):59–68, 2006.
- [13] T. M. I. Huerta, M. Holte and J. González. Detection and removal of chromatic moving shadows in surveillance scenarios. *ICCV*, pages 1499–1506, 2009.
- [14] D. H. K. Kim, T. Chalidabhongse and L. Davis. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 11(3):172–185, 2005.
- [15] H. P. M.K. Moghimi. Shadow detection based on combinations of hsv color space and orthogonal transformation in surveillance videos. In *Proceedings of Iranian Conference on Intelligent Systems*.
- [16] C. C. M.T. Yang, K.H. Lo and W. Tai. Moving cast shadow detection by exploiting multiple cues. *Image Processing*, 2:95–104, 2008.
- [17] S. Nadimi and B. Bhanu. Physical models for moving shadow and object detection in video. *IEEE TPAMI*, 26(8):1079–1087, 2007.
- [18] K. Onoguchi. Shadow elimination method for moving object detection. *ICPR*, 1:583–587, 1998.
- [19] N. A. Q. Yang, K.H. Tan. Shadow removal using bilateral filtering. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 2012.
- [20] M. P. R. Cucchiara, C. Grana and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE TPAMI*, 25(10):1337–1342, 2003.
- [21] M. P. A. P. S. S. R. Cucchiara, C. Grana. Improving shadow suppression in moving object detection with hsv color information. *C. A.F. editor, Proc. IEEE ITSC*, pages 334–339, 2001.
- [22] R. K. S. Surkutlawar. Shadow suppression using rgb and hsv color space in moving object detection. *International Journal of Advanced Computer Science and Applications*, 2013.
- [23] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.
- [24] Y. Weiss. Deriving intrinsic images from image. *ICCV*, 2:68–75, 2001.
- [25] G. J. X. Dong, K. Wang. Moving object and shadow detection based on rgb color space and edge ratio. In *Proceedings of CISIP'09*.
- [26] A. B. Y. Ivanov and J. Liu. Fast lighting independent background subtraction. *IJCV*, 37(2):199–207, 2000.

- [27] K. I. Y. Matsushita, K. Nishino and M. Sakauchi. Illumination normalization with time-dependent intrinsic images for video surveillance. *IEEE TPAMI*, 26(10):1336–1347, 2004.

# A Complete Classification of Occlusion Observer's Point of View for 3D Qualitative Spatial Reasoning

Chaman Lal Sabharwal

Missouri University of Science and Technology

Rolla, MO- 63128, USA

[chaman@mst.edu](mailto:chaman@mst.edu)

## Abstract

*Qualitative Spatial Reasoning (QSR) theories have applications in areas such as geographic information systems (GIS), robotics, biomedicine and spatial databases. Several region connection calculi have been proposed for use in this capacity. Primarily the existing QSR theories have been applied to 2D data. Yet the ability to perform qualitative reasoning over a collection of 3D spatial objects is desirable. Over the past two decades several theories have appeared for accurately representing and acquiring 3D spatial knowledge: LOS-14, its extension ROC-20, occlusion calculus OCC, fourteen occlusion states OCS-14, and a recent visual VRCC-3D+ with 17 occlusion predicates. Each has a positive impact. However there are still some issues and ambiguities that require unambiguous ontology and resolution. In this paper, we provide a new set of self-documenting predicates for 3D complete spatial relations including occlusion. In addition we provide new heuristics for eliminating the time consuming computations by employing efficient data structures. This improvement will greatly enhance the usefulness and usability of aforementioned systems.*

Keywords: geographical information systems, robotic navigation, spatial objects, graphics, occlusion

## 1. Introduction

Most of the theories about space and time study the quantitative aspects of a problem, whereas the qualitative calculi allow for rather inexpensive reasoning about entities located in space. For example, some of spatial reasoning is implemented for handling geographical information systems (GIS) queries efficiently [1], and such reasoning is used for robotic sensors, biomedicine, spatial networking, and cognitive sciences.

Although reasoning over two dimensions is sufficient for many applications [2], other spatial reasoning applications need to consider information in more than two dimensions. Thus the ability to perform qualitative spatial reasoning over a collection of 3D objects is necessary. The 3D spatial reasoning involves the visualizing and then manipulating spatial relations. A 2D QSR system cannot be utilized for such tasks. Robots see and interpret the world with data acquired through sensors.

Further, in order to determine occlusion, the view reference point, the plane of projection, and the type of

projection must be known. Over the past two decades several theories have appeared for accurately representing and acquiring 3D spatial knowledge: Galton developed Line of Sight method with 14 occlusion relations LOS-14, in 1994 [3], its extension Region Occlusion Calculus with 20 occlusion relations ROC-20 was designed by Randell et al. in 2001 [4], Kohler developed the occlusion calculus OCC in 2002 [5], RCCD-3D by Albath et al. in 2010 [6], Guha et al. Designed OCS-14 (*Occlusion States 14*) in 2011 [7] and at the same time independently Sabharwal et al. developed VRCC-3D+ in 2011 [8]. Each has a positive impact. However there were some issues that required resolution. Recently, Elloe et al. attempted to resolve such issues and improve upon the computational aspects in 2014 [9]. In their attempt, they introduced two predicates for depth as discussed in Section 4, which turned out to be obscure and inefficient. In this paper, we provide a completely new set of self-documenting predicates for occlusion relations superseding the aforementioned work. In addition we provide efficient heuristics for eliminating the time consuming ray tracing used in performing obscuration computations. These improvements will greatly enhance the usefulness and usability of the aforementioned systems. The same reasoning works for mobile objects when the observer is stationary.

This paper is organized as: Section 2 describes mathematical concepts and region connection calculus background Section 3 clarifies the graphics concept of closer, occlude, obscure, in front, Section 4 describes recent implementation issues in handling occlusions problems, Section 5 describes complete, consistent and new set of occlusion relations using first order logic, Section 6 discusses efficient implementation techniques, Section 7 is on conclusions and future work.

## 2. Background

### A. Mathematical Concepts

Basic mathematical concepts are the same as in the point set topology. The definition of connectedness in region connection calculus is slightly different. 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. In mathematics, a set is *connected* if it cannot be the union of disjoint open sets. For example, the set  $(0,1) \cup (1,2)$  is disconnected as  $(0,1)$  and  $(1,2)$  are open sets. In RCC, regions  $A$  and  $B$  are *weakly* connected if  $\bar{A} \cap \bar{B} \neq \emptyset$ . Thus  $(0,1) \cup (1,2)$  is connected in RCC because

$[0,1] \cap [1,2] \neq \emptyset$ . This is equivalent to (1)  $C(A,A)$ , and (2)  $C(A,B) \iff C(B,A)$  for any two regions A and B.

### 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 [10, 11]. Conceptually, for any two distinct objects, there are three possibilities on broad level: (a) *one is outside the other*, that results in the discrete spatial relation, DR(DiscRete), (b) *one overlaps the other across boundaries*. This means PO(proper overlap), (c) *one is inside the other*, that means EQ(equal) or PP(proper part). To make relations jointly exhaustive and pairwise distinct (JEPD), we have converse relation denoted PPc(proper part converse),  $PPc(A,B) \iff PP(B,A)$ . These five relations constitute RCC5 relations. For additional detail on discrete, specifically, DR is split into DC(disconnected) and EC(externally connected). For proper part, PP is split into TPP (tangential proper part) and NTPP (non-tangential Proper part). Similarly for proper part converse, PPc, we have converse relations TPPc and NTPPc. These eight relations constitute RCC8 relations of region connection calculus.

RCC8 was formalized by using first order logic [10] or 9-Intersection model [11], see Fig. 1. The intersections of interior (Int), boundary (Bnd), and exterior (Ext) of one object with the other object form 9-Intersections: IntInt, IntBnd, IntExt, BndInt, BndBnd, BndExt, ExtInt, ExtBnd, ExtExt.

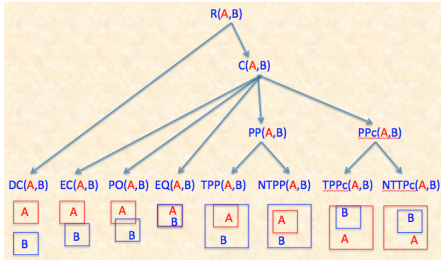


Fig. 1. RCC8 Relations in 2D.

Whereas a 2D object is in a plane, a 3D object is in space. The simplest examples of 3D objects are a pyramid, a cuboid, 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 the rest of this discussion, we will base our analysis on convex objects.

RCC8 spatial relation for 3D objects is incomplete without occlusion consideration. VRCC-3D+ is a region connection calculus that qualitatively determines the spatial relations between 3D objects, both in terms of connectivity and obscuration [6, 8, 9]. 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. The relative depth is denoted by tri-valent parameter InFront. Fifteen obscuration relations were defined in VRCC-3D+ [8,9]. 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 hybrid occlusion relation specifies both a connectivity relation and an obscuration relation.

## 3. Occlusion Concepts

Basically occlusion of one object by another object depends on the observer location relative to the objects. Our approach is to derive spatial obscuration relations and classification from projection of 3D objects on a 2D projection plane and relative distance of the objects from the observer. The 2D plane is used to determine the existence of occlusion. The depth parameter, InFront, coupled with projections determine the type of occlusion. Obscuration predicates are based on two parameters: projection in a plane and depth (distance of the object from viewpoint).

In general, there are two types of projections: Parallel and Perspective. Parallel projection loses the depth concept in the projection, because the all projectors are parallel. We use perspective view for accurate depth representation. The terms "inFront", "occlusion", "obscuration", "closer" are closely related. In natural language, the term inFront between two objects A and B is synonymously interpreted as "A is in front of B", "A occludes B", "A obscures B", "A is closer than B". However, there are subtle differences between these, see examples below. Here we describe these terms precisely with the help of examples.

### A. Examples of occlusion detection supporting qualitative spatial reasoning

In computer graphics, there is a distinction between terms occlusion and obscuration. Occlusion means opaque (hidden, obstructed), whereas obscuration means unclear (hazy, vaguely transparent). For this reason, the graphics community prefers the term occlusion to obscuration. We are using both terms interchangeably to mean opaque.

#### A1. Point Occlusion

If V (viewer), a, b are collinear points in that order and the distance of a is smaller than the distance of b from viewer V, then the three terms are equivalent: point a is closer than point b or a occludes b or a is in front of b, see Fig. 2(a). However if both points a and b are equidistant from V, then none is in front of the other, see Fig. 2(b). Two points are equidistant, coincident, none occludes the other. This is *not* mutual occlusion as explained in Section 3.A2. If V, a, b are not collinear, then b can be closer than a without b occluding a, see Fig. 2(c). Thus a point can be closer without being in front of the other.

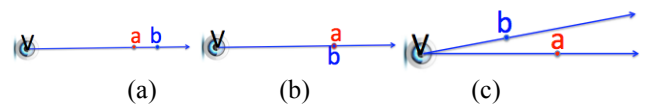


Fig. 2 (a) View point V, points a and b; a occludes b, a is in front of b, a is closer than b to the viewpoint. (b) View point

V, points a and b; a and b are equidistant, none is in front of the other. (c) View point, points a and b, b is closer than a, but b does not occlude a, b is not in front of a.

## A2. Object Occlusion

In 2D, to determine whether an area A occludes an area B, there are several configurations of A and B. If a ray from V does not intersect any of the two regions, then it does not shed light on the occlusion. If for *every* ray from V, it intersects exclusively one object and not the other object, then no object occludes the other, see Fig. 3(a, b). An object A can be closer than object B without obscuring the object B, see Fig. 3(a).

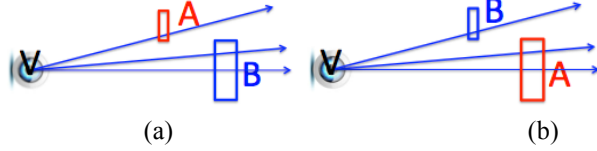


Fig. 3 Objects do not intersect (a) A is closer than B (b) B is closer than A.

We cannot make any conclusion from a ray if the ray intersects only one object exclusively and not the other object, see Fig. 4(b).

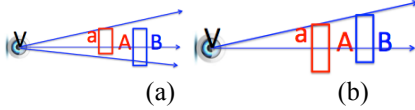


Fig. 4. (a) ray from V intersects A, but not B, another ray from V intersects B, but not A, still A occludes B (*partially*). (b) ray from V intersects both, the front point occludes the other point, still A occludes B (*completely*).

If for *every* ray from V that intersects the interiors of both objects A and B at points a and b, with a in front of b, then the object A occludes object B, or A is in front of B are synonymous, Fig. 5(a, b). Also object A is closer to V than object B is to V, Fig. 5(a), A *partially* obscures B.

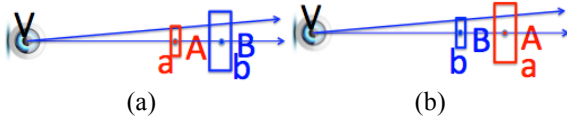


Fig. 5. (a) A is in front of B, A occludes B, A is closer than B is, (b) B is in front of A; B occludes A; B is closer than A is.

If there are two rays that intersect objects A and B, such that for one ray, the intersection with A is closer than intersection with B, and for the second ray, the intersection with B is closer than intersection with A, then the object A partly occludes B and conversely. They *mutually* occlude each other, see Fig. 6. None alone occludes the other completely or partially.

We require that for *every* ray from V that intersects both the objects, the two intersections are used to make judgment about the relative depth of the objects. If the ray intersection occurs at multiple points, only the closest intersections are

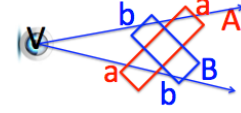


Fig. 6. Mutual Occlusion.

significant for occlusion detection. For example, if the ray intersection is Vbaab, we need to process only first two intersections, in this case, Vba, so the point b occludes point a see Fig. 7; If the ray intersection is Vabab, we process the first two intersections, in this case, Vab. So the point a occludes the point b.

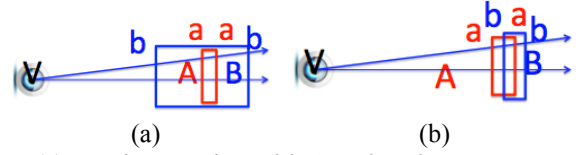


Fig. 7. (a) Ray intersection with A and B shown at two points each so that ray is Vbaab. However only Vba is sufficient for occlusion determination between A and B. (b) Ray intersection with A and B shown at two points each so that ray is Vabab. However only Vab is sufficient for occlusion determination between A and B.

*Partial or full occlusion:* If for *every* ray from view point V that intersects both A and B, and ray points  $a \in A$  are closer than corresponding points  $b \in B$ , then A occludes (*partially* or *completely*) B, see Fig 5(a) and Fig. 7(b).

*Mutual or Equal.* Similarly it can be seen that if on *every* ray from view point V that intersects both A and B, and if sometimes a is closer and sometimes b is closer, then A and B *mutually* occlude each other else if points a, b are equidistant from V, then it is termed as *equal* occlusion for spatial reasoning, this is a clear distinction between *mutual* and *equal* occlusion predicates. This is a conscious decision made for spatial reasoning because it provides expressive power at no cost.

## 4. Problems

In Section 1, we mentioned that there are issues in the implementation of occlusion in VRCC-3D+ [9]. Let us first describe the problems and then we will present new formulation and new crisp occlusion relations in Section 5.

**First Problem.** Let A, B be 3D objects,  $(x, y)$  be a projection point in the projection plane, a line of sight from camera C (Center of Perspective Projection) through  $(x, y)$  intersects objects at points  $f_A(x, y)$  and  $f_B(x, y)$ . Elloe et al. [9] define occlusion by means of two predicates  $o$  and  $o_c$ . The definition of predicate  $o$  is the following equation [9]

$$o(A, B) = \begin{cases} T: \exists x, y \in C - f_A(x, y) < |C - f_B(x, y)| \\ F: \text{otherwise} \end{cases}$$

and converse  $o_c(A, B) = o(B, A)$ . This definition of occlusion is trivially inaccurate. By comparing points



on only one ray, one cannot claim any thing about the whole object. For example, the occlusion detection is inconclusive by this definition as in Fig. 6, there are points on A and B, in one case point a is in front of point b and in the other case b is in front of a.

**Second Problem** Each occlusion relation is determined with a 5 parameters (IntInt, IntBnd, BndInt, o, o<sub>c</sub>) instead of 4 parameters (IntInt, IntBnd, BndInt, InFront). This increases computational time complexity by 20%. Increased computation slows down the interaction time and is detrimental to usability of the system. Instead of two new predicates to replace a single predicate InFront, the tri-valent predicate InFront can be made quad-valent to accommodate *mutual* for accurate interpretation of InFront.

**Third Problem** Elloe et al. [9] point out that with their approach, 15 occlusion relations have been reduced to 12 relations. Their table 1 below is listing of 12 relations. A closer look at the table indicates that some of their relations are disjunctions. In a table each row represents a unique relation rule, the number of rows in the table is 12+1+1+2+1=17. This contradicts the assertion that it is a smaller set of occlusions than 15 unique relations. For efficiency consideration, larger set of predicates amounts to larger computation time. Baring, the disjunctions, we will show that same tasks can be accomplished with 7 crisp relations instead of 12.

**Fourth Problem** There is no consistent ontology for naming the occlusion relations, see table 1. We devise a complete, systematic comprehensible list of occlusion relations see table 2. Also there are repeated occlusion ray tracing computations that can be eliminated. As it is, the occlusion computation is quite inefficient in [9].

Table 1. Full obscuration relation set with identified converse relations. cited[9]

	IntInt	IntExt	ExtInt	o	o <sub>c</sub>	Converse
nObs_e	F	T	T	F	F	nObs_e
pObs	T	T	T	T	F	pObs_c
pObs_c	T	T	T	F	T	pObs
pObs_e	T	T	T	F	F	pObs_c
pObs_m	T	T	T	T	T	pObs_m
eObs	T	F	F	T	F	eObs_c
eObs_c	T	F	F	F	T	eObs_c
eObs_e	T	F	F	F	F	eObs_e
eObs_m	T	F	F	T	T	eObs_m
cObs	T	T	F	T	F	cObs_c
cObs_c	T	T	F	F	T	cObs
cObs_e	T	T	F	F	F	cObs_e

We will present improved and enhanced set of crisp self-documenting predicates. Consequently these issues will disappear de facto in Section 5.

## 5. Completeness of Spatial Object Occlusions

In this section, we describe occlusion predicates gradually as follows: (1) redefine InFront accurately, (2) describe occlusion relation in natural language, (3) define occlusion predicates using first order logic, and (4) describe predicates via a table whose each row is rule for occlusion determination and classification.

In general, occlusion analysis is performed in two steps. First, the RCC5 relation is computed between the projections in 2D. Second, the qualitative spatial distance between the viewer and 3D objects is determined. The projection alone is not sufficient to determine obscuration. In Sabharwal et al. [8], the predicate InFront is used with value Y if A is closer than B; N is used if B is closer than A, and E is used imply that A and B are equidistant. As shown in the examples in Section 2, though three relations are accurate for single points, but they are not exhaustive for objects. There is a possibility that the objects cross, see Fig. 6. This leads to inaccurate classification due to tri-valent interpretation of the InFront predicate.

However, it is noted that the predicate InFront is sound in principle, but incomplete and inaccurate in implementation. We propose that predicate InFront accommodate *mutual* occlusion explicitly, when objects cross each other. In our further discussion, we will have InFront represent four possibilities; “A is in front of B”, “B is in front of A”, “A and B are equidistant from V”, and “A and B *mutually* obscure each other to indicate that A obscures B *partly* and B obscures A *partly* exclusively”. Using four values of InFront, we will update occlusion relations accordingly. The characterization of crisp occlusion relations is detailed in table 2.

### 5.1 Description of Depth Parameter InFront

Let V be the viewer, let A<sub>p</sub> and B<sub>p</sub> be the projections of A and B on the projection plane. The observer captures the scene that is in field of view, FOV. For occlusion purposes, the *viewer sees* objects through the *window* A<sub>p</sub> ∪ B<sub>p</sub> only in the projection plane. The ray from V, to intersect both objects, is through points in A<sub>p</sub> ∩ B<sub>p</sub>. Any reference to objects, A and B for InFront predicate, is reference to the part of objects seen via *only* A<sub>p</sub> ∩ B<sub>p</sub>. Let P(x,y) be a point in A<sub>p</sub> ∩ B<sub>p</sub>, let the ray VP intersect A and B at points f<sub>A</sub>(x,y) and f<sub>B</sub>(x,y) closest to the viewer. For qualitative distance, the value of InFront is formally stated as follows.



---

**Algorithm** for depth *InFront* determination

---

**If**  $A_p \cap B_p = \emptyset$   
    then there is no obscuration: *InFront* = "na"  
**elseif**  $\forall (x,y) \in A_p \cap B_p$ ,  
    **if**  $f_A(x,y) = f_B(x,y)$ , then A and B are equidistant:  
        *InFront* = "E"  
    **elseif**  $f_A(x,y) \leq f_B(x,y)$ , then A obscures B: *InFront* = "A"  
    **elseif**  $f_A(x,y) \geq f_B(x,y)$ , then B obscures A: *InFront* = "B"  
**elseif**  $\exists (x,y), (x',y') \in A_p \cap B_p$  such that  
     $f_A(x,y) < f_B(x',y')$ , and  $f_A(x',y') > f_B(x,y)$  then A and B  
    mutually obscure each other: *InFront* = "M"  
**end**

---

*Caution:* For the sake of simplicity, we may loosely write *InFront* = A, B, E, M instead of *InFront* = "A", "B", "E", "M".

Now the quad-valent distance parameter *InFront* is accurate description depth relation parameter. To make the occlusion relations self-documenting, we denote the occlusion predicate as  $xObs\_z(A,B)$ . Since some relations have converse while others do not, to make it completely symmetric, we have (1)  $z=a$  for "A is in front of B", (2) for converse, "B in front of A", we use  $z=b$ ; (3) for equality,  $z=e$  if "A and B are equidistant", and (4)  $z=m$  for *mutual* occlusion when "partly A is in front of B and partly B is in front of A". This way it is easier to comprehend the predicates when  $z$  is used to describe depth. Clearly there is distinction between *mutual* and *equal* occlusion for QSR. For natural language expressiveness in classification, we will use two distinct predicates, one for *mutual* and one for *equal*, instead of combining them into one as has been done in the past [8].

**5.2 Occlusion Predicates in Natural Language.** The system has to interpret data as viewed by the observer. Most difficult part is the representation of spatial occlusion predicate with complete expressive power. With  $xObs\_z(A,B)$ , in essence  $x$  refers to the type of occlusion,  $n$ ,  $p$ ,  $m$ ,  $e$ ,  $c$ , exclusively and  $z$  refers to qualitative distance of the objects from the viewer. There are four types of distances for objects and five types of obscuration. Out of 20 relations, some combinations are impossible, only the possible combinations are described here.

The discussion of obscuration predicate is incomplete without reference to projections per se. It is not possible to know the type of occlusion apriori. As we define the term  $xObs\_z(A,B)$  lucidly, we refine occlusion by integrating contribution of RCC5 relations, namely,  $DR(A_p, B_p)$ ,  $PO(A_p, B_p)$ ,  $EQ(A_p, B_p)$ ,  $PP(A_p, B_p)$ , and  $PPc(A_p, B_p)$ . As such, we upgrade the predicate  $xObs\_z$  to  $xObsy\_z$  where  $y$  refers to the RCC5 relation in projection plane,  $x$  refers to type of projection, and  $z$  refers to relative distance parameter *InFront*.

The complete listing of the predicates is given in Section 5.3 and for visual inspection a table 2 is given in Section 5.4. This approach eliminates errors and leads to efficient reasoning.

Now for  $x = n$ ,  $y$  in  $nObsy\_z(A,B)$  corresponds to the relation  $DR(A_p, B_p)$ . In this case  $nObs$  will be true independent of the value of  $z$ , so  $z$  is not applicable for this. Therefore four versions of  $nObsy\_z$  can be simplified to a single version  $nObsDR$  and the value of the *InFront* is "na",

Now for  $x = p$ ,  $y$  in  $pObsy\_z(A,B)$  corresponds to the relation  $PO(A_p, B_p)$ ,  $PP(A_p, B_p)$ , and  $PPc(A_p, B_p)$ . From  $PO(A_p, B_p)$  there are two relations for *InFront* value A, B. There are two other relations, one from  $PP(A_p, B_p)$  with *InFront* equal to A, one from  $PPc(A_p, B_p)$  with *InFront* equal to B. There are 4 *partial* obscurations in all. They are named descriptively where these come from.

Now for  $x = e$ ,  $y$  in  $eObsy\_z(A,B)$  corresponds to the relation  $EQ(A_p, B_p)$ ,  $PO(A_p, B_p)$ ,  $PP(A_p, B_p)$ ,  $PPc(A_p, B_p)$ . In each case, there is one predicate for equal obscuration with *InFront* value E. There are 4 *equal* obscuration predicates.

Now for  $x = m$ ,  $y$  in  $mObsy\_z(A,B)$  corresponds to the relation  $EQ(A_p, B_p)$ ,  $PO(A_p, B_p)$ ,  $PP(A_p, B_p)$ ,  $PPc(A_p, B_p)$ . In each case, there is one predicate for mutual obscuration with *InFront* value M. There are 4 *mutual* obscuration predicates.

Now for  $x = c$ ,  $y$  in  $cObsy\_z(A,B)$  corresponds to  $EQ(A_p, B_p)$  or  $PPc(A_p, B_p)$  or  $PP(A_p, B_p)$ . From  $EQ(A_p, B_p)$ , there are two relations for *InFront* value A, B. There are two other relations, one from  $PP(A_p, B_p)$  with *InFront* equal to B, one from  $PPc(A_p, B_p)$  with *InFront* equal to A. There are 4 *complete* obscurations in all.

### 5.3 Occlusion Predicates using first order logic

Now that we have described the obscuration predicates in natural language, we will define them formally in first order logic as follows. The complete and comprehensive occlusion relations  $nObs$ ,  $pObs$ ,  $eObs$ ,  $mObs$ ,  $cObs$  supported with  $y$  for RCC5 relation between projections and  $z$  for parameter *InFront* are coherently denoted by  $xObsy\_z$ .

There is one  $nObs$  occlusion relation:

$$nObsDR(A,B) \equiv DR(A_p, B_p) \wedge (InFront(A,B) == "na")$$

There are 4 types of  $pObsy\_z$  occlusion relations. There are two predicates from  $PO(A_p, B_p)$  with parameter *InFront* values A, B for partial obscuration. There is one predicate from  $PP(A_p, B_p)$ , with *InFront*=A and there is one predicate from  $PPc(A_p, B_p)$ , with *InFront*=B for partial obscuration, namely:

$$\begin{aligned} pObsPO\_a(A,B) &\equiv PO(A_p, B_p) \wedge (InFront(A,B) == "A") \\ pObsPO\_b(A,B) &\equiv PO(A_p, B_p) \wedge (InFront(A,B) == "B") \\ pObsPPc\_b(A,B) &\equiv PPc(A_p, B_p) \wedge (InFront(A,B) == "B") \\ pObsPP\_a(A,B) &\equiv PP(A_p, B_p) \wedge (InFront(A,B) == "A") \end{aligned}$$

There are 4 types of  $eObsy\_e$  occlusion relations. For  $y$ , there corresponds one predicate from each  $PO(A_p, B_p)$ ,  $EQ(A_p, B_p)$ ,

$PP(A_p, B_p)$ , and  $PPc(A_p, B_p)$  with  $InFront=E$  when objects A and B are equidistant, namely :

$$\begin{aligned} eObsEQ\_e(A, B) &\equiv EQ(A_p, B_p) \wedge (InFront(A, B) == "E") \\ eObsPO\_e(A, B) &\equiv PO(A_p, B_p) \wedge (InFront(A, B) == "E") \\ eObsPPc\_e(A, B) &\equiv PPc(A_p, B_p) \wedge (InFront(A, B) == "E") \\ eObsPP\_e(A, B) &\equiv PP(A_p, B_p) \wedge (InFront(A, B) == "E") \end{aligned}$$

There are 4 types of  $mObs\_m$  occlusion relations. For y, there is one predicate from each  $PO(A_p, B_p)$ ,  $EQ(A_p, B_p)$ ,  $PP(A_p, B_p)$ , and  $PPc(A_p, B_p)$  with  $InFront=M$  when objects A and B are properly cross, namely :

$$\begin{aligned} mObsEQ\_m(A, B) &\equiv EQ(A_p, B_p) \wedge (InFront(A, B) == "M") \\ mObsPO\_m(A, B) &\equiv PO(A_p, B_p) \wedge (InFront(A, B) == "M") \\ mObsPPc\_m(A, B) &\equiv PPc(A_p, B_p) \wedge (InFront(A, B) == "M") \\ mObsPP\_m(A, B) &\equiv PP(A_p, B_p) \wedge (InFront(A, B) == "M") \end{aligned}$$

There are 4 types of  $x=c$  occlusion relations. There are two from PP and PPc with values B and A and there are two from EQ with values of  $InFront$  as A, B, for complete obscuration.

$$\begin{aligned} cObsPPc\_a(A, B) &\equiv PPc(A_p, B_p) \wedge (InFront(A, B) == "A") \\ cObsPP\_b(A, B) &\equiv PP(A_p, B_p) \wedge (InFront(A, B) == "B") \\ cObsEQ\_a(A, B) &\equiv EQ(A_p, B_p) \wedge (InFront(A, B) == "A") \\ cObsEQ\_b(A, B) &\equiv EQ(A_p, B_p) \wedge (InFront(A, B) == "B") \end{aligned}$$

This is a complete classification of seventeen JEPD unique occlusion relations, see Table 2.

RCC5	Infront			
	A	B	E	M
	n			
	DR	p	p	e
	PO	c	c	e
PP	p	c	e	m
	PPc	c	p	e

Table 2 Let x be table entry,  $y=RCC5(A_p, B_p)$ ,  $z=InFront(A, B)$ , then  $xObs\_z(A, B)$ . The first row entry indicates, there is no occlusion irrespective of  $InFront$  value.

#### 5.4 Tabular form of Occlusion Predicates

We have completely described the occlusion predicate  $xObs\_z(A, B)$ , see Table 2. Now we have one predicate for nObs, four predicates each for pObs, eObs, mObs, and cObs. There are seventeen JEPD predicates in all, described in natural language and in first order logic. We can reduce them to 7 by suppressing the y but adding the details of y in table entries:  $xObs\_z$ , see Table 3. This is essential for display, but not necessary for development.

## 6. Implementation Consideration

For any theoretical development, its practical usefulness depends on the implementation followed its use in client applications. Clearly we have improved the theoretical

representation of the solution to computation of obscuration classification relations. We presented crisp ontology for obscuration relations:  $xObs\_z$ .

Table 3 Complete Set of Occlusion Predicates

$xObs\_z$	IntInt	IntBnd	BndInt	InFront	converse
nObs	F	F	F	na	nObs
pObs_a	T	F	T	A	pObs_b
pObs_b	T	T	F	B	pObs_a
eObs_e	T	F	F	E	eObs_e
mObs_m	T	F	T	M	mObs_m
cObs_a	T	F	F	A	cObs_b
cObs_b	T	F	T	B	cObs_a

First viewpoint and viewplane are selected. The objects are projected on the view plane. With the projections of the objects, RCC5 relations are determined using IntInt, IntBnd, and BndInt predicates with the 2D projections  $A_p$  and  $B_p$  of the 3D objects A and B, respectively.

The 5-step *algorithm* for obscuration detection becomes:

**Algorithm** for  $xObs\_z$  determination

**Input:** objects A and B, view point V and projection plane P.

**Output:** predicate  $xObs\_z$

1. Project Objects A and B, determine  $A_p$  and  $B_p$
2. Determine RCC5 relations between  $A_p$  and  $B_p$
3. Determine InFront parameter values
4. Integrate steps 2 and 3 to classify the obscuration type

There are standard algorithms for step 1, and 2. The step 3 is most complex and computation intensive in practice. In order to determine the obscuring object, as shown in Section 5.4, a semi-infinite ray is drawn from viewer through points in  $A_p \cap B_p$  and analyzed for intersection with the objects. This is a computation intensive step as it is repeated thousands of times depending on digitization of projection plane.

Computations of ray intersections in step 3 can be eliminated altogether by judiciously performing the step 1. As soon as the projection is computed, we know the functional relation between objects and their projections: A to  $A_p$ , and B to  $B_p$ . We can record it in step 1 to use it in step 3 as a lookup table to avoid repeated ray intersections. This can be done with an appropriate intelligent grid data structure that keeps

track of the closest *intersection* points on the objects. Now for step 3, we can look up the computed value for each  $A_p$  grid point. This eliminates tens of thousands of ray-object intersection computations.

By using this heuristic the algorithm can be implemented very efficiently. Test case was written in Python and implemented on Apple, using synthetic data of 500 objects of various shapes. The simulation showed a remarkable improvement. Computation efficiency will increase significantly if one object information is reused for obscuration with several other objects in the application.

### 6.1 Hybrid Spatial and Occlusion Relations

The same reasoning also works for mobile objects and stationary observer. Topological relation is a static RCC8 relation for 3D objects. Static relation is independent of the observer, it is the same for every observer. The occlusion relation ( $xObs_y_z$ ) is the spatial dynamic relation as seen by the viewer. The dynamic occlusion relation varies from viewer to viewer location. By consolidating the two, we have complete hybrid spatial relations. If  $R$  is an RCC8 relation in 3D, and  $xObs_y_z$  is occlusion relation, then  $R$  and  $xObs_y_z$  in tandem coalesce to represent coherent spatial relation,  $R\_xObs_y_z$ .

For the sake of simplicity and space availability, we suppress  $y$ , and display the relations in the form  $R\_xObs_z$ . There are 8 RCC8 connectivity relations and 7  $xObs_z$  occlusion relations. Not all obscuration relations are physically possible with each RCC8 relation. There are 23 hybrid relations: 5 DC, 5 EC, 6 PO, 1 EQ, 2 TPP, 2 TPPC, 1 NTPP, 1 NTPPC relations, see Fig. 8.

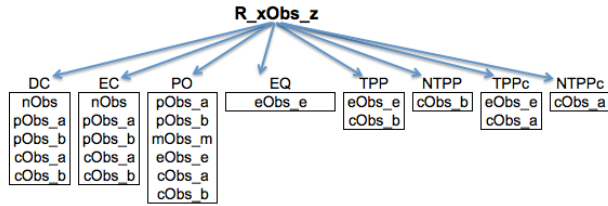


Fig. 8. A hierarchy tree of composite spatial relations.

## 7. Conclusion and Future Directions

We have given a complete description and classification of qualitative spatial occlusion relations for 3D objects as seen by an observer. The same reasoning works well with the objects that are mobile and the observer is stationary. The spatial relations are self-documenting and easy to understand. We optimized the set of occlusion relations from 12 to 7 and reduced the composite relations from 34 to 23. These computations are performed repeatedly in any application. This development will be useful in GIS, robotic sensors for

navigation, biomedicine, and related areas. Conceptual neighborhoods and composition tables are integral part of any qualitative spatial reasoning system, we plan to develop these ideas to produce conceptual neighborhood graphs and composition tables. This work also supersedes the existing 3D spatial reasoning systems.

## 8. References

- [1] B. Bennett.: *Spatial Reasoning With Propositional Logics*, Proceedings of the 4th International Conference on Principles on Knowledge Representation and Reasoning (KR-94), Bonn, Germany, pp. 165-176, 1994.
- [2] Galata, A., Cohn, A., Magee, D., and Hogg, D.: *Modeling interaction using learnt qualitative spatio-temporal relations and variable length markov models*. In *ECAI*, pp. 741-746, 2002.
- [3] A.P. Galton, Lines of Sight, in *AISB Workshop on Spatial and Spatio-Temporal Reasoning*, 1994.
- [4] D.A. Randell, M. Witkowski, and M. Shanahan, *From Images to Bodies: Modelling and Exploiting Spatial Occlusion and Motion Parallax*, *IJCAI-01*, pp. 57-63, 2001.
- [5] Kohler, C.: *The Occlusion Calculus*. In: Cognitive Vision Workshop. ICVW '02, pp. 1-6, 2002.
- [6] J. Albath, J. Leopold, and C. Sabharwal, Visualization of Spatio-Temporal Reasoning Over 3D Images, *Proceedings of the 2010 International Workshop on Visual Languages and Computing* (in conjunction with the 16<sup>th</sup> International Conference on Distributed Multimedia Systems), pp. 277-282, 2010.
- [7] Guha, P., Mukerjee, A., Venkatesh, K.S.: OCS-14 : *You Can Get Occluded in Fourteen Ways*. In: Walsh, T. (ed.) *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*. pp. 1665-1670, 2011.
- [8] C. L. Sabharwal, J. L. Leopold, and N. Eloë: *A More Expressive 3D Region Connection Calculus* In The 2011 International Workshop on Visual Languages and Computing (in conjunction with the 17th International Conference on Distributed Multimedia Systems (DMS'11)), pp. 307-311, 2011.
- [9] N. Eloë, C. Sabharwal and J. Leopold: *A More Efficient Representation of Obscuration for VRCC-3D+ Relations*, *Polibits Journal of Computer Science*, pp. 29-34, 2014.
- [10] D.A. Randell, Z. Cui, A.G. Cohn, *A Spatial Logic Based on Regions and Connection* KR92:165-176, 1992.
- [11] M. J. Egenhofer, R. Franzosa, *Point-Set topological Relations*, *International Journal of Geographical Information Systems* 5(2), pp.161-174, 1991.

# BopoNoto: An Intelligent Sketch Education Application for Learning Zhuyin Phonetic Script

Paul Taele and Tracy Hammond

Sketch Recognition Lab  
Texas A&M University  
College Station, TX 77843 USA  
{ptaele,hammond}@cse.tamu.edu

**Abstract**— The zhuyin phonetic script not only provides greater learning benefits compared to romanization systems for Chinese Mandarin language students with existing English fluency, but also allows students to take advantage of its use in Taiwan where it exists in both official and popular usage. However, while pen-enabled computing educational applications can assist traditional pedagogical approaches for accelerating mastery of the script, existing approaches are either constrained in providing writing assessment, catered to native language users, or are less flexible in recognizing more natural writing. We present BopoNoto, an intelligent sketch education application for language students to learn zhuyin. Our application provides a sketching interface for practicing the symbols, and a sketch recognition system for assessing the visual and technical correctness of their input. From our evaluations, BopoNoto successfully demonstrates strong results in understanding and assessing students’ written input.

**Keywords**— *sketch recognition; intelligent user interface; intelligent tutoring system; Chinese Mandarin; zhuyin; bopomofo*

## I. INTRODUCTION

The written Chinese language differs from its writing counterparts in western languages such as English in that Chinese consists of characters that are inherently non-phonetic [2,33], and so mastering the pronunciation of characters in written Chinese is highly challenging for language students with primarily English language fluency [33]. There however exists different phonetic systems for written Chinese that map the non-phonetic Chinese characters to their equivalent pronunciations, including the popular hanyu pinyin that is a romanization system conventionally taught in American Chinese language classes due to students’ familiar with alphabet letters already in English [12]. However, hanyu pinyin has its own disadvantages that can be challenging for native English users learning Chinese Mandarin, such as hanyu pinyin’s orthographic rules and pronunciations not allowed or existing, respectively, in English [10].

One alternative to romanized phonetic scripts such as hanyu pinyin is the zhuyin phonetic system (Figure 1), known more formally as zhuyin fuhao, colloquially as bopomofo, and officially as Mandarin Phonetic Symbols I (MPS1). Like other native non-romanized East Asian language phonetic scripts such as hiragana and katakana in Japanese and hangul in Korean [2,16], zhuyin was designed specifically to map directly and uniquely to the sounds in Chinese Mandarin [6,11]. Furthermore, zhuyin is officially and commonly used by people in Taiwan [1,24], and is continually taught to both domestic

elementary students [17] and foreign university students [21] alike, so students familiar with the script – especially those living and studying already in Taiwan – can also take advantage of being able to understand them from their surroundings and with locals there.

B: ㄅ	P: ㄆ	M: ㄇ	F: ㄈ	D: ㄉ
T: ㄊ	N: ㄋ	L: ㄌ	G: ㄍ	K: ㄎ
H: ㄏ	J: ㄐ	Q: ㄑ	X: ㄒ	ZH: ㄗ
CH: ㄔ	SH: ㄕ	R: ㄖ	Z: ㄗ	C: ㄘ
S: ㄙ	A: ㄚ	O: ㄛ	E: ㄜ	EH: ㄝ
AI: ㄞ	EI: ㄟ	AU: ㄠ	OU: ㄡ	AN: ㄢ
EN: ㄣ	ANG: ㄤ	ENG: ㄥ	ER: ㄦ	Y: ㄩ
	U: ㄨ		YU: ㄩ	

Figure 1. The thirty-seven symbols that represent the phonetic system of zhuyin (i.e., zhuyin fuhao, bopomofo, Mandarin Phonetic Symbols I).

While zhuyin provides several important advantages to such language students with primarily existing English fluency, mastering the script involves an initially steeper learning curve, a relatively larger set of symbols, completely unfamiliar symbols, and more diverse visual complexity compared to the English letters found in its romanized phonetic script counterparts [1], where the last two features stem from the symbols’ historic origins in Chinese characters [24]. Traditional pedagogical approaches for learning zhuyin thus has parallels with related non-phonetic scripts in East Asian languages, which rely on instructors and course materials introducing rote memorization and written techniques of stroke count, order, and technique (e.g., [20]).

The importance of writing technique practice for written East Asian scripts such as zhuyin includes enabling students to improve the aesthetics of their writing [13], but unsupervised writing practice outside of the classroom – using traditional pedagogical approaches such as rote memorization – may hinder students in memorizing the symbols effectively if they unintentionally develop bad writing habits [19]. Computer-assisted educational tools have strong potential in aiding instructors in addressing such concerns by providing emulated supervision outside of the classroom when instructors are less accessible outside of class, but available educational apps

provide minimal or lacking writing feedback, relevant language handwriting interfaces focus more on best interpreting what the user wrote, and related current sketch research systems require that students write their input in a specific way to receive accurate assessment.

In this paper, we therefore introduce an intelligent sketch application called BopoNoto, which is a portmanteau of the script name ‘bopomofo’ and a common East Asian pronunciation of the word ‘note’. Our application enables students to practice learning the zhuyin phonetic script through direct writing practice and accompanying automated feedback on the visual and technical correctness of their written symbols. Our application provides students with different writing interfaces to practice their knowledge of individual symbols and the correct phonetics mapping of prompted Chinese characters, and allows them to automatically receive and review mistakes in their symbol writing.

## II. RELATED WORK

As an intelligent sketch application for the zhuyin phonetic script, we discuss related work from language education apps for learning zhuyin and other related written East Asian language scripts, handwriting input systems for inputting written East Asian languages, and sketch recognition interfaces for educational domains including written East Asian languages.

### A. Language Education Apps

With the growing ubiquity of mobile devices, developers have tapped into the technology to create digital educational interfaces including learning zhuyin (e.g., [4]). However, many of these accessible mobile app interfaces are constrained to traditional flash card and binary feedback learning, and also lack direct writing input and assessment. Other interfaces for learning symbols in other written East Asian language scripts have taken more unique approaches in creating engaging interactive group learning experiences with mobile devices [9,25,30], but do not address individual learning practice that incorporates a writing practice modality with automated assessment. More recent educational mobile apps (e.g., [23]) have moved beyond the limitations of prior flash card-based apps to include writing with assessment, but are still constrained to tracing practice and binary feedback.

### B. Handwriting Input Systems

Peripherally related to educational writing interfaces for East Asian languages such as zhuyin are intelligent handwriting input systems that enable users to naturally write out the symbols on touch- and stylus-enabled computing devices. Advances from the machine learning research community in recognizing written East Asian language symbols have been explored for decades [15], and the lessons from these research efforts have led to the development of robust handwriting input systems for written East Asian language symbols that anticipated the rising popularity of touchscreens. These works include Synaptics [18] for early-generation stylus-driven touchscreens, Microsoft [22] for tablet notebooks, and Google [7] for smartphones. However, the focus of these handwriting input systems involves best interpreting the intentions of users’ written input such as messy and incomplete input. Such systems therefore cater to input from users with existing fluency of written East Asian symbols such

as zhuyin, and may not be appropriate to novice students who have not yet properly established good writing habits.

### C. Sketch Recognition Interfaces

Two particularly related systems include Hashigo [27], an intelligent sketch education interface for helping students practice Japanese kanji; and LAMPS [28], an analogous interface for Chinese zhuyin. The strengths of these systems are that students can write on these interfaces and immediately receive feedback on the visual and technical correctness of their input. However, these systems primarily rely on geometric sketch recognition approaches, and thus perform less optimally for students’ written input that are considered still visually correct but cannot be cleanly segmented into smaller geometric components for recognition and assessment. In the case of LAMPS for zhuyin symbols, the authors highlight limitations in recognizing symbols that cannot be geometrically approximated by straight polylines, such as those containing arcs, curves, and dots.

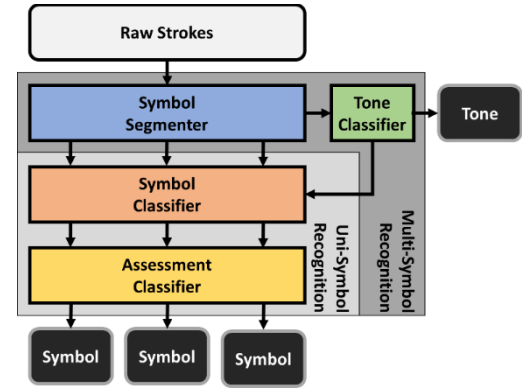


Figure 2. An overview of BopoNoto’s recognition system for classifying both individual and multiple zhuyin phonetic symbols.

## III. RECOGNITION SYSTEM

Designing the intelligent sketching interface for the BopoNoto application to robustly recognize students’ zhuyin phonetic symbols that are appropriate for classroom submission motivated us to develop a recognition system consisting of an optimized collection of gesture and sketch recognition techniques and heuristics, as shown by the system overview in Figure 2.

### A. Symbol Segmentation

The first layer of BopoNoto’s recognition system is symbol segmentation (Figure 3), which first takes the raw strokes from the students’ collection of one or more written symbols and then segments them into their individual symbols. The design of our symbol segmenter takes lessons from both the drawing behavior cues from cognitive theory research [32] and Asian calligraphy [29] for the symbol segmentation process, which involves windowing the strokes left to right and assuming an approximate square block boundary for candidate symbols to segment.

Based on these assumptions from the literature, our algorithm first determines the candidate symbol’s potential boundary size by calculating the length threshold of either the height of the entire raw strokes’ bounding box height or the width of the largest raw stroke. We desired the latter to handle



cases when the input consists of symbols with very short heights. We next iterate through and temporarily store each raw stroke from left to right based on the order that the user wrote them, and then calculate the bounding box width of all the iteratively stored strokes. If this width exceeds 125% of the length threshold, we then segment the collected strokes as a candidate symbol, reset the stored strokes, and shift the process to the next unadded strokes. Otherwise, the process continues for the next stroke until all symbols are segmented into a list of symbols for the next layer of the recognition system.

One special case involves symbols where a stroke is added to the bottom-left of the written symbol, which causes the segmenter to possibly prematurely segment the symbols in the middle of the collected strokes, since the width may exceed the boundary width from the left side of the candidate symbol. In order to alleviate this special case, we automatically add all strokes that are left of the previous added stroke, since we assume that such strokes will not form a distinct symbol due to drawing assumptions.

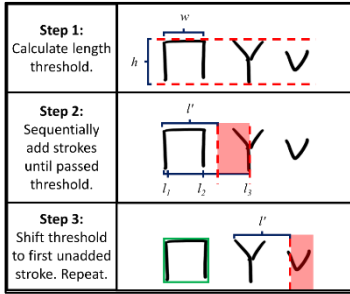


Figure 3. The general three-step process for segmenting raw strokes into a list of zhuyin symbols.

### B. Tone Symbol Recognition

Immediately following the symbol segmenter, we then take the last symbol from the collected list and test whether the symbol is instead one of the four explicit tone marks found in the zhuyin phonetic system to represent tonal sounds in Chinese Mandarin (Figure 4).

Since the four explicit zhuyin tone marks visually consist of either segmented lines or dots, we apply geometric shape tests from relevant sketch recognition algorithms. For the straight lines that form the rising second tone and the falling fourth tone, we utilize line tests from ShortStraw [35]. The process first involves removing potential garbage points from the stroke that occurs from users pressing down and lifting up their finger or stylus from the touch surface (i.e., 5% from the beginning and end of the stroke). Next, we calculate the distance of the ideal line formed from the endpoints and the path length of the stroke, then take the ratio of the lesser value to the greater value. If this ratio exceeds 0.8, we then measure the angle formed from the stroke’s endpoints. If the angle is within a threshold angle range of 45° or 120°, we classify the stroke as either a rising second or a falling fourth tone, respectively.

Our classifier then proceeds to test the stroke as a low third tone, which visually consists of a line that is initially angled down and then angled back up. One potential solution is to utilize a number of corner segmentation algorithms to detect the three corners (i.e., bottom left, top, and bottom right) in the

stroke. However, we alternatively applied geometric heuristic tests on the stroke to avoid the overhead associated with incorporating a separate corner segmenter by simply trimming the garbage points, locating the furthest left, top, and right points, and then applying similar line tests on these substrokes for angles of 120° and 60°, respectively. If the stroke passes these conditions, we classify it as a low third tone.

Finally, we tested the stroke for its potential as a neutral fifth tone – which is visually represented as a dot – by using insights from the Dahmen scribble recognizer [5]. Specifically, we first calculate the width and height of the stroke’s bounding box, then calculate the path length of the stroke, and lastly calculate the density of the stroke by taking the quotient of the stroke path length and stroke bounding box area. If this density ratio exceeds 0.3, the stroke is then classified as a neutral fifth tone. Otherwise, the stroke remains unclassified and the process proceeds to the next layer of the recognition system.



Figure 4. The four tone marks in the zhuyin phonetic script (L-R): the rising second tone, the low third tone, the falling fourth tone, and the neutral fifth tone. The high first tone is implicitly implied if the phonetic symbols lacks a tone mark in zhuyin.

### C. Symbol Classification

The crux of the recognition system lies in the individual zhuyin symbols, where segmented symbols from the symbol segmenter and possibly the tone classifier subsequently proceed through the symbol classifier. The aim of the symbol classification for our BopoNoto application is to ensure that it can classify students’ symbols that are visually similar to model written symbols. As a result, we first recruited a university Chinese Mandarin language instructor from Taiwan with several years of classroom experience to provide us with model written symbols. The data collection of these model written symbols involved prompting the instructor to write five iterations of each zhuyin symbol for a total of  $5 \times 185 = 377$  writing samples.

From the language instructor’s model written symbols, we group these sketches as training data, particularly as templates for the template matching process. Our reasoning is that template matching potentially enables us to more easily identify written input that is pedagogically visually correct instead of simply determining the input’s best interpretation found in other machine learning techniques that may not penalize for students’ sloppy writing. We therefore initially experimented with various naive and modified Euclidean and Hausdorff distance-based template matching algorithms on the individual symbols for their classification feasibility, and empirically observed that existing approaches in the literature that were successful for smaller sets of symbols and were appropriate as touchscreen gestures had performed less optimally for a larger zhuyin phonetic symbol set of nearly forty symbols and of relatively greater visual complexity.

We eventually developed a two-part template matching algorithm designed to perform well in classifying zhuyin symbols by relying on two different metrics: a specialized



Hausdorff distance-based score and a stroke points coverage ratio (Figure 5).

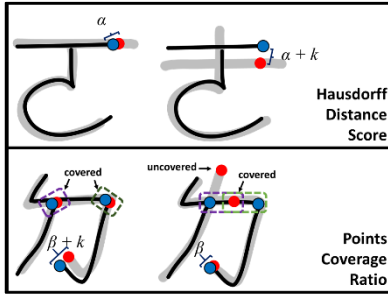


Figure 5. The individual symbol classification process that relies on both a Hausdorff distance score and a points coverage ratio.

### 1) Hausdorff Distance Score

The first metric is the Hausdorff distance score, and we calculate this by normalizing (i.e., resample, scale, and translate) the individual symbol's raw strokes and classifying them using a specialized Hausdorff distance-based template matcher [26] derived from existing template matchers (e.g., [14]) and that was adopted in this paper for relatively more complex symbols such as those found in zhuyin. One such special case is for a zhuyin phonetic symbol consisting of a single horizontal line, which has issues for general template matching algorithms due to rescaling it into a square bounding box [34]. To compensate for this special case, we do a pre-test by first determining if the symbol is single-stroke, a line, and horizontal. If so, we automatically classify this stroke as that particular zhuyin phonetic symbol.

The next step involves the actual score calculation conventionally performed in Hausdorff distance template matching, which involves iterating through each resampled point in the written symbol, locating the closest corresponding point in the candidate model template, calculating the Euclidean distance, and then taking a running sum of that distance into the combined sum of distances. We then calculate the score from the equation below [26], which is a scaled variation of the score employed in [34] but for relatively more complex symbols.

### 2) Points Coverage Ratio

The second metric that we use in the symbol classification process specifically addresses issues of visually similar symbols, which occurs for a non-trivial number of zhuyin symbols. We initially observed that misclassification cases occurred for symbols that were visually similar due to the fact that the subset of points from model template of an incorrect symbol type overlaps more closely than the entire set of points from the closest model template of the correct symbol type. Due to this observation, we set forth to calculate the ratio of points from the model template that was paired as the shortest distance to the points in the student's written input.

With the information from the second metric, we then go back to the sorted metric from the first metric and locate the model template with the highest coverage ratio in the top 10% score-ranked model templates. The zhuyin symbol type of the model template that satisfies the conditions from these two metrics is eventually used to classify the student's written input.

## D. Assessment Classification

Once the symbol is visually classified from the previous layer of the recognition system through the symbol classifier, we finally assess the technical correctness of the written symbol for visually correct symbols. This assessment consists of three technical correctness tests for the following: stroke count, stroke order, and stroke direction.

### 1) Stroke Count Test

The first test is in determining the correct number of strokes. To perform this, we simply count the number of strokes from the students' written input to those from the model template. If there is a mismatch in the count, the written input is assessed as having an incorrect stroke count.

### 2) Stroke Order Test

For symbols that pass the stroke count test, we then perform a stroke order test. This process first involves pairing the strokes from the students' written input to their equivalent strokes from the model template. In order to do so, we approximate optimal pairings of stroke pairs between the students written input and the model template by taking the start, middle, and end point of the written input and model template, and calculate the summed distances from all three pair of points. Whichever stroke's trio of points from the model template has the shortest distance from that written input stroke is therefore paired to that model template's stroke. The process continues until all strokes from the written input are paired to their equivalent strokes in the model template.

The next step then relies on sorting the pairs of strokes from between the written input and model template in ascending temporal order of the model templates strokes, since we assume that the model template will have the correct temporal ordering. From this sorting information, we also check to see if the written input's strokes are also correctly sorted in ascending time. If not, we record the index of the first instance of the temporally misplaced stroke and assess the written input as having incorrect stroke order.

### 3) Stroke Direction Test

The last assessment check is in the correctness of the stroke direction. For cases where the written input is assessed as having correct stroke order, we rely on the list of temporally-ordered paired strokes by iteratively comparing the endpoints of the model template's stroke with those of the student's written input. We first try to find the corresponding endpoints of the model template to the written input by first determining whether the endpoints of the written input's stroke approximates an ideal vertical line. If so, we then pair the topmost point of the written input's stroke to the topmost point of the model template's stroke and likewise for the bottommost point. Otherwise, we pair the corresponding endpoints by the leftmost and rightmost points.

Once the endpoints of the corresponding strokes from the written input and model template are mutually paired with each other, we then perform a Boolean check on the temporal ordering of the endpoints of the model template stroke and then on the written input. That is, we check if the top or left endpoint temporally occurs before the bottom or right endpoint, respectively, for the model template stroke, and likewise for the written input stroke. If the Boolean check is equivalent for both

the model template and the written input for that particular stroke, we then consider that stroke to have correct stroke direction. We iteratively apply this stroke direction test on each stroke of the written input, and classify the entire symbol as having correct stroke direction. Otherwise, we record the index of the first instance of the incorrect stroke direction and label the symbol as such.

#### IV. SKETCHING INTERFACE

We subsequently incorporate our recognition system that automatically assesses students' visual and technical correctness on their written zhuyin symbols input into two intelligent sketch user interfaces: one for writing practice on prompted individual symbols and one for writing practice on the phonetic symbol mapping of prompted Chinese characters.

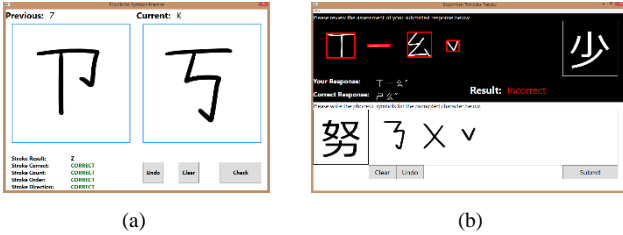


Figure 6. BopoNoto's different writing interfaces: (a) symbol writing practice interface and (b) phonetics writing interface.

##### A. Symbol Writing Practice Interface

The first intelligent sketch user interface in the BopoNoto application allows students to practice writing individual symbols and subsequently receive visual and technical correctness, which is appropriate for students who are initially learning about the zhuyin phonetic script or who wish for a refresher of the zhuyin phonetic symbols directly (Figure 6a). This particular interface consists of randomly displaying the approximate English transliteration of the symbol for the user to write; the user then provides their solution in a sketching canvas on the right side of the interface. The user has access to conventional interaction buttons to undo or clear their strokes, and can submit their solution for the interface.

Following the submission of their written input, the left side of the interface subsequently displays a visual representation of any mistakes in their sketch for review (see Figure 7), while also displaying textual information listing the technical correctness of their submitted written input.

##### B. Phonetics Writing Practice Interface

The second intelligent sketch user interface allows students to practice writing the phonetics of different Chinese characters (Figure 6b). That is, for each prompted Chinese character, the user must write the phonetic symbols associated with the pronunciation of that character. This particular interface is well-suited for students who have become more confident in their familiarity of the zhuyin phonetic symbols.

The specifics of this particular interface follows similarly to the symbols writing practice interface, where users first write their written multi-symbol input and then submit their answer for feedback. Their submitted answer is then highlighted to visually demonstrate how closely proportioned their individual

symbols are, as well as indicating how their input compares to the expected model solution.

##### C. Assessment and Feedback Display

For the symbols writing practice interface, we provide two types of feedback for communicating the assessment of the students' written zhuyin phonetic symbol input. The first way is through text that lists the different written technique assessments and corresponding correctness level. The second way is through visual cues that overlap the original written input with either highlighted strokes and points or examples of the expected model template solution.

Examples demonstrating the specific visualizations of the four assessment cases of incorrect visual structure or written technique can be found in Figure 7. For incorrect stroke count, the user is shown both an example of the expected model template solution, as well as highlighted stroke endpoints from both the written input and model template, where the blue endpoints on the model template indicate the correct stroke. For incorrect stroke direction, the first temporally incorrect stroke is highlighted in red. For incorrect stroke direction, the temporally incorrect starting endpoint of the stroke and the stroke itself is highlighted as an orange point and a red stroke, respectively. For incorrect symbols, the submitted written input is overlaid with the strokes from a model template solution as orange strokes. Lastly, correct submitted written input – not shown in Figure 7 – does not provide any visual cues as a way to notify the user that there are no errors within their solution.

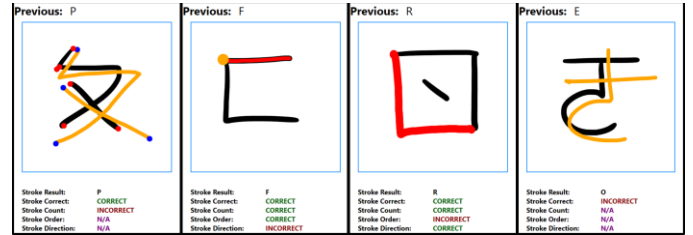


Figure 7. Visualizations of the different assessment cases of incorrect visual structure or written technique (L-R): incorrect stroke count, incorrect stroke direction, incorrect stroke direction, and incorrect symbol.

#### V. RESULTS AND DISCUSSION

Since our BopoNoto application is an intelligent sketch user interface, we approach evaluating our application from its recognition performance, where we evaluate how well it classifies users' written symbols and compare it to related zhuyin sketch education application LAMPS [28]; and also from its interaction performance, where we evaluate how well it assesses written symbols and whether users understood how to use the writing practice interfaces.

##### A. Recognition Evaluation

###### 1) Overall Symbol Recognition Performance

For evaluating the performance of BopoNoto's recognition system, we recruited ten participants – three females – who consisted of one former Chinese as a Second Language (CSL) instructor with native zhuyin fluency, four Taiwanese international engineering graduate students with native zhuyin fluency, and two American university students with no prior exposure with zhuyin, and three American university students

with similar lack of experience but existing experience other East Asian written language scripts. We prompted each participant to provide five writing samples of each of the zhuyin phonetic symbols for a total of  $37 \times 5 \times 10 = 1850$  writing samples, and provided them with instructions prior to the data collection dependent on their existing zhuyin fluency. For our native zhuyin writers, we only stated that the data will be used for an educational writing application, while the non-native zhuyin writers were provided a list of zhuyin phonetic symbols only as reference for providing the data.

Following the completion of the data collection, we subsequently tested the written symbols with BopoNoto’s recognition system. From our recognition, we observed that our system recognized users’ written symbols very well overall, with misclassifications occurring only on three different symbols – B, P, and L – that were misclassified at least once for symbols K, OU, and D, respectively. Overall recognition for all symbols tested from the users’ collected test data exceeded 99% using an all-or-nothing evaluation approach [35].

## 2) Recognition Performance versus LAMPS

We conducted an additional evaluation of BopoNoto by testing its recognition system on related work LAMPS on zhuyin phonetic symbols, where we used the written symbols test data from the former CSL instructor’s provided written symbol input. While we evaluate the overall recognition performance directly on the recognition system for BopoNoto, we take a different approach with LAMPS and present a weaker constraint that involves testing LAMPS’ accuracy on how well corner-finding algorithms can segment the strokes. The reason is that since LAMPS relies exclusively on geometric sketch recognition for classifying the written zhuyin phonetic symbols, we can therefore provide a recognition performance upper bound based on whether the correct number of corners were located. This is because LAMPS assumes that corner information helps determine the written input’s segmented line components that approximate zhuyin phonetic symbols.

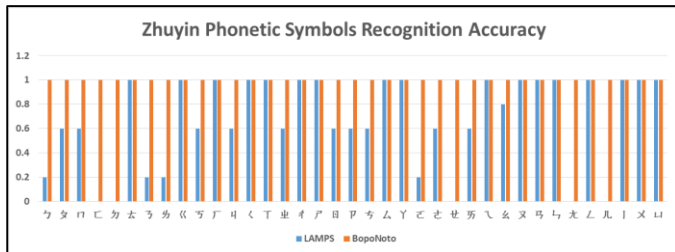


Figure 8. Side-by-side comparison of recognition results between LAMPS and BopoNoto on zhuyin phonetic symbols from a language instructor.

BopoNoto's recognition system was tested directly while LAMPS' recognition system was tested on number of correct line corners from a state-of-the-art corner finding algorithm.

The results of the second recognition performance is shown in Figure 8, where BopoNoto excels on successfully classifying the entire zhuyin phonetic symbols. On the other hand, the recognition performance from LAMPS solely on corner finding information performed less successfully. We speculate that the recognition rate was lower for LAMPS due to two factors: 1) the participants in the original LAMPS publication were given specific instructions to write the symbols as if demonstrating to

someone not familiar with them, while we gave no specific constraints, and 2) the participant for BopoNoto was not given specific writing instructions and chose to write the symbols with a more natural and curvier and casual style that happens to be problematic to geometric sketch recognitions.

### B. Interface Evaluation

In evaluating the interfaces used in BopoNoto, we take a more informal approach by first asking users with lacking native zhuyin writing experience to write the symbols without reference to the correct stroke order. For the five domestic student participants, we reused their training data by sending the data to the assessment classifier. The goal was to observe symbols made with incorrect written technique and whether BopoNoto can identify these misclassifications. After processing the data and manually identifying the technically incorrect symbols, we ran these technically incorrect symbols to the assessment classifier that was then able to replicate identifying all of these symbols as technically incorrect and its associated type of technical error.

Lastly, we briefly observed whether the participants were able to understand how to interact with the two writing practice interfaces, specifically whether they knew where to write their input and read their writing assessments. From our observations and informal inquiries of the participants following the study, we discovered that the participants had overall either figured out how to use the interface immediately or quickly figured it out after a brief trial and error with the interactive areas.

## VI. FUTURE WORK

With the lack of intelligent sketch user systems focused on writing practice of East Asian language scripts such as zhuyin, one of our main aims of this research work was to determine whether we would be able to develop a robust recognition system and working sketch education application specific to written zhuyin phonetic symbol instruction. We believe that from our evaluations that we have succeeded and hope to continue the momentum from this research work to develop more encompassing interfaces for written zhuyin instruction, such as the addition of tutorial modes and the inclusion of phonetics writing practice on more sets of Chinese characters.

Additionally, we would like to take the lessons and energies of our research work to collaborate actively with Chinese Mandarin instructors so that our application more closely tie into existing curriculum plans.

Lastly, we hope to extend the accomplishments of our recognition system by adapting them to other written East Asian language scripts that may be even more visually complex. There is interesting potential in observing how well the recognition system and interaction capabilities are similarly appropriate and compatible with Korean hangul and Japanese hiragana.

## VII. CONCLUSION

In this paper, we describe our intelligent sketch education application work called BopoNoto, which is designed to assist students in more optimally learning how to write zhuyin phonetic symbols. Our application provides a sketch interface that prompts users to write their answer out, and then our application’s recognition system can automatically assess the

visual and technical correct of students' written input. From our evaluations, we discover that our application succeeds in robust recognition and understandable interaction of zhuyin phonetic symbols.

## REFERENCES

- [1] Ager, S. Zhuyin fuhao / bopomofo. <http://bit.ly/1I2OSL4>, 2015. [Online; accessed 31-May-2015].
- [2] Campbell, L. Historical Linguistics: An Introduction, 3 ed. The MIT Press, Cambridge, 2013.
- [3] Cheema, S., and LaViola, J. Physicsbook: A sketch-based interface for animating physics diagrams. In Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces, IUI '12, ACM (New York, NY, USA, 2012), 51–60.
- [4] Chiu, C. Learn bopomofo! <http://bit.ly/1SRzTXw>, 2011. [Online; accessed 31-May-2015].
- [5] Dahmen, K., and Hammond, T. Distinguishing between sketched scribble look alike. In Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 3, AAAI'08, AAAI Press (2008), 1790–1791.
- [6] DeFrancis, J. Chinese Language: Fact and Fantasy, 1 ed. University of Hawaii Press, Honolulu, 1984.
- [7] Deselaers, T., Keysers, D., Rowley, H., Wang, L.-L., Carbune, V., Popat, A., and Narayanan, D. Google handwriting input in 82 languages on your android mobile device. <http://bit.ly/1FM4Gvg>, 2015. [Online; accessed 31-May-2015].
- [8] Dixon, D., Prasad, M., and Hammond, T. icandraw: Using sketch recognition and corrective feedback to assist a user in drawing human faces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, ACM (New York, NY, USA, 2010), 897–906.
- [9] Edge, D., Searle, E., Chiu, K., Zhao, J., and Landay, J. A. Micromandarin: Mobile language learning in context. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, ACM (New York, NY, USA, 2011), 3169–3178.
- [10] Everson, M. E. Issues in chinese literacy learning and implications for teacher development. In Issues in Chinese Language Education and Teacher Development, P. Duff and P. Lester, Eds. University of British Columbia: UBC Centre for Research in Chinese Language and Literacy Education, 2008, 70–78.
- [11] Government Information Office Republic of China. Taiwan Yearbook 2006: The People & Languages. Government Information Office Republic of China, Taipei, 2006.
- [12] He, W. W., and Jiao, D. Curriculum design and special features. In Teaching and Learning Chinese: Issues and Perspectives, J. Chen, C. Wang, and J. Cai, Eds. Information Age Publishing, Charlotte, 2010.
- [13] Heisig, J. W. Remembering the Kanji I: A Complete Course on How Not to Forget the Meaning and Writing of Japanese Characters, 4 ed. Japan Publications Trading Company, Tokyo, 2001.
- [14] Kara, L. B., and Stahovich, T. F. An image-based, trainable symbol recognizer for hand-drawn sketches. Computers & Graphics. 29, 4. (Aug. 2005), 501–517.
- [15] Liu, C.-L., Jaeger, S., and Nakagawa, M. Online recognition of chinese characters: The state-of-the-art. IEEE Trans. Pattern Anal. Mach. Intell. 26, 2 (Jan. 2004), 198–213.
- [16] Lunde, K. CJKV Information Processing: Chinese, Japanese, Korean & Vietnamese Computing, 2 ed. O'Reilly Media, Sebastopol, 2009.
- [17] Mandarin Daily News. Chinese education online - bopomofo. <http://www.mdnkids.com/BoPoMo/>, 2015. [Online; accessed 31-May-2015].
- [18] Matic, N. P., Platt, J. C., and Wang, T. Quickstroke: An incremental online chinese handwriting recognition system. In Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02) Volume 3 - Volume 3, ICPR '02, IEEE Computer Society (Washington, DC, USA, 2002), 30435–.
- [19] McNaughton, W., and Ying, L. Reading & Writing Chinese: Traditional Character Edition, 2 ed. Tuttle Publishing, North Clarendon, 1999.
- [20] National Taiwan Normal University. Practical Audio-Visual Chinese 1, 2 ed. Zheng Zhong/Tsai Fong Books, Taipei, 2008.
- [21] National Taiwan Normal University. Mandarin training center. <http://bit.ly/1FmNBaJ>, 2015. [Online; accessed 31-May-2015].
- [22] Pittman, J. A. Handwriting recognition: Tablet pc text input. Computer 40, 9 (Sept 2007), 49–54.
- [23] Sigma Sky, LLC. write bopomofo. <http://apple.co/1I2X11V>, 2014. [Online; accessed 31-May-2015].
- [24] Su, Q. G. Bopomofozhuyin fuhao: an alternative to pinyin. <http://abt.cm/1GTaebc>, 2008. [Online; accessed 31-May-2015].
- [25] Syson, M. B., Estuar, M. R. E., and See, K. T. Abkd: Multimodal mobile language game for collaborative learning of chinese hanzi and japanese kanji characters. In Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 03, WI-IAT '12, IEEE Computer Society (Washington, DC, USA, 2012), 311–315.
- [26] Taele, P., Barreto, L., and Hammond, T. Maestoso: An intelligent educational sketching tool for learning music theory. In Proceedings of the Twenty-Seventh Innovative Applications of Artificial Intelligence Conference, IAAI '15 (2015), 3999–4005.
- [27] Taele, P., and Hammond, T. Hashigo: A next-generation sketch interactive system for japanese kanji. In Proceedings of the Twenty-First Innovative Applications of Artificial Intelligence Conference, IAAI'09, AAAI Press (2009), 153–158.
- [28] Taele, P., and Hammond, T. Lamps: A sketch recognition-based teaching tool for mandarin phonetic symbols i. J. Vis. Lang. Comput. 21, 2 (Apr. 2010), 109–120.
- [29] Takezaki, K., and Godin, B. An Introduction to Japanese Kanji Calligraphy. North Clarendon, Tokyo, 2008.
- [30] Tian, F., Lv, F., Wang, J., Wang, H., Luo, W., Kam, M., Setlur, V., Dai, G., and Canny, J. Let's play chinese characters: Mobile learning approaches via culturally inspired group games. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, ACM (New York, NY, USA, 2010), 1603–1612.
- [31] Valentine, S., Vides, F., Lucchese, G., Turner, D., hoe Kim, H., Li, W., Linsey, J., and Hammond, T. Mechanix: A sketch-based tutoring system for statics courses. In Proceedings of the Twenty-Fourth Innovative Applications of Artificial Intelligence Conference, IAAI '12 (2012), 2253–2260.
- [32] van Sommers, P. Drawing and Cognition: Descriptive and Experimental Studies of Graphic Production Processes. Cambridge University Press, Cambridge, 1984.
- [33] Wang, H.-H., and Honig, A. S. What difficulties do children experience while learning to read and write chinese. In Teaching and Learning Chinese: Issues and Perspectives, J. Chen, C. Wang, and J. Cai, Eds. Information Age Publishing, Charlotte, 2010.
- [34] Wobbrock, J. O., Wilson, A. D., and Li, Y. Gestures without libraries, toolkits or training: A \$1 recognizer for user interface prototypes. In Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, UIST '07, ACM (New York, NY, USA, 2007), 159–168.
- [35] Wolin, A., Eoff, B., and Hammond, T. Shortstraw: A simple and effective corner finder for polylines. In Proceedings of the Fifth Eurographics Conference on Sketch-Based Interfaces and Modeling, SBM'08, Eurographics Association (Aire-la-Ville, Switzerland, Switzerland, 2008), 33–40.

# Visually Mapping Requirements Models to Cloud Services

Shaun Shei, Aidan Delaney, Stelios Kapetanakis and Haralambos Mouratidis

## Abstract

*We extend an existing visual language for requirements modelling to model the requirements of cloud services. To achieve this we demonstrate how candidate cloud services can be identified from existing visual requirements models. We further extend the meta-model of the visual language to include cloud requirements in order to migrate our candidate service to a cloud provider.*

## 1 Introduction

Cloud computing allows the provision of a wide range of services through the abstraction of physical and virtual resources. This offers seemingly unlimited scalability, availability, flexibility and dynamic provisioning through a pay-per-use model. However, some organisations are still hesitant to fully commit to this technology due to negative publicity regarding data-breaches [1, 2], security leaks [3] as well as interoperability and compatibility issues when migrating towards cloud environments [4, 5, 6]. Cloud computing is built upon and extends several established concepts and technologies such as Service-Oriented Architecture (SOA), distributed computing and virtualization. Moreover, through extension of existing technologies we also inherit the security issues and vulnerabilities of each [7]. This creates a complicated scenario where we need to consider security from multiple perspectives. One method for tackling this problem is to adopt an agent oriented software engineering approach [8], where the focus is placed on analysing components and properties to elicit requirements for a software system. There is currently a lack of standardised modelling languages and approaches to holistically capture cloud computing environments in the context of software security. The existing notations capture aspects of different service models such as Software-as-a-Service [6], Platform-as-a-Service [9] or Infrastructure-as-a-Service [10]. There is a lack of a holistic modelling language that captures both the customer requirements

and cloud services corresponds to the need for secure software systems in the industry [11]. Security is a concept that is often tacked on after the design and deployment of software systems, where security mechanisms are introduced in response to vulnerabilities as they appear. Security-by-design is a branch within recent research efforts [12], where the goal is to obtain a clear understanding of security issues early in the software development process. The Secure Tropos methodology provides a modelling language that represents security requirements through security constraints, allowing developers to model software systems and its organizational environment using actors, goals and relational links such as dependencies. The contributions presented in this paper are as follows:

- We define a pattern for service identification based on grouping a goal, plans and resources in software systems.
- We then model an initial description of properties required by services when migrated to cloud environments.

The rest of this paper is organized as follows. In section 2 we present an overview of the the Secure Tropos visual language. In section 3 we provide a standard definition of a software service and demonstrate how such services can be identified from a Secure Tropos requirements model. Section 4 extends the meta-model for Secure Tropos to incorporate cloud computing requirements as identified from the literature and describes how to adapt the identification of services towards a cloud environment. Finally, in section 5 we present our conclusion and future work.

## 2 Secure Tropos Notation

Secure Tropos is a requirements engineering methodology aimed at fully capturing the properties of software systems and the organizational environment, focusing on modelling security [13]. The language extends the concepts of (social) actor, goal, task, resource and social dependency from the i\* modelling language



and redefining existing concepts introduced in the Tropos language and development process [14]. The Secure Tropos methodology closely follows the software development life-cycle with emphasis on security and privacy requirements, allowing the developer to incrementally create and refine models of the system-to-be during the analysis and design stage.

The Secure Tropos notation is fully defined in [13]. Here we present an outline of the subset of the notation used in this paper. The concrete notation is presented within *views*, where each view denotes a specific phase of activity in the modelling process. We now discuss Secure Tropos Views.

## 2.1 Organisational View

The diagram in Figure 1 illustrates the main nodes of an organisational view of Secure Tropos. It depicts a node-link diagram enclosed in a bounding rectangle. The nodes in the node-link diagram vary in shape according to the type of Secure Tropos element that they depict. The links similarly vary.

The circular node depicts an *actor*. An actor can be a physical or abstract manifestation, with strategic goals and intentions. An example actor labelled “Lecturer” can be seen in Figure 1.

The semi-oval node depicts a *goal*. Goals represent an actors strategic interests, which can be decomposed into sub-goals and combined using Boolean operations. An example goal labelled “Get student academic achievements” can be seen in Figure 1. Goals are linked through a *Dependency* link, depicted by one semi-circles on each side of the goal element.

A *Dependency* link indicates that an actor depends on another actor in order to achieve some goal/plan or to obtain a resource, where the direction the semi-circle is pointing towards denotes the dependee. An example dependency link can be found linking the goal “Get student academic achievements” with the actor “University of Brighton” who depends on the actor “Lecturer” to achieve the goal.

*Security Constraints* are depicted by the octagon node. Security constraints define security requirements through a set of restrictions that limit the way goals can be carried out. An example of a security constraint “Keep account access secure” can be found from the actor “Lecturer” to the goal “Access student records”.

## 2.2 Security Requirements View

The diagram in Figure 2 illustrates the security requirements view, which provides a detailed analysis of the organisational view. This view depicts a node-link

diagram enclosed in a bounding circle, defined by an actor that is delegated as the solution “system”. Several new elements are introduced in this view.

The elongated hexagon node depicts a *Plan*. A plan specifies the details and conditions under which a goal or measure is operationalised. “Lecturer fill in form” is an example of a plan.

The rectangle node depicts a *Resource*. Resources represent a physical or virtual entity. Resources can be linked to goals using a *Requires* link. An example of a resource is “Lecturer Notes” which is linked to the goal “Get academic achievements” via a requires link. The requires link indicates that the goal requires certain resources in order to be satisfied.

The pentagon node depicts a *Threat*. A threat indicates the potential loss or problems that can put the system at risk. The “Man-in-the-middle” is an example of a threat, which is linked via the *Impacts* link to the goal “Get student details”.

The *Impacts* link indicates the presence of a threat targeting a goal.

The hexagon node depicts a *Security Objective*. An example of a security objective is “Ensure data is kept private”, which is linked to the security constraint “Keep personal details private” and “Keep student records private” via *Satisfies* links and the *Security Mechanism* “Secure connection” via the *Implements* link.

The *Satisfies* link indicates that the security objective satisfies the given security constraint.

The hexagon node with two parallel horizontal lines depicts a *Security Mechanism*. A security mechanism is a method or procedure that enforces security objectives. “Secure connection” is an example of a security mechanism, which is linked via the implements link to the security objective “Ensure data is kept private”.

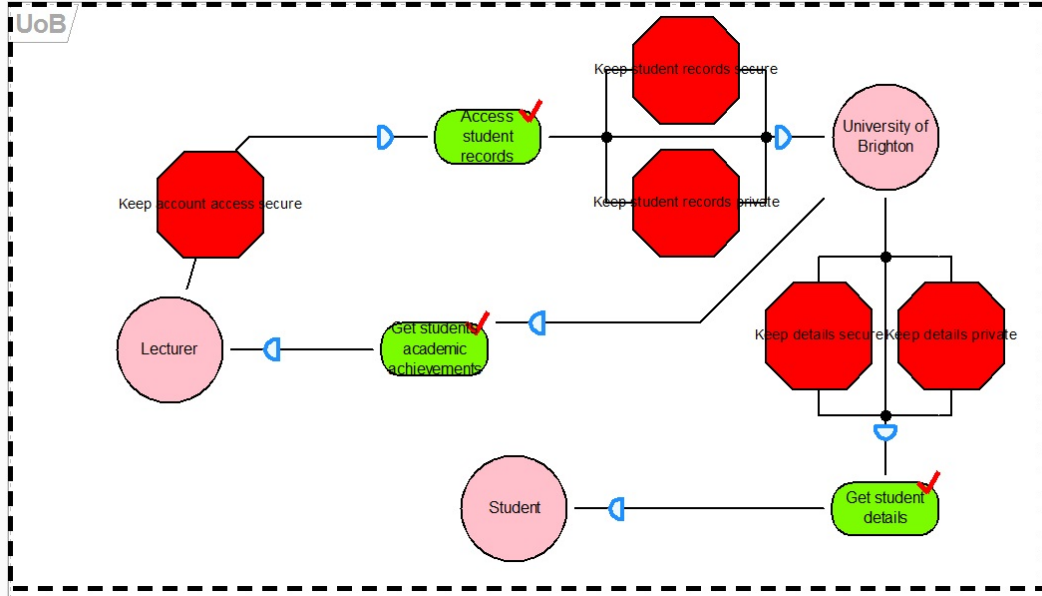
The example organisational and security views demonstrate the syntactic richness of Secure Tropos. Of which, we have provided outline explanations of a subset. We will now use this subset to identify services in Secure Tropos models.

## 3 Service

Before we can begin the process of determining what properties and aspects to capture when modelling cloud services, there is a need to obtain a concrete definition on what a service is. The common definition of a service can be given as “The performance of work (a function) by one for another”. However, service, as the term is generally understood, also combines the following related ideas [15]:

- The capability to perform work for another





**Figure 1. An example of an organisational view in Secure Tropos**

- The specification of the work offered for another
- The offer to perform work for another

Our interpretation for a computing service is based on definitions provided by IT standards bodies, specifically from the “Organization for the Advancement of Structured Information Standards” (OASIS). OASIS defines a service as “A mechanism to enable access to one or more capabilities, where the access is provided using a prescribed interface and is exercised consistent with constraints and policies as specified by the service description” [15].

**Capability** The capability of a service represents the ability to do something. Capabilities are a way to meet the needs of an agent by fulfilling their requirements. A capability has no function on its own, therefore service functionalities in the service description link together with the capabilities in order to fulfill a purpose.

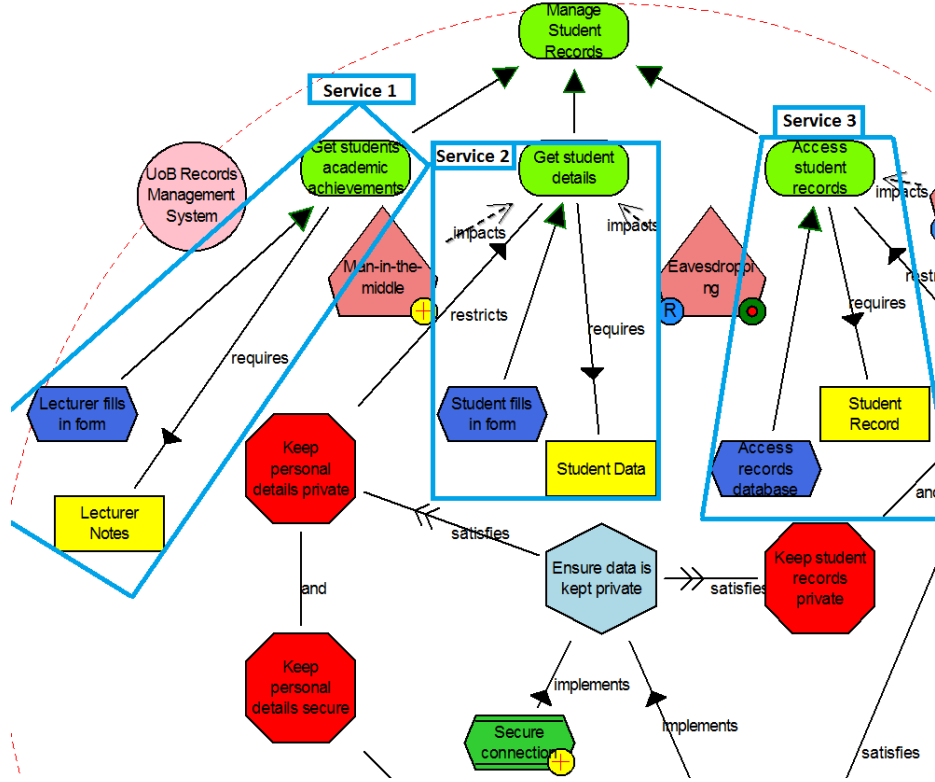
**Interface** The OASIS definition states that interfaces provide access to the capabilities of a service. The interface describes the means of interacting with a service and the actions involved in using a capability. These actions are initiated through specific protocols, commands, and information exchange specified by service functionalities in the service description. This is essentially the information that needs to be provided to the service in order to access its capabilities and receive a result.

**Service Description** In order to ensure that services are published and visually available for access, the interaction with services are described in the service description in terms of inputs, outputs and associated semantics.

### 3.1 Service Identification

Based on the OASIS definition, we propose that in the context of our work a service can be identified and modelled through a combination of a goal, plans and resources. **Goals** provide a description of the main underlying capability of the service, which relies on the functionality provided by the plan and the attributes defined in the resource to give the service description. Each **plan** describes a functionality of the service, which is required to give purpose to the service capability through the interface. The **resources** provide a specification of the information required by the interface in order to access the capabilities of the service.

This pattern can be identified in Secure Tropos by indicating a goal, searching for dependants of the goal and including any plans and resources linked to the goal. The most basic form of the pattern includes a goal, plan and resource. As goals increase in complexity, additional plans and resources may be added in order to specify additional functionalities. A simple example denoting the identification of a service based on the goal “Get student details”, plan “Student fills in form” and resource “Student Data” is shown in Figure



**Figure 2. An example of a security requirements view in Secure Tropos.**

2. We explain the key characteristics required to identify a service in the following sub-sections, justifying the goal, plan, resource pattern.

### 3.1.1 Capability and Functionality

Goals represent an actors strategic interest, which can be defined as the user requirements. In terms of the service definition, goals provides a description of the main underlying capability of the service. In the example shown in Figure 2, the goal “Get student details” describes the need to get student details. However on its own, this underlying capacity does not describe how this is fulfilled, nor does it specify the inputs or outputs when accessing the interface. The functionality is described by the plan, which is a description of what something does and how to achieve that goal. The plan provides parameters for a particular function in terms of behaviour and purpose. In essence, the capability of the service is described by the goal and the functionality is described by the plan. Thus by linking a function to a capability to describe how and what to do, we come closer to the definition of a service.

### 3.1.2 Interface

The plan describes the actions required for interacting with the service while the resource describes the information required for the interface to access the service capabilities. For example in Figure 2, the plan “Student fills in form” indicates that the interface has to have a functionality that captures the action of students filling in forms. The resource “Student Data” is also required in order to specify and store the data that was captured by the functionality of the service defined through the plan.

### 3.1.3 Service Description

Goals provide a conceptual part of the service description by describing what the service is supposed to accomplish and the conditions for using the service through capabilities. The service description also relies on the plan to define the semantics of interaction with the service, in addition to the resource in order to specify the attributes of the input and outputs. For example in Figure 2, the goal “Get student details” indicates that the service description will publish the fact that this service will get student details. The plan “Student fills in form” tells us the semantics of the interaction

and the resource “Student data” describes the inputs and outputs of the service in the service description.

Having established a visual Goal-Plan-Resource pattern in the Secure Tropos diagram, we now proceed to incorporate cloud requirements into Secure Tropos.

## 4 Incorporating in Secure Tropos

We validate our proposed work through constructed examples based on existing systems. In our case study, we create a scenario based on migrating an university records management system to the cloud.

### 4.1 Extensions to the Meta-Model

We extend the Secure Tropos meta-model to include additional attributes to model cloud requirements. Figure 3 shows a portion of the meta-model illustrating our extensions.

**Resource** The resource object is extended to include fields specifying the category of the resource, type, specifications, region, owner and classification. These notions are based on requirement-level properties defined in both CloudML [16], a domain-specific modelling language that specifies the provisioning, deployment, and adaptation concerns of cloud systems at design-time [5] and Cloud Computing Ontology (CoCoOn) [17], an ontology-based system for describing cloud infrastructure. As an example consider the region property from CloudML and CoCoOn, which describes the geographical location of the hardware component and is used to determine security/privacy jurisdiction and legislation. An example of a property from CloudML that is not requirement-level is *PrivateIP*, as this is implementation-specific. The extended resource will be deployed in the physical layer when modelling cloud services, providing a fine-grained view of the infrastructure required to enact services defined in the software system layer. This also provides the foundation for performing security analysis on cloud-specific threats and vulnerabilities and the modelling of data-flow. **Service Definition** Goal, plan and resource are existing concepts in the Secure Tropos meta-model. We propose the identification of a service based on the Goal-Plan-Resource pattern and extend the meta-model as shown in Figure 3, where we indicate that a service is an aggregation of a single goal, one or many plans and one or many resources.

The migration link is indicated by the shaded box in Figure 3, which links one service to one or many cloud actors.

### 4.2 Organisational View

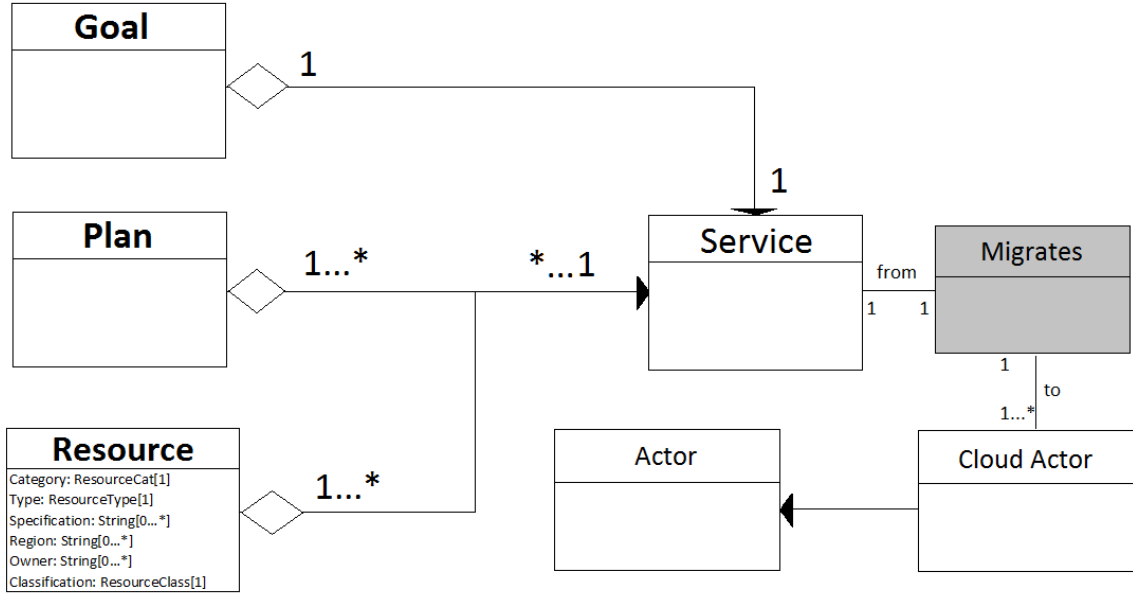
An example demonstrating the organisational view is illustrated in Figure 1, which shows the dependency relational link between different actors and the security constraints imposed based on goals. The security constraints are identified from the perspective of the dependent and dependee with regards to the goal. In this example, the *Lecturer* actor depends on the *University of Brighton* actor to satisfy the *Access student records* goal. The Lecturer has a security constraint that they should keep their account access secure while the *University of Brighton* actor has the security constraints of both keeping student records private AND secure.

### 4.3 Security Requirements View

The security requirements view in Figure 2 shows a wide range of elements that can be modelled in order to analyse the security requirements of a software system. The primary goal “Manage Student Records” has three sub-goals, in this example the “Get Student Details” sub-goal is examined in more detail. This sub-goal requires the resource “Student Data” and the plan “Student Fills in Form” in order to satisfy its requirements. It also has the security constraints of keeping personal details private and secure and is impacted by two threats; “Man-in-the-Middle” and “Eavesdropping”. Each of the sub-goals also define a service with its corresponding plan and resource links, as indicated by the bounding box and service labelling.

### 4.4 Defining and Migrating Services to the Cloud

Our proposed process allows the developer to identify and indicate a set of components that conceptually contribute towards a service. The constructed group is then linked via the migration relationship to a cloud service defined in another view, the cloud service view. In Figure 2, we have indicated that the goal “Get Student Details”, the plan “Student Fills in Form” and the resource “Student Data” contribute towards the definition of a service. Based on the goal description, the functionality of the service is to obtain details from students. The plan indicates that the service will include the capacity to obtain student data from forms that are filled in by the student, possibly through an form defined by the interface. The required input will be student data which the resource describes in full detail, including properties such the owner of the data, how the data is stored and the specifications of the data.



**Figure 3. Extensions to the meta-model to identify services from goal, plans and resources, define the migration link from service to a cloud actor and add attributes to resource.**

In order to model and analyse software systems for cloud environments, we need to create a model for describing cloud services and the components involved in the definition of these services. This would include the software applications deployed to address the problem or tasks that the service is trying to solve or achieve, specifications of the resources required to execute the software and identifying the data that will be processed to determine the flow of data. The resources can then be categorised further into network, compute and storage requirements, depending on the capabilities the service requires.

There are several possible scenarios where migrating to the cloud is applicable. In order to model cloud services, we propose a two-layer model; the software system and the physical layer. The **software system** describes the programmatic implementations of the functions offered by the services. The software system contains applications and services which provide solutions to the client requirements. Descriptions in each of the components within the software layer define a dependency link with required resources, which are provided in the physical layer. The **physical layer** contains resource elements that describe the storage, compute and network components required to satisfy the service requirements (Goal) based on the requirements of the stakeholder. Each one of the components in the physical layer describes an aspect of the infras-

tructure required to define a cloud computing system.

Each cloud service will include deployment models, service models and specifications for resources, based upon the user requirements and restrictions identified in the early stages of the requirements modelling. This process allows us to define the exact requirements when planning for resource provision, utilisation and optimisation.

## 5 Conclusion

We have discussed the current progress and challenges for modelling secure software system in this paper, emphasising the need for a modelling language that is able to holistically capture properties that define a cloud computing system based on client requirements. To address this gap, we define the properties and attributes required to model software services and cloud services. We define a pattern to group interdependent properties for the migration towards cloud services in a cloud computing environment, based on the strategic interests of stakeholders. These properties consist of the primary goal which defines what the main functionality of the service should be, plans that tell us what the capabilities of the service should be in order to fulfil the functionalities defined in the goal and the resources that are required by the service in order

to perform its functions. We validate our proposed research through extensions to the security requirements modelling language Secure Tropos and provide a case study based on modelling the migration of an university records management system to a cloud computing environment.

In order to obtain a holistic view of security vulnerabilities and threats, we will build on the cloud service view to examine each component within the cloud service based on different cloud models. We can further extend the security attacks view to include cloud specific security vulnerabilities, threats and mechanisms to mitigate these attacks. The nature and approach of attacks changes dynamically according to a wide variety of parameters, such as the service model, deployment model and in scenarios involving deployment of services to multiple clouds or service providers.

## References

- [1] T. H. Depot, "The home depot reports findings in payment data breach investigation," 2014.
- [2] A. Pavel, "Amazon.com server said to have been used in sony attack," May 2011.
- [3] Cloud Security Alliance, "Security research alliance to promote network security," *Network Security*, vol. 1999, no. 2, pp. 3–4, 1999.
- [4] A. Bergmayr, H. Brunelière, J. L. C. Izquierdo, J. Gorroñogoitia, G. Kousiouris, D. Kyriazis, P. Langer, A. Menychtas, L. Orue-Echevarria, C. Pezuela, and M. Wimmer, "Migrating legacy software to the cloud with ARTIST," *Proc. European Conference on Software Maintenance and Reengineering, CSMR*, pp. 465–468, 2013.
- [5] N. Ferry, A. Rossini, F. Chauvel, B. Morin, and A. Solberg, "Towards model-driven provisioning, deployment, monitoring, and adaptation of multi-cloud systems," *Proc. IEEE Sixth International Conference on Cloud Computing*, pp. 887–894, 2013.
- [6] S. Frey and W. Hasselbring, "The cloudmig approach: Model-based migration of software systems to cloud-optimized applications," *International Journal on Advances in Software*, vol. 4, no. 3 and 4, pp. 342–353, 2011.
- [7] M. Armbrust, O. Fox, R. Griffith, A. D. Joseph, Y. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, *et al.*, "Above the clouds: A Berkeley view of cloud computing," *University of California, Berkeley, Tech. Rep. UCB*, pp. 07–013, 2009.
- [8] N. R. Jennings, "Agent-oriented software engineering," in *Multiple Approaches to Intelligent Systems*, pp. 4–10, Springer, 1999.
- [9] E. Kamateri, N. Loutas, D. Zeginis, J. Ahtes, F. D'Andria, S. Bocconi, P. Gouvas, G. Ledakis, F. Ravagli, O. Lobunets, and K. a. Tarabanis, "Cloud4SOA: A semantic-interoperability paas solution for multi-cloud platform management and portability," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8135 LNCS, pp. 64–78, 2013.
- [10] R. G. Cascella, C. Morin, P. Harsh, and Y. Jegou, "Contrail: A reliable and trustworthy cloud platform," in *Proc. 1st European Workshop on Dependable Cloud Computing*, p. 6, ACM, 2012.
- [11] W. Madsen, *Trust in Cyberspace*, vol. 1999. 1999.
- [12] P. Devanbu, S. Stubblebine, and S. S. Premkumar T. Devanbu, "Software engineering for security - a roadmap," *Icse*, pp. 227–239, 2000.
- [13] H. Mouratidis and P. Giorgini, "Secure Tropos: A Security-Oriented Extension of the Tropos methodology," *International Journal of Software Engineering and Knowledge Engineering*, vol. 17, pp. 285–309, Apr. 2007.
- [14] A. Bandara, H. Shinpei, J. Jurjens, H. Kaiya, A. Kubo, R. Laney, H. Mouratidis, A. Nhlabatsi, B. Nuseibeh, Y. Tahara, T. Tun, H. Washizaki, N. Yoshioka, and Y. Yu, "Security patterns: comparing modeling approaches," 2010.
- [15] C. M. MacKenzie, K. Laskey, F. McCabe, P. Brown, and R. Metz, "Reference Model for Service Oriented Architecture," *Oasis*, 2006.
- [16] E. Brandtzæg, S. Mosser, and P. Mohagheghi, "Towards cloudml, a model-based approach to provision resources in the clouds," in *8th European Conference on Modelling Foundations and Applications (ECMFA)*, pp. 18–27, 2012.
- [17] M. Zhang, R. Ranjan, A. Haller, D. Georgakopoulos, M. Menzel, and S. Nepal, "An ontology-based system for cloud infrastructure services' discovery," in *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2012 8th International Conference on*, pp. 524–530, IEEE, 2012.

# Pupils' Collaboration around a Large Display

R. Lanzilotti\*, C. Ardito\*, M.F. Costabile\*, A. De Angeli<sup>°</sup>, G. Desolda\*

\* Dipartimento di Informatica, Università di Bari Aldo Moro, Italy

<sup>°</sup> Dipartimento Ingegneria e Scienza dell'Informazione, Università di Trento, Italy

{name.surname}@uniba.it, °antonella.deangeli@disi.unitn.it

**Abstract**— Collaboration is acknowledged as a key element of learning. Thus, it is valuable to develop Information and Communication Technology applications that, being implemented on proper devices, can support collaborative learning. Large multi-touch displays appear to encourage collaboration by offering users a shared environment to act upon. However, little knowledge is available on the actual influence of this technology on human behavior and more empirical evidence is needed to better understand its capability to foster collaboration. This paper provides a contribution in this direction by presenting a field study that helps understand collaborative practices of pupils playing an educational game that runs on a multi-touch large display. This study involved 98 fifth-graders at a primary school (average age 10 years old). Results confirmed the potential of large displays as useful devices for collaborative learning.

**Keywords**—Technology-enhanced learning; large multi-touch display; educational game; field study.

## I. INTRODUCTION

In the last few years, several publications reported and discussed studies investigating how Information and Communication Technology (ICT) can support children learning (e.g. [1-3]). Along this direction, we focused on primary school education in ancient history, especially because Italy has a rich source of cultural heritage, with historical sites dating back to centuries B.C. Our approach extends formal learning activities (classroom activities), with field visits and games to be played in situ [4] or in the school [5]. We focused on collaborative games, which are considered a valuable form of informal and collaborative learning, able to capture pupils' attention and engage them (e.g., [1]). Numerous benefits have been identified for collaborative learning. In particular, it helps developing a social support for learners and creating learning communities (*social benefits*); it generates positive attitude towards teachers and reduce anxiety (*psychological benefits*); it actively involves students in the learning process and improves classroom results (*academic benefits*) [6]. Specifically, collaborative games require different skills at the same time and each player can practice those ones felt to be the most congenial. By acting together, pupils can solve problems, overcoming difficulties thanks to their joint efforts. In addition, games create a pleasant learning environment, in which pupils learn with fun. Innovative technologies, such as mobile devices, interactive large displays and multimodal interfaces, can make the environment more engaging [1]. For example, the excursion-game technique discussed in [4] uses mobile

devices and permits the development of educational pervasive games to be played by groups of young visitors, exploring sites of cultural interest, such as archaeological parks.

Literature reports examples of applications of large displays that foster synchronous and co-located collaboration [7]. Their size has the potential to accommodate more people in front or around the display. Thus, they can stimulate simultaneous participation in learning activities, promote engagement, allow to share control and responsibility over the manipulation of learning material [8]. However, more empirical evidence is needed to better understand the collaborative practices of learners around a large display and how such practices are organised and managed by learners. This paper gives a contribution to this issue by presenting a field study, whose goal was to investigate how pupils behave around multi-touch large displays. Small groups of pupils were observed while playing an educational game on a multi-touch vertical display installed in their school laboratory. The study involved a total of 98 pupils of six different fifth-grade classes at a primary school. We also compare and contrast our results with related work, analyzing how display features may affect pupils' collaboration.

The paper is organized as it follows. Section II discusses related work. The motivation for developing the educational game is presented in Section III. Section IV describes the study and discusses its results. Finally, Section V concludes the paper.

## II. DISPLAY FEATURES AFFECTING COLLABORATION

Large interactive displays are increasingly used; a comprehensive review is reported in [7]. By analyzing the literature, features of such displays that affect people collaboration are briefly discussed in this section.

### A. Simultaneous use actions

Over a decade ago, Scott and his colleagues proposed a set of guidelines for co-located collaboration around a tabletop display [9]. A guideline that clearly emerges from users' feedback and is relevant for any type of large display is "support simultaneous use actions". This is a primary requirement, since people collaborating in a task want to interact simultaneously with artifacts on the display. It affects both hardware and software. The hardware must provide either multiple input devices or touch screens capable to detect multiple touches occurring at the same time. The software must support the interactions with



multiple components at the same time. A notable example is *DiamondTouch*, a tabletop display that, beside allowing multiple concurrent users, is capable of identifying the touches of a specific user [10].

Another system that permits multi-user simultaneous interaction is *CityWall*: it detects more finger touches at the same time as well as hand touches. For example, in order to rotate a picture shown on the display, the user puts his hand on top of the picture and rotates his hand. The large display of *CityWall* allows people to browse photos and videos downloaded from social networks like YouTube and Flickr [11]. A study was performed with the system installed in a shop window next to a café in the center of Helsinki, Finland [12]. Data on 1199 people were collected. They showed that the large display enabled collaborative activities of different groups, who used the system in parallel, possibly for different tasks. In several cases, groups of strangers ended up socializing and having fun together, even if they started interacting separately. Several other studies performed in the last years show that the multi-user interaction possibilities offered by large display allow different people to work in parallel [7].

### B. Display set-up

A feature that has a considerable influence on people collaboration is the display set-up, which refers to the physical installation of the display, characterized by size, orientation and shape [7]. Vertical flat displays are the most used so far. The horizontal orientation, i.e., a tabletop display, is frequent in locations like offices and museums, since users can interact for a long time while comfortably sitting around the display. Thus, it appears especially appropriate for collaborative tasks [13]. However, it also presents some inconveniences, because two users sitting on the opposite sides of a tabletop display see contents reversed with respect to each other. An interesting proposal for overcoming this problem has been recently presented in [14]: by wearing active shutter glasses, each user can see her/his private view, while the shared view is visible to everybody. Thus, users can collaborate in the manipulation of the shared information. Rogers and Lindley studied collaboration around vertical and horizontal large interactive displays [15]. They found that horizontal surfaces better support collaborative activities that closely couple the resources used and/or created during the various activities. On the other hand, vertical displays are better at providing a shared surface that allows a group of people to view and annotate information to be talked about and referred to. The vertical display gives all viewers the same perspective of the task and provides a holistic view of the data. A more recent study challenged this view [16]: by analyzing the users' performance and their satisfaction with vertical and horizontal set-ups, the authors reported a strong preference for the horizontal set-up (probably due to decreased fatigue). People who preferred the vertical surface appreciated that it provides a better overview and that hands less likely occlude objects on the screen.

Other studies experimented displays with diagonal

orientation, showing that they improve the interaction of the user(s) placed at the lowest side, since the display provides a better viewing angle, but greatly limits the collaboration with users standing at the other sides, even if they have a common task to perform [13] [17].

Systems with spherical or cylindrical displays have also been proposed; they were compared in [18]. With both spherical and cylindrical displays, each user can see at most one half of the display. Thus, these displays, rather than fostering collaboration, allow different users to interact without disturbing each other and preserving their privacy.

### C. Application purpose

The effect on collaboration determined by the purpose of the application running on a large display, i.e. related to the tasks users perform, is evident in the field study reported in [19]; it investigated the use of *Tourist Planner*, an application on a tabletop display, installed in the tourist information center in Cambridge (UK), which supported groups of people planning their trips. The study showed that the system acted as an aggregator for a group of people coming together in the information center. They interacted together with *Tourist Planner*, collaborating to define their itinerary, rather than being dispersed in the environment, as usually occurred.

Some applications were designed to foster classroom collaborative activities through multi-touch tables. Piper and Hollan, for example, compared the use of the digital material through the tabletop with the use of paper material, in order to investigate how the study practices are influenced, including student participation and cooperation, in a neuroscience class of 20 undergraduates at the University of California [20]. Results revealed that the large, shared nature of the tabletop display allows student to have equal access to material and engage in parallel activities. Moreover, the use of the digital material on the tabletop provided greater playfulness and enjoyment were noticed.

*TablePortal* is a system that allowed teachers to manage and monitor collaborative learning of students working with multi-touch tables [21]. The teacher used a separate table to communicate with the students' tables; in this way, teachers and students could work together on their multi-touch tables and collaborate on learning tasks. Observations in a real context showed an enhanced level of teacher's awareness, flexible monitoring, and a positive impact on social interactions in the classroom.

Another example of successfully use of multi-touch technology for learning is reported in [22]. Comparing pair-programming learning through multi-touch technology with learning using a desktop, this study revealed that students performed better working at the multi-touch table, because it encouraged collaboration and helped people expressing their potential.

### D. Location

Nowadays, large displays are installed in different

locations, either public (a city street, a museum, a university campus, etc.) or semi-public (an office, a school, etc.). The location strongly affects the behavior of people in the environment around the display. In semi-public locations, most people know each other and could thus be less inhibited to work together [23]. If the display is installed in a location where people do not know each other, an initial time span is observed in which people are a bit reluctant to approach the display [24], but, once they start interacting, in most cases they soon socialize and collaborate with each other.

### III. THE EDUCATIONAL GAME

The value of games for learning purposes is predicted by several pedagogical theories and confirmed by some studies showing how ICT can be used to engage learners with e-games (e.g., [2]). In her recent book, Oviatt reports an in-depth review of educational technology [1]. She provides evidence that interactive games lead to learning improvements of about 10–40% when compared with traditional lessons.

In our research, we built upon the Discovery Learning technique, defined by Bruner in his Constructivism theory [25], to propose an educational format that integrates formal learning (classroom lessons) with informal and technology-based learning [5]. The educational format organizes learning activities in three phases, in which pupils acquire new information: 1) attending the lesson(s) by their teacher in the classroom (*symbolic phase*), 2) acting in a real context (*active phase*); and 3) interacting with technological tools to manipulate visual images or tactile objects (*iconic phase*).

The format was experimented with pupils aged 10 years old. The goal was to foster history learning, stimulating pupils' interest in cultural heritage. The three learning activities were the following: 1) pupils attended a classroom lesson, in which the teacher provided basic notions on the history of Egnathia, an ancient Roman city in Southern Italy (symbolic phase); 2) pupils visited the archaeological park of Egnathia, in which they observed its ancient monuments and objects (active phase); 3) pupils played an educational game, called *History Puzzle*, implemented on a large multi-touch display available in the school laboratory (iconic phase).

In collaboration with teachers and pedagogues, we decided to adopt a collaborative learning approach, whose benefits are discussed by many researchers (e.g., see [6]). Specifically, collaborative learning refers to an educational method that involves groups of individuals, acting together in small groups to achieve a common learning goal. Consequently, we implemented *History-Puzzle* on a large interactive display, in order to allow small groups of pupils to play together. According to the simultaneous use actions requirement, our system recognizes the gestures performed simultaneously by the hands of multiple pupils.

The game proposes puzzles that pupils have to solve in order to discover monuments and/or other objects, which

were presented by the teacher during the lesson and that pupils had seen during a visit to the archaeological park of Egnathia. The puzzles show the 3D reconstruction of the points of interest in Egnathia, allowing pupils to appreciate the original look-and-feel of archaeological ruins. The figure to be discovered by solving the puzzle is shown at the center of the screen. The nine square tiles of the puzzle contain incomplete sentences reporting historical notions about the selected place. For each tile of the puzzle, the players choose the tile with the rest of the sentence among those displayed on the left and right sides of the puzzle, and drag it onto the puzzle tile. In the example of Figure 1, which refers to the puzzle of the “Basilica Episcopale” (“Civil Basilica”), the tile with the sentence “Era un edificio con ...” (“It was a building composed of...”) is associated to the tile “... tre navate” (“three naves”), located at the top right corner of the display, which correctly completes the sentence. If the selected association is the right one, the tile reveals one ninth of the 3D reconstruction of the original place.

A score of 5 points is awarded if the tiles are correctly associated, while the current score is reduced by 2 points every time the pupils move a tile onto the wrong one. This score mechanism stimulates pupils to reflect upon their actions and leads them to discuss together the tiles they have to associate. The current score is permanently displayed at the bottom of the screen, just under the puzzle (e.g., “Punteggio: 30” in Figure 1). In order to make the game more challenging, some tiles report false answers or answers that do not match any of the nine incomplete sentences in the puzzle. When the puzzle is completed, a new screen proposes various multimedia contents related to the building, such as sounds, videos, images and texts. Then, the system returns to a screen showing the map of the park, where participants can choose the next puzzle. Once the game is over, the final score is displayed. *History-Puzzle* was deployed on a MultiTouch Ltd 46-inch large Full HD LCD display, with vertical orientation.



Figure 1. The “Basilica Episcopale” puzzle.

### IV. FIELD STUDY

The data presented in this paper were collected as part of a wider study, conducted from November 2011 to January 2012, whose main goal was to evaluate the learning effectiveness of the educational format described at the

beginning of the previous section. As reported in [5], the study showed that pupils were actively engaged in all educational activities and that the game was a valid means for consolidating the acquired knowledge. In this paper we concentrate specifically on the iconic phase, in which groups of pupils interacted with History-Puzzle. This analysis had another specific goal, namely to investigate how pupils behave around the multi-touch large display. This analysis was not reported in [5] and it is described in this paper. From now on, the study we refer to is about pupils playing with History-Puzzle, addressing the latter specific goal.

#### A. Participants and procedure

The study involved six classes of fifth-graders at the primary school “Clementina Perone” in Bari, Italy. A total of 98 pupils (50 girls, average age 10 years old) participated in the study, as part of their school activities. The overall study described in [5] involved a total of 107 pupils, but 9 of them did not participate in the iconic phase, since they were not at school when this phase was performed. The participants were divided into 22 groups (12 groups of four and 10 of five pupils). The groups were decided by the schoolteachers, who also aimed to guarantee homogeneity in terms of the pupils’ cognitive and social development. Parental consensus was obtained prior to the study.

On November 24th and 25th, the groups took turns to go to the school laboratory to play History-Puzzle, 10 groups the first day, the remaining 12 groups the day after. Three researchers were involved in the study. One of them interacted with the pupils, explaining how to interact by hand gestures and the task objectives. The other two observed and provided technical support if needed. Each group had to solve three puzzles. The interaction with the multi-touch display was videotaped by two cameras. Camera 1 was installed on a tripod about two meters away from the display to film the pupils’ behaviour. Camera 2 was placed on top of the display to film the pupils’ faces while they interacted with the system. Pupils’ comments and utterances were captured by an audio recorder next to Camera 2. Moreover, two research assistants noted down the main events and provided help when explicitly requested or when pupils were not able to continue playing.

#### B. Data analysis

The collected data included videotaping of the groups’ interactions and notes from observation in the laboratory. In order to better analyze pupils’ behavior, three researchers transcribed the videos, literally noting down all intelligible speech and details of all instances of their interaction with the multi-touch display. Moreover, contextual information was coded: for each group member the level of involvement in the game and his/her position in front of the display were considered. The transcripts were analyzed by a thematic analysis following a semantic approach: themes were identified within the explicit or surface meaning of the data, based on what participants said or did during the game [26]. Each researcher independently produced the transcripts, and

60% of the results were double-checked for reliability, leading to an initial value of 85% for all measures reported in this article. Discrepancies were solved by discussion.

The analysis highlighted two important themes related to group strategy and anti-social behavior. Group strategy addresses the pupils’ behavior while playing together as a group, showing common patterns and reflecting on their causes and consequences. The anti-social behavior theme provides an overview of the cases when pupils did not directly contribute to the game, either because they chose not to or because they were hampered by the others. This theme also investigates the causes of conflicts during playing or deliberate attempts to disrupt the game.

#### C. Results

Three researchers analyzed a total of 5 hours 46 minutes of videos recorded during the study. This included the arrival of the groups, game presentation, time each group interacted with the display to play the game, i.e., to solve the puzzles, look at the multimedia shown at the end of each puzzle and see their final score. In particular, the 22 groups took 3 hours 58 minutes to reassemble 65 puzzles (3 puzzles per 21 groups, plus 2 puzzles for a group who experienced technological problems). On average, a group spent 4,15 minutes to reassemble the first puzzle, 3,37 minutes to reassemble the second puzzle, 2,6 minutes for the third one.

##### *Group strategy*

The game was composed of two consecutive activities: *tile association* and *tile positioning*. *Tile association* was a problem-solving activity: pupils looked at the display discussing within the group the solution of the riddle, which would identify the correct association between tiles. *Tile positioning* was the physical action performed to overlap the tiles. *Tile association* required reading information from the display and identifying the correct answer; *positioning* required pointing and dragging.

Generally, the groups were well organized in performing the game. They mainly applied the following strategy. For each individual tile, they first read the riddle and discussed the answers. *Positioning* tended to occur only after reaching group agreement. Only 3 groups experienced some problems when solving the first puzzle. They appeared disorganized and unable to choose how to analyze tiles, which tiles to associate and who should move them. At the beginning, these pupils mainly played individually, reading and moving different tiles of the puzzle with little or no group interaction. Yet this state of affairs tended to disappear quickly and, afterwards, these groups identified a better strategy for solving the remaining two puzzles.

*Tile association* was a clear collaborative activity, articulated as follows: 1) as soon as a pupil started to read the text in a puzzle tile, the whole group read it in unison; 2) a pupil proposed an external tile to be associated; 3) pupils discussed together if they confirmed the proposal, 4) the tile to be moved was chosen; otherwise, they went back to step 2). Excerpt 1 reports a discussion occurring in Group 20.

Each pupil is denoted by a capital letter.

#### EXCERPT 1 - Group 20

A reads the text in a puzzle tile: "The Amphitheatre was...". Then A reads the text of a tile to be associated "Rectangular".

C: "Do you know the Amphitheatre? It was elliptical! It was not rectangular, as you said! So the right tile is this! [indicating a specific tile]."

Generally, tile association was performed by the whole group without touching the display. If a pupil moved a tile without the permission of his/her classmates, he/she was scolded by the group, especially if the tile was wrong. Only in 6% of the cases (i.e. to solve 4 puzzles), a group split into two sub-groups, each of which took care of associating half of the tiles. Specifically, pupils divided the display space into two parts and each sub-group concentrated on the tiles displayed in front of them. The pupil behavior in each sub-group was the same. According to [27], this strategy is defined cooperative (pupils cooperate): group members split the work, solve sub-tasks either individually or in subgroups and then assemble the partial results into the final output. Instead, the strategy is collaborative (pupils collaborate) if group members do the work 'together'.

*Tile positioning* was performed by one pupil on behalf of the group. Two behaviors were observed: *spontaneous* and *organized*. The spontaneous behavior refers to the situation in which tile positioning was not defined a priori and a child naturally moved it. Thus, after tile association, either the pupil who identified the tile would move and overlap it on the puzzle tile or any pupil would move the identified tile because he/she was the closest to the display or the fastest to act (see Excerpt 2). This behavior was observed in solving 31% (i.e. 19 puzzles) of the puzzles. It is worth mentioning that this spontaneous behavior did not generate any confusion or disturbance in the group; instead, it was a sign that the group actually collaborated in harmony.

#### EXCERPT 2 - Group 7

The group discusses tile association, selecting the correct tile to be moved, as indicated by child C who reads the tile text: C: "Curbstone!"

D is the closest to the tile with text "Curbstone".

C tells D: "Give me "Curbstone"."

D takes the tile and moves it closer to C.

C completes the tile positioning.

Excerpt 3 shows a situation in which a child complained because he/she was not allowed to position a tile.

#### EXCERPT 3 - Group 15

The group reads a puzzle tile in unison.

A selects an external tile and reads its text.

B moves another tile and the association is not accepted.

C moves the same tile of B and again the association is not accepted.

D identifies another tile.

B, who is the closest to this tile, moves it

D exclaims: "Oh, nooo. I should move it!"

To avoid complaints, pupils turned to an explicitly organized behaviour, taking turns in positioning a selected tile. This often occurred around 60% (i.e. 37 puzzles) of the puzzles. Excerpt 4 illustrates a typical case.

#### EXCERPT 4 - Group 15

The group is discussing a tile association.

D autonomously positions a tile.

B to D: "Please do not do that again!"

A: "Let's take turns!"

D: "That's right!"

A: "OK. The order is: me, Francesco, Giovanna and Vito."

In 10% (i.e. 6 puzzles) of the puzzles, a different turn-taking was observed. Specifically, a pupil selected a puzzle tile to be associated. The discussion began and the group selected the external tile providing the correct association. The pupil moved the selected tile. Excerpt 5 illustrates this behavior.

#### EXCERPT 5 - Group 18

C: "Francesco [i.e. D] is the first: he must read and move the tile. Then, Giulia, ..."

D: "They could go in through two different entrances..."

A: "Rectangular!" [pointing to the tile]

C: "Nooo, Symmetrical!!!" [pointing to the tile]

B: "Yes, symmetrical!"

D moves the tile.

In some cases, pupils had to solve technological problems. The display used during the study is less sensitive at the borders and in a central strip about two centimeters wide. This makes the interaction with the displayed objects more difficult at these points. Once they understood how solving such problems, pupils collaborated to overcome them. Figure 2 illustrates an example in which two girls moved the tile together to overcome the less sensitive central strip.



Figure 2. Girls move together a tile to overcome a technological problem.

#### Anti-social behavior

Generally, all pupils participated in the game, collaborating with the others in the group. However, a careful analysis of group dynamics revealed some episodes when pupils appeared not to be collaborating for a short



time, but soon they were stimulated by the other group members to be more active. The following three different behaviors were identified: 1) *hindered*, i.e., a pupil tried to interact with the display, but the others hampered her/him; 1) *disturbing*, i.e. a pupil bothered the group or encumbered the game activities, especially because he/she wanted to be the only one to interact with the display; 3) *onlooking*, i.e. a pupil watched the others playing (discussing and interacting with the display) without contributing. Table I reports the number and the percentage of the episodes in which such behaviors were observed.

TABLE I. NUMBER AND PERCENTAGE OF EPISODES REVEALING NOT COLLABORATING BEHAVIORS.

<i>Pupil behavior</i>	<i>Episodes</i>	
	<i>N</i>	<i>%</i>
Hindered	5	19%
Disturbing	10	37%
Onlooking	12	44%
<b>Total</b>	<b>27</b>	<b>100%</b>

A hindered pupil was observed only in 19% of the not collaborating behavior episodes. Figure 3 shows a girl who would like to interact with the display, but her classmates do not allow her to reach a better position to touch the display. However, she was active in tile association providing suggestions for selecting the tile.



Figure 3. The girl at the back is hindered by her classmates.

A disturbing behavior emerged in 37% of the not collaborating episodes. The disturbing pupil tried to prevail over his/her classmates with gestures like blocking the others to prevent them from touching the display (Figure 4). However, he/she was always scolded by his/her classmates and went back to a more collaborative behavior.



Figure 4. The girl at the center pushes away her classmates in order to interact with the display.

Finally, an onlooking behavior was noticed in remaining 44% of the episodes. Typical behavior of the onlooking pupil is having his/her hands behind his/her back (as in Figure 5) or in his/her pockets. Soon classmates tried to stimulate the onlooking child to be more active. In all cases, he/she returned to actively participate in the game.

Other episodes of anti-social behaviors occurred when pupils quarreled during the game. We observed only 15 episodes of conflicts among pupils. Specifically, 10 conflicts were related to social interaction and arose when a pupil did not observe his/her turn in interacting with the display, for example, because he/she was closer to the tile to be moved than the pupil whose turn it was. However, the other group members scolded him/her and re-established the right order, as illustrated in Excerpt 7.

#### EXCERPT 7 - Group 23

B: "Please, we have to go slowly!"

C: "We have to take turns! Simona starts!"

E tries to move a tile.

C to A: "Oh, Diego it isn't your turn! You must not touch the display! Simona has to move the tile!"



Figure 5. An onlooker attitude of the boy with his hands behind his back.

The remaining 5 conflicts arose because of the pupils' position in front of the display (physical space). Such conflicts occurred more in groups of 5 (4 conflicts in 2 groups of 5, 1 conflict in 1 group of 4), also because the display was not large enough to comfortably accommodate all pupils in front of the display. Specifically, in the groups of 4, pupils were well positioned in front of the 46-inch display next to each other. Thus, they maintained their position since they could read the tiles and interact with all the objects shown on the display. Rarely, a pupil at the side of the display moved towards the center.

In groups of 5, it happened that a pupil was forced to stand behind the others or in a peripheral position, not convenient for interacting with the display, thus he/she tried to reach a better position by pushing the other pupils. In fact, it was observed that pupils being at the center were more active, since they could easily read the text in the tiles and reach them. Figure 6 shows a sequence of images in which a girl initially behind her classmates tries to acquire a better position to be able to interact with the display. However, pupils were able to manage such conflicts autonomously: the teacher intervened just twice to deal with physical space conflicts.

#### D. Discussion

The analysis of the data collected during the study confirmed what other authors report in the literature (e.g., [20-22]), namely that educational applications running on large multi-touch displays provide a shared experience for learners by facilitating social interaction and collaboration among them. In order to play the game, pupils worked together, solved problems emerging during the game, and exchanged information among themselves. Thus, with respect to the objective of our study, which intended to investigate about pupils' collaboration behavior, we can conclude that the educational game implemented on the large display fostered collaboration, particularly in the problem solving activities related to tile association.

The obtained results also permit further comparisons with related literature, by analyzing in more details which

display features primarily acted as collaboration promoters. Table II summarizes these findings. As discussed in Section II, support for simultaneous use actions is considered as a primary requirement for allowing people to collaborate in the interaction with the display. In our study, however, students spontaneously swapped to a sequential use pattern in the interaction with the technology, thus showing that the simultaneous use actions requirement was marginal. Several factors can explain this behavior, including the task at hand, but we believe that the technology by itself could have afforded this behavior [28].

TABLE II. DISPLAY FEATURES FAVOURING COLLABORATION.

<i>Display feature</i>	<i>Collaboration promoter</i>
Simultaneous use actions	Marginal at this stage
Display orientation	Yes
Display size	Yes
Application purpose	Yes
Location	No evidence

Pupils appeared to be very excited by the multi-touch technology and all of them were keen on moving the tiles. They spontaneously adopted a sequential approach only to avoid conflict and to equally interact with the display. The group collaborated in the identification of the right answer, and then a designated "user" mediated the interaction with the display. Another reason for the sequential tile positioning was related to group performance. If a tile was improperly moved, the group score decreased. Indeed, the group required the control of tile positioning. If a member moved a tile without the group consensus, the others members scolded him/her. Nevertheless, the possibility of performing simultaneous actions was exploited to overcome technological issues: children effectively collaborated to overcome a problem, as in the case of pupils moving the tile together in Figure 2.

Concerning the display set-up, orientation and size revealed as collaboration promoters (see Table II). The vertically positioned display favored the view of the visualized elements to the players, who were all able to read the tile contents and collaborate in tile association. The display used in the study was of 46-inch size; it worked very well for groups of 4 pupils, since they could stand in front of the display very comfortably. In groups of 5, pupils moved more frequently to obtain a better position to interact with the display. Indeed, more conflicts about physical space occurred in groups of five.

History-Puzzle was designed with the purpose of promoting pupils' collaboration. It succeeded in this, since the results showed that pupils strongly collaborated, primarily in tile association. Moreover, as reported in [5], teachers highlighted that the multi-touch display favored pupils' inclusion (i.e. the involvement in the school activities of all pupils, regardless of social, cultural and personal differences [29]). Teachers remarked that even those pupils, who in class seem disinterested and tend to distract, actively participated in the game, provided appropriate answers and collaborated in the group activities with enthusiasm.





Figure 6. A girl trying to acquire a better position to interact with the display.

Based on our study, there is not evidence that the location influenced pupils' collaboration, because no comparison between different locations was carried out. However, teachers commented that, when working in the laboratory with desktop computers, usually pupils work individually or, sometimes, in pairs [5]. Once the large display was available in the laboratory, they enjoyed working in larger groups and collaborating with their peers.

## V. CONCLUSIONS

This paper has presented the results of a field study whose aim was to investigate collaborative practices of pupils playing an educational game implemented on a multi-touch large display. The study confirmed that such displays encourage collaborative activities. It also showed that size and orientation of the display and purpose of the application running on it were the features that most affected pupils' collaboration.

The performed study was based on qualitative data. Some researchers claim that quantitative methods are better than qualitative ones, since the former provide objective measurements and enable replication of studies, while the latter build on subjective interpretation. This is not true if qualitative data are analysed with methods that ensure the necessary objectivity and soundness [30], as done in the presented study. It is actually suggested to first perform qualitative research when the objective is to explore a new area of interest and to possibly discover diversity and variety [31]. Once enough insight is gained, it is possible to frame the design and analysis of a quantitative study to provide better indications of the magnitude of the researched phenomenon. We are confident that the research presented in this paper will stimulate further work toward a deeper understanding of the influence of large display on collaboration activities.

## ACKNOWLEDGMENTS

This work is partially supported by the Italian Ministry of University and Research (MIUR) under grants PON 02\_00563\_3470993 "VINCENTE" and PON04a2\_B "EDOC@WORK3.0", and by the Italian Ministry of Economic Development (MISE) under grant PON Industria 2015 MI01\_00294 "LOGIN".

## REFERENCES

- [1] S.L. Oviatt: "The future of educational interfaces" (Routledge Press, 2012).
- [2] D.W. Shaffer: "How computer games help children learn" (Macmillan, 2006).
- [3] P. Di Bitonto, T. Roselli, V. Rossano, E. Frezza and E. Piccinno, "An educational game to learn type 1 diabetes management", *Proc. DMS '12*, 2012, KSI Press, pp. 139-143.
- [4] C. Ardito, M.F. Costabile, A. De Angeli and R. Lanzilotti, "Enriching exploration of archaeological parks with mobile technology", *ACM Trans. Comput.-Hum. Interact.*, vol. 19, no. 4, 2012, pp. 1-30.
- [5] C. Ardito, R. Lanzilotti, M.F. Costabile and G. Desolda, "Integrating traditional learning and games on large displays: an experimental study", *Educational Technology & Society*, vol. 16, no. 1, 2013, pp. 44-56.
- [6] M. Laal and S.M. Ghodsi, "Benefits of collaborative learning", *Procedia - Social and Behavioral Sciences*, vol. 31, no. 0, 2012, pp. 486-490.
- [7] C. Ardito, P. Buono, M.F. Costabile and G. Desolda, "Interaction with large displays: a survey", *ACM Computing Survey*, vol. 47, no. 3, 2015, pp. 38.
- [8] I. Jamil, K. O'Hara, M. Perry, A. Karnik, M. Marshall, S. Jha, S. Gupta and S. Subramanian, "Dynamic Spatial Positioning: Physical Collaboration around Interactive Table by Children in India", *Human-Computer Interaction - INTERACT 2013*, LNCS 8120, 2013, Springer, pp. 141-158.
- [9] S.D. Scott, K.D. Grant and R.L. Mandryk, "System guidelines for co-located, collaborative work on a tabletop display", *Proc. ECSCW '03*, 2003, Kluwer Academic Publishers, pp. 159-178.
- [10] P. Dietz and D. Leigh, "DiamondTouch: a multi-user touch technology", *Proc. UIST '01*, 2001, ACM, pp. 219-226.
- [11] P. Peltonen, A. Salovaara, G. Jacucci, T. Ilmonen, C. Ardito, P. Saarikko and V. Batra, "Extending large-scale event participation with user-created mobile media on a public display", *Proc. MUM '07*, 2007, ACM, pp. 131-138.
- [12] P. Peltonen, E. Kurvinen, A. Salovaara, G. Jacucci, T. Ilmonen, J. Evans, A. Oulasvirta and P. Saarikko, "It's Mine, Don't Touch!: interactions at a large multi-touch display in a city centre", *Proc. CHI '08*, 2008, ACM, pp. 1285-1294.
- [13] C. Shen, K. Everitt and K. Ryall, "UbiTable: impromptu face-to-face collaboration on horizontal interactive surfaces", *Ubiquitous Computing - UbiComp 2003*, LNCS 2864, 2003, Springer, pp. 281-288.
- [14] R. Lissermann, J. Huber, M. Schmitz, J. Steimle and M. Mühlhäuser, "Permulin: mixed-focus collaboration on multi-view tabletops", *Proc. CHI '14*, 2014, ACM, pp. 3191-3200.
- [15] Y. Rogers and S. Lindley, "Collaborating around vertical and horizontal large interactive displays: which way is best?",

- Interacting with Computers*, vol. 16, no. 6, 2004, pp. 1133-1152.
- [16] E.W. Pedersen and K. Hornbæk, "An experimental comparison of touch interaction on vertical and horizontal surfaces", Proc. *NordiCHI '12*, 2012, ACM, pp. 370-379.
  - [17] W. Buxton, G. Fitzmaurice, R. Balakrishnan and G. Kurtenbach, "Large displays in automotive design", *IEEE Comput. Graph. Appl.*, vol. 20, no. 4, 2000, pp. 68-75.
  - [18] H. Benko, A.D. Wilson and R. Balakrishnan, "Sphere: multi-touch interactions on a spherical display", Proc. *UIST '08*, 2008, ACM, pp. 77-86.
  - [19] P. Marshall, R. Morris, Y. Rogers, S. Kreitmayer and M. Davies, "Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface", Proc. *CHI '11*, 2011, ACM, pp. 3033-3042.
  - [20] A.M. Piper and J.D. Hollan, "Tabletop displays for small group study: affordances of paper and digital materials", Proc. *CHI '09*, 2009, ACM, pp. 1227-1236.
  - [21] I. AlAgha, A. Hatch, L. Ma and L. Burd, "Towards a teacher-centric approach for multi-touch surfaces in classrooms", Proc. *ITS '10*, 2010, ACM, pp. 187-196.
  - [22] A. Soro, S.A. Iacolina, R. Scateni and S. Uras, "Evaluation of user gestures in multi-touch interaction: a case study in pair-programming", Proc. *ICMI '11*, 2011, ACM, pp. 161-168.
  - [23] C. Ardito, M.F. Costabile, R. Lanzilotti, A. De Angeli and G. Desolda, "A Field Study of a Multi-Touch Display at a Conference", Proc. *AVI'12*, 2012, ACM, pp. 580-587.
  - [24] H. Brignull and Y. Rogers, "Enticing people to interact with large public displays in public spaces", Proc. *INTERACT '03*, 2003, IOS Press, IFIP, pp. 17-24.
  - [25] J. Bruner: "Acts of Meaning" (Harvard University Press, 1990).
  - [26] V. Braun and V. Clarke, "Using thematic analysis in psychology", *Qualitative Research in Psychology*, vol. 3, no. 2, 2006, pp. 77-101.
  - [27] P. Dillenbourg, "What do you mean by collaborative learning?", *Collaborative-learning: Cognitive and computational approaches*, 1999, Elsevier, pp. 1-19.
  - [28] W.W. Gaver, "The affordances of media spaces for collaboration", Proc. *CSCW '92*, 1992, ACM, pp. 17-24.
  - [29] D. Ianes, "The Italian model for the inclusion and integration of students with special needs: Some issues", *Transylvania Journal of Psychology, Special Issue No. 2, Supplement No.*, vol. 1, 2006, pp. 117-127.
  - [30] Y. Rogers, H. Sharp and J. Preece: "Interaction Design: Beyond Human - Computer Interaction" (Wiley, 2015, 4th edn).
  - [31] Y. Dittrich, M. John, J. Singer and B. Tessem, "For the Special issue on Qualitative Software Engineering Research", *Information and Software Technology*, vol. 49, no. 6, 2007, pp. 531-539.

# Experiencing a new method in teaching Databases using Blended eXtreme Apprenticeship.

Vincenzo Del Fatto, Gabriella Dodero  
Faculty of Computer Science  
Free University of Bolzano - Bozen  
Bolzano - Bozen, ITALY  
vincenzo.delfatto@unibz.it, gabriella.dodero@unibz.it

Roberta Lena  
Dipartimento Istruzione e Formazione italiana,  
Provincia Autonoma di Bolzano  
Bolzano - Bozen, ITALY  
roberta.lena@scuola.alto-adige.it

**Abstract**—The traditional approach of teaching Databases requires a great effort on the initial aspects of modeling and design that can negatively affect student's motivation. This consideration led us to propose a method able to motivate students. The paper presents an innovative method of teaching a Database course, as well as the results of its experience in an Italian high school. This method is based on a blended approach of the Extreme Apprenticeship methodology, jointly with a specific organization of the course topics, which, compared to the traditional organization, has been strongly adapted to the paradigm of Learning by Doing. Good results in terms of students' performance and in terms of level of learning perceived by the students have been achieved. The perceived level of learning achieved by the students has been measured through a questionnaire administered at the beginning, in the middle and at the end of the course.

*Blended eXtreme Apprenticeship; DataBase Teaching; Learning by Doing*

## I. INTRODUCTION

Traditionally, in Database (DB) courses design and modeling phases are covered in the initial part of the course; and the practical use, through exercises on a Data Base Management System (DBMS), is usually scheduled at the end of the course. Therefore, DB learning usually starts from abstract and complex design aspects, and ends with simple data applications. This approach requires a great initial effort on modeling and design, which are fully understood by the students only by applying them, at the end of all thematic blocks. This approach has strong negative consequences on teaching efficacy. These considerations are behind our proposal for a method that keeps students' motivation high, combined with a high level of learning perception of course topics. The method is based on a blended approach of the eXtreme Apprenticeship (XA) methodology, complemented by an organization of topics strongly adapted to the learning by doing paradigm. In the proposed method, all phases of DB design are experienced in parallel, being reinforced from the beginning by a large number of practical exercises.

The paper is structured as follows. Section II presents the basic concepts of the XA methodology and a brief analysis of the traditional DB teaching. In Section III the context of the experience and the reasons that led us to the proposal of the new method are presented. In Section IV the structure of the proposed method is detailed. Section V presents the results of the experience. In section VI the final considerations and future work are presented.

## II. BACKGROUND

### A. eXtreme Apprenticeship

An innovative teaching methodology called eXtreme Apprenticeship (XA) was recently applied at the Free University of Bozen-Bolzano. This methodology has been developed in 2010 at the University of Helsinki, and applied in Introduction to Programming courses [14], showing significant improvements over traditional formats of teaching [12, 15]. The basic principles of XA are the following:

- learning through practice (Learning by Doing),
- formative assessment, carried out through a continuous bidirectional feedback between teacher and student.

XA is based on Cognitive Apprenticeship [4], which refers to the teaching method in old workshop, where the master first exemplifies the tasks, then drives the apprentices progressively to acquire autonomy [5]. XA is divided into three phases:

1. Modeling: The teacher provides, through working examples, a model of how an expert does the job.
2. Scaffolding: the student performs, after modeling phase, a number of exercises under the guidance of the master. Teacher support is based on Vygotsky's *Zone of Proximal Development* [16].
3. Fading: gradual reduction of the teacher support.

By carrying out a large amount of exercises with relatively small goals, the student has a continuous perception of his

cognitive progress, with a positive effect on self-esteem and self-efficacy. The teacher supports the students' motivation providing feedback, useful for improving his learning. Exercises proposed by XA contain in their text those theoretical basic information that is strictly necessary for immediately starting with the practice, and to gradually introduce, within the next exercises, the concepts necessary to achieve the intended cognitive objectives.

The adoption of XA in University courses led to a decrease in dropout rate, and an increase in the percentage and in grades of students who passed the exams [6, 7, 8, 9, 12, 15]. Dodero and Di Cerbo [9] developed a blended version of XA, as in previous blended teaching experiences [11]. In the blended approach an on-line setting has been implemented, where scaffolding is provided to the students by individual, asynchronous feedback messages. The results obtained with blended XA were positive, and comparable to those obtained with in-presence tuition and individual real time feedback by the original XA approach [12, 14, 15].

### B. Teaching Databases

Traditionally, DB teaching proposes the phases of conceptual, logical and physical design in consecutive and separate blocks. This approach is based on the engineering principle of splitting what has to be represented in a DB (conceptual design) from its implementation (logical and physical design). Each phase provides a detailed and exhaustive discussion of every topic. The practice on a DBMS software is usually scheduled at the end of such a theoretical part, or it is interleaved with the theoretical part, resulting loosely interconnected with the theory. This teaching technique is commonly applied in Academia [2, 13] and moderately applied in high schools [1, 10] where other unplugged methodologies, typically targeted to primary schools, are also sometimes used [3].

However, a detailed discussion proposed in consecutive, loosely connected blocks can negatively impact on learning, namely on attention level and motivation of the students. Most students do not grasp an overview of all three topics, and often acquire just technical, non-connected skills. Moreover, the first thematic blocks of the DB theory require a sustained effort and mental abstraction by students, who only at a later stage will have a feedback through DBMS interaction.

## III. A NEW TEACHING METHOD FOR DATABASES

This section presents the context of our experience and the reasons that led us to the proposal of the new method.

### A. The context of the experience

The proposed method was applied in an evening class of an Italian High School in Finance and Marketing Administration. The evening course provides a selected number of topics and a reduced amount of lectures, offering a Computer Lab for just two hours a week. A key feature of the evening school is the extreme variety in terms of type and level of education of its participants. The class consists of 23 students including:

- a group of young age students, with good computer science skills;
- a group of older working students of varying age (from 20 to 50 years old), with extremely heterogeneous computer science skills;

Some issues related to this context are listed below:

- the diversity of skills and maturity of the students,
- the evening hours, which affect the level of attention and fatigue,
- the motivation, that depends on the achievement of very different objectives.

### B. Choosing the Method

In such a heterogeneous context, where students' fatigue and motivation play a huge role in the dynamics of the lecture, the choice of teaching methods is crucial. Below we discuss the reasons that led us to adopting a new method, based on blended XA and on a revision of content scheduling.

#### 1) Methodological Aspects

Firstly, literature has shown that XA strengthens students' motivation [7, 14]. They are encouraged both through the proposed exercises with small cognitive goals, and by teacher's support. Using XA, students perceive the continuing evolution of their learning, supporting their self-efficacy and self-esteem. Second, the method is based on formative assessment of the student. Through a continuous bidirectional feedback with the teacher, formative assessment leads all students to achieve good basic skills, because they receive, step by step, all information needed to achieve the cognitive goal, and to acquire good practices. Finally, XA is based on Learning by Doing, a modality that is suited to learning in the evening hours. Learning through computer lab activities emphasizes the centrality of the student, and fosters meta-cognitive and self-assessment attitudes, allowing the student to verify the result of his action in the lab. XA, as proposed in [14], prescribes that all the exercises are carried out in the lab, in presence of a teacher who provides real-time support. Dodero and Di Cerbo [9] proposed a blended version of XA, where activities are scheduled both in presence of the teacher and as on-line activities. In Blended XA, at first students attend a lab session with scaffolding, and later, there is a gradual decrease of in presence scaffolding in favor of on-line scaffolding. In the experience described in this paper the following blended XA strategy has been adopted: each week, in addition to the two hour lab, students were given a few exercises to be solved as homework, with on line support from the teacher, through a Moodle LMS platform.

#### 2) Scheduling Course Content

Our method provides an innovative reorganization of course content, which differs from the traditional organization of many of DB courses, organized in phases according to a "horizontal" temporal sequence, ranging from the conceptual level to the physical level, and ending with a practical experience using a DBMS software. The new organization of topics is carried out by proposing exercises that engage the students to "vertically" work on the DB, ranging from the

conceptual design to the practical DBMS exercise in the same lab hour. In this way, students can grasp, from the very beginning, a complete overview of course topics, and understand how each topic contributes to the overall design of a DB.

#### IV. THE NEW METHOD STRUCTURE

This section details the structure of the proposed method. In our experience we adopted a new organization of topics, that vertically crossed the various DB design phases. As prescribed by XA, theoretical lectures have been replaced by lab sessions providing exercises. In each lab session, the first exercise was presented to the students, as a modeling phase done by the teacher: the students repeated the teacher's actions, first by drawing on paper the conceptual and logical models, then by implementing the practical part using pgAdminIII and PostgreSQL. The next exercises were carried out with the gradual decreasing support of the teacher (scaffolding and fading phase). To implement XA, the exercises have to:

- gradually propose higher cognitive goals,
- promote the acquisition of mastery to perform a task.

Therefore the exercises contained theoretical information and were designed as repetitive, to a certain extent. In the experience, three groups of exercises have been proposed for each topic, according to the following structure:

- the first exercise: an exercise with theoretical references, proposed in modeling phase. The teacher shows how she performs the exercise, leading the student to grasp the most significant aspects and to acquire proper procedures:
- the second group of exercises (Now you try it!): an exercise similar to the first one, carried out by the student with the support of the teacher. The exercise may contain new theoretical references and repeats the mechanisms proposed in the first exercise, with minor cognitive deviations. The intervention of the teacher can be scaffolding or fading;
- the third group of exercises (Try again!): further exercises requiring repetition of knowledge and skills already applied in the first two groups of exercises. Also in this case the intervention of the teacher can be scaffolding or fading.

Each exercise required the modeling of a reality of interest, adopting the following structure:

- ER Model;
- Relational Model;
- Creating tables (using SQL);
- Operations on tables (insertions, deletions, queries).

Each exercise was designed as a sequence of smaller exercises, to engage students at different stages of the design phase, and in DB manipulation and interrogation. As an example, an exercise proposed during the third lab was the following:

**EXERCISE** – Let us create a DB for a library that manages DVDs rental.

- Draw the ER diagram and the relational schema of the following database entities: (i) DVD, with *Cod\_dvd*, *Title* and *Duration* attributes; (ii) CATEGORY, with *Cod\_category* and *Name* attributes.
- Write the SQL code to create the DVD table, taking into account the following properties and constraints: (i) *Cod\_dvd*: character type (5), not null; (ii) *Title*: character type (50), not null; (iii) *Duration*: integer, not null; (iv) Primary key constraint.
- Write the SQL code to create the CATEGORY table taking into account the following properties and constraints: (i) *Cod\_category*: type character (3), not null; (ii) *Name*: type character (20), not null; (iii) Primary key constraint.
- Write the SQL code to insert in the DVD table the following data:

<i>Cod_dvd</i>	<i>Title</i>	<i>Duration</i>
D0100	The Blues Brothers	133
D0101	The Great Dictator	125
D0102	The Silence of Lambs	118

- Write the SQL code to insert in the CATEGORY table the following data:

<i>Cod_category</i>	<i>Name</i>
C001	Horror
C002	Thriller
C003	Comedy
C004	Action

- Extract Title and Duration from DVD table and sort by ascending title.
- Extract from the DVD table all data containing the word "Blues".

In each lab students solved 3 exercises, and got 1 exercise as homework, to be delivered within 5 days. Before the next lab, the homework was corrected by the teacher and, if failed, there was still time to improve and deliver a flawless exercise. Students were free to choose the format and the method of delivery: as a text file, or on paper; by e-mail or by uploading the exercise in the personal folder of the school's LMS (Moodle). During the scaffolding and fading phases, the assessment of exercises resulted in a formative assessment, which is not intended to assign a grade, but to identify areas for improvement and to implement corrective strategies. For this reason the grades ranged between two possible values:

- 1: indicates that the learning objectives of the exercise have been met by the student.

- 0: indicates, on the contrary, that the learning objectives of the exercise were not met by the student.

## V. RESULTS

This Section presents the results of the experience, which was conducted from December 2014 to April 2015, as 2 lab hours and a blended homework each week.

### A. Level of learning perceived by the students

At the beginning, in the middle, and at the end of the course, an anonymous questionnaire was submitted to students in order to detect the initial, intermediate and final level of knowledge in DB topics in terms of perceived level of learning by the students. The questionnaire asked the student to self-assess on a Likert scale from 1 (level zero) to 5 (excellent level) w.r.t the following topics:

- Experience with a DBMS
- Ability to query a DB
- Knowledge of ER model
- Experience with SQL

Results from the initial questionnaire show that knowledge about the proposed topics was minimal; it mostly concerned tables and queries (see Fig. 1 and Fig. 2). Both the logical and conceptual design of a DB and the standard SQL language were unknown to most of the class (see Fig. 3 and Fig. 4). All the bar charts in this section show the number of students on the Y-axis and the level of perceived learning on the X-axis.

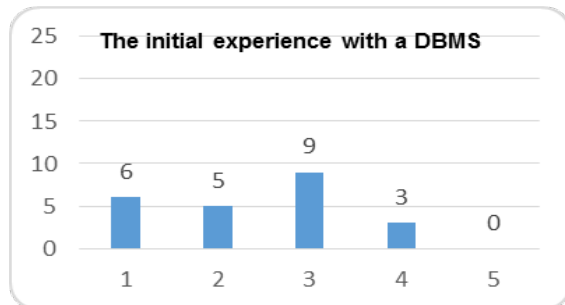


Figure 1. The initial experience with a DBMS perceived by the students

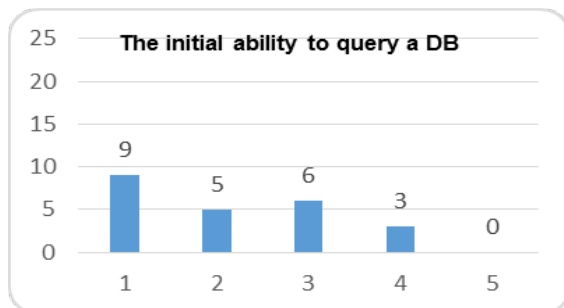


Figure 2. The initial ability to query a DB perceived by the students

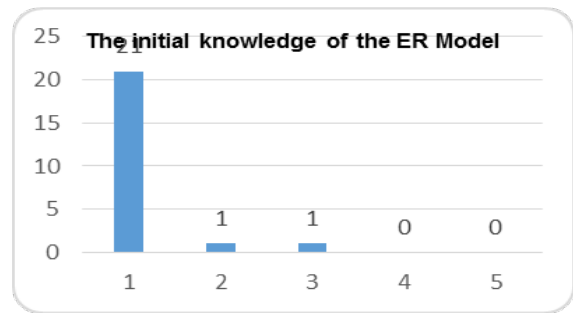


Figure 3. The initial knowledge of the ER Model perceived by the students

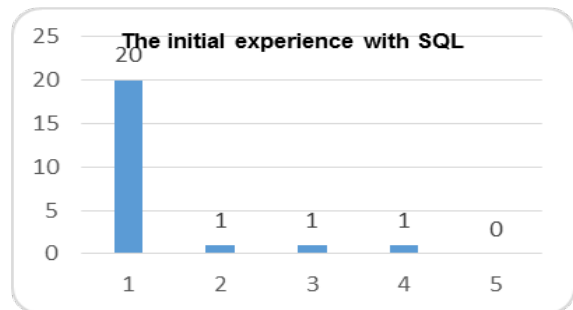


Figure 4. The initial experience with SQL perceived by the students

The same questionnaire was proposed to students for a mid-term evaluation. Results of the intermediate questionnaire show a significant increase in terms of perceived improvement, as well as a similar distribution of perceived levels, for each of the four topics proposed (see Fig. 5, Fig. 6, Fig. 7 and Fig. 8).

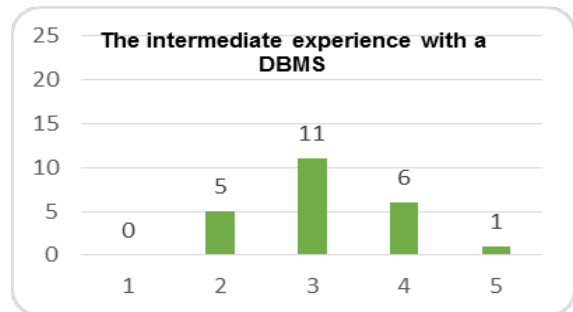


Figure 5. The intermediate experience with a DBMS perceived by the students

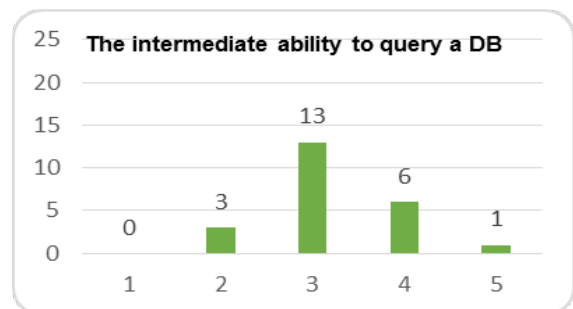


Figure 6. The intermediate ability to query a DB perceived by the students



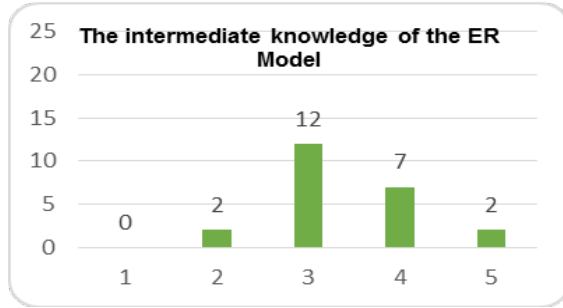


Figure 7. The intermediate knowledge of the ER Model perceived by the students

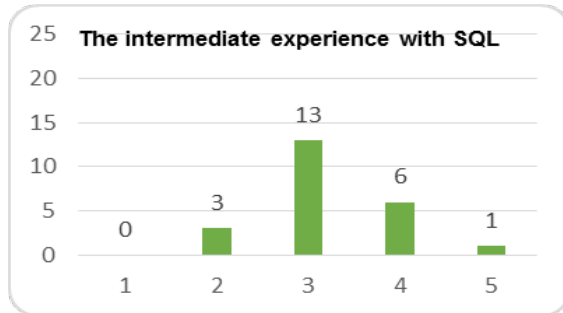


Figure 8. The intermediate experience with SQL perceived by the students

This good result is due to the new organization of the course topics: the phases of the DB design, and the use of a DBMS, have been proposed in parallel; thus, students' mastery about different aspects has increased simultaneously. Results of the intermediate questionnaire show that, for all topics, the number of students with the same level of perceived learning is homogeneous. The effectiveness of the simultaneous development of interconnected skills is even more evident noticing that in the last two topics (see Fig. 7 and Fig. 8), respectively, knowledge of the ER model and experience with SQL, the intermediate distribution shows a clear improvement compared to the initial situation (see Fig. 3 and Fig. 4). A further important result of the experience is highlighted in the intermediate questionnaire, where results show a Gaussian distribution shifted on medium-high values (3 and 4).

The same questionnaire was proposed to students for a final evaluation. It reveals two significant results of the experience. For all topics, the number of students with the same level of perceived learning is homogeneous (see Fig. 9, Fig. 10, Fig. 11 and Fig. 12). Also this result relates to the new organization of course topics, proposed in parallel. Results of the final questionnaire show that level of learning and self-efficacy perception of the students has a Gaussian distribution shifted on medium-high values (3, 4 and 5), which is higher w.r.t intermediate results. This depends on the application of the XA methodology and particularly to formative assessment that allows the whole class to obtain good results

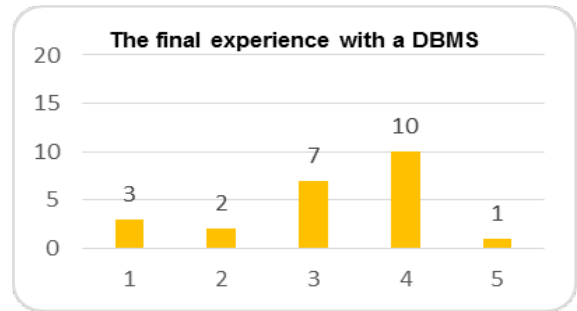


Figure 9. The final experience with a DBMS perceived by the students

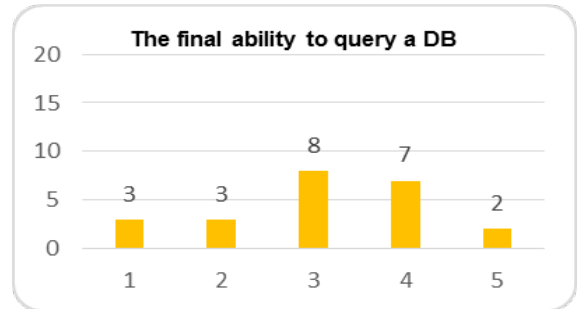


Figure 10. The final ability to query a DB perceived by the students

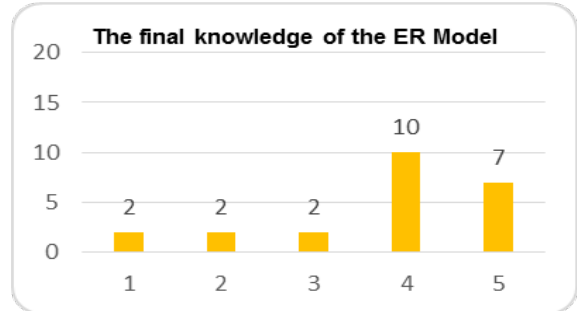


Figure 11. The final knowledge of the ER Model perceived by the students

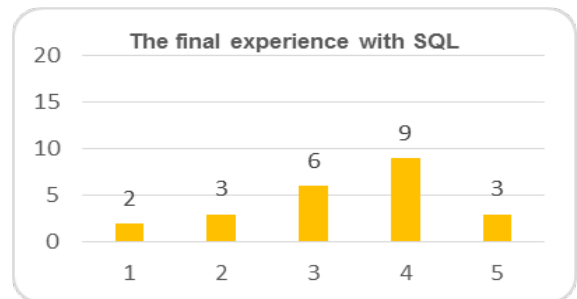


Figure 12. The final experience with SQL perceived by the students

For a better understanding of improvements in experience results in terms of level of learning perceived by the students in the initial, intermediate and final questionnaire, the following figures show a comparison among the specific questionnaire topics.

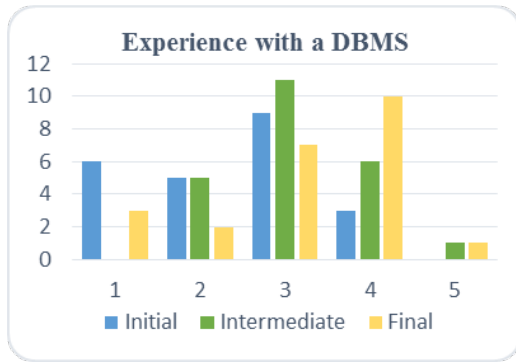


Figure 13. Comparison among initial, intermediate and final experience with a DBMS perceived by the students

Fig. 13 shows the comparison among initial, intermediate and final experience with a DBMS perceived by the students. About 80% of students are between 1 and 3 (low-medium values) in the initial questionnaire, and between 3 and 5 (medium-high values) in the intermediate and final questionnaires.

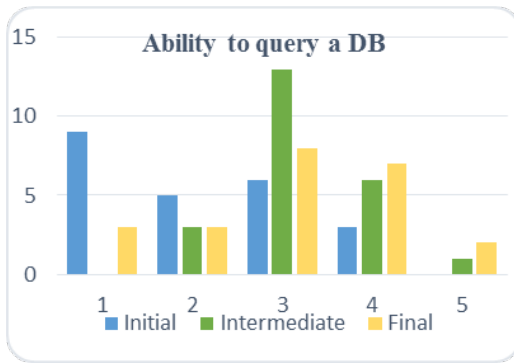


Figure 14. Comparison among initial, intermediate and final ability to query a DB perceived by the students

Fig. 14 compares initial, intermediate and final ability to query a DB as perceived by the students. About 85% of students are between 1 and 3 (low-medium values) in the initial questionnaire, and between 3 and 5 (medium-high values) in the intermediate questionnaire. In the final questionnaire about 75% of students is between 3 and 5 (medium-high values).

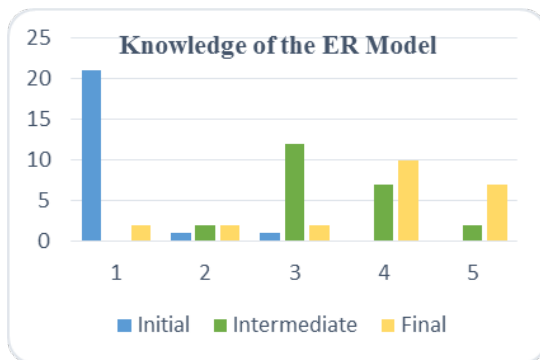


Figure 15. Comparison among initial, intermediate and final knowledge of the ER Model perceived by the students

Fig. 15 compares initial, intermediate and final knowledge of ER Model perceived by students. In particular, all the students are between 1 and 3 (low-medium values) in the initial questionnaire and about 80% of students are between 3 and 5 (medium-high values) in the intermediate and final questionnaires.

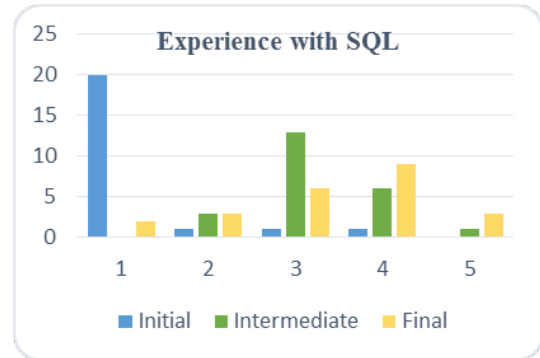


Figure 16. Comparison among initial, intermediate and final experience with SQL perceived by the students

Fig. 16 compares initial, intermediate and final experience with SQL perceived by students. About 100% of students is between 1 and 3 (low-medium values) in the initial questionnaire and about 80% of students are between 3 and 5 (medium-high values) in the intermediate and final questionnaires.

We remark that, for a correct interpretation of these results, the context of the experience has to be kept in mind. There was a single teacher in a class of 23 students, some of whom seldom attending the lectures. Optimal support for such a class would be typically given by one teacher every 10-15 students. Results achieved in the experience have shown, already at midterm, the effectiveness of the adopted method in terms of both summative evaluation results and self-efficacy perception of the students. Certainly, a better student/teachers ratio would allow for a shift on higher levels of the final results.

## B. Summative Assessment

As per school quality plan, two summative assessments were planned, respectively, in the middle and end of the course. The format of the summative assessment was similar to those of the past years, i.e. no different performance criterion was applied because of XA. Fig. 17 shows the distribution of the grades of the intermediate and final assessments. The grades of both summative assessments, mostly in the medium-high range, with at least 70% of the grades between 7/10 and 9/10, suggest that the good perception of students, described in the previous sub-section, was transformed into a successful final evaluation (see Fig. 17, 2° assessment). As in the University courses, this good result is a consequence of the use of XA. Formative assessment fosters mastery acquisition in performing the task, allowing all students in the class to achieve good cognitive results.

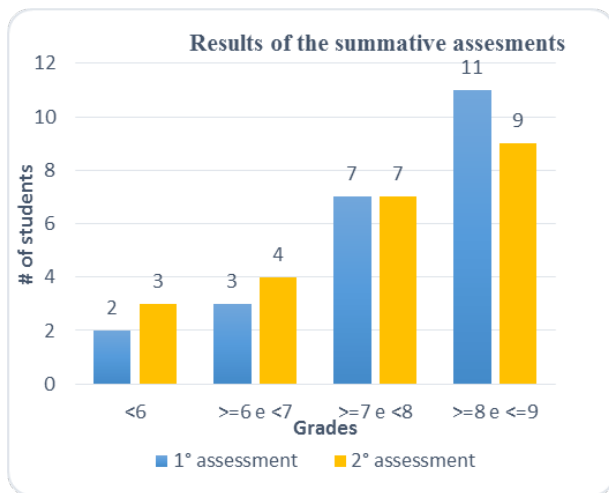


Figure 17. Distribution of grades in the two assessments

It is worth nothing that the trend of grades is very similar in the intermediate and final assessment. This highlights how XA achieves good results in a short period of time. The small decrease in grades in the second and final assessment is due to the complexity of the exercises. Finally, the systematically low grades were scored by students who seldom attended the course, compromising the creation of an effective relationship with the teacher and the outcome of the methodology. These results are very positive, especially in the light of the initial background of the students, which revealed a total lack of knowledge about the DB design and the SQL language.

## VI. CONCLUSIONS AND FUTURE WORKS

In this paper a new teaching methodology for Database courses and its experience in an evening class of an Italian High School in Finance and Marketing Administration have been presented. In previous experiences, the application of both XA and blended XA has shown that this approach is suitable for different domains (Math, Programming, Operating Systems) as well as for different levels of students (High schools and Academia). In this work we describe as the blended XA together with a new organization of thematic blocks of exercises has achieved good cognitive results in a short time for most of the class also in the Database domain. Results have been measured in two ways, considering grades scored in two summative assessments, respectively, in the middle and end of the course, and collecting answers to a questionnaire to detect the level of perceived knowledge of DB topics by the students. The questionnaire was distributed three times, at the beginning, in the middle and at the end of the course. Results are very positive with respect to the initial situation, characterized by minimal knowledge on the subject. Finally, the experience showed that this approach performs best when teacher support is adequate. Therefore, in large high school classes the method is best applicable where co-teaching is possible. As future work, we plan to decrease the teacher's effort spent in formative assessment through automatic or

semiautomatic exercise correction tools, to better support the teacher in the application of the method, and guaranteeing better scalability in large classes.

## REFERENCES

- [1] A. V. Aho, J. D. Ullman. Foundations of Computer Science. Computer Science Press, Inc., New York, NY, USA. 1992.
- [2] P. Azteni, S. Ceri, S. Paraboschi, R. Torlone. Basi di dati. Seconda edizione. Milano:McGraw-Hill. 1999
- [3] T. Bell, I. Witten, M. Fellows. Computer science unplugged. (002) [http://csunplugged.org/wp-content/uploads/2015/03/CSUnplugged\\_OS\\_2015\\_v3.1.pdf](http://csunplugged.org/wp-content/uploads/2015/03/CSUnplugged_OS_2015_v3.1.pdf)
- [4] A. Collins, J. Brown, S. Newmann. Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. Champaign, University of Illinois at Urbana-Champaign. 1987.
- [5] A. Collins, "Cognitive apprenticeship," The Cambridge handbook of the learning sciences, p. 4760, 2006.
- [6] V. Del Fatto, G. Dodero, R. Gennari, R. Assessing Student Perception of Extreme Apprenticeship for Operating Systems 14th IEEE ICALT conference. Atene. 2014.
- [7] V. Del Fatto, G. Dodero, R. Gennari. How Measuring Student Performances Allows for Measuring Blended Extreme Apprenticeship for Learning Bash Programming. In Computers in Human Behavior Journal, Elsevier, in press in 2015. DOI: 10.1016/j.chb.2015.04.007
- [8] V. Del Fatto, G. Dodero and R. Gennari. Operating Systems with Blended Extreme Apprenticeship: What Are Students' Perceptions. In Interaction Design and Architecture Journal (IXD&A), special issue, 2015.
- [9] G. Dodero, F. Di Cerbo, Extreme Apprenticeship Goes Blended: An Experience, ICALT 2012, IEEE International Conference on, Advanced Learning Technologies, IEEE, 2012, 324-326.
- [10] P. Gallo, P. Sirsi. Cloud, Informatica-secondo biennio, Istituti tecnici – settore economico, indirizzo Amministrazione, Finanza e Marketing. 2012.
- [11] D. Garrison and H. Kanuka, "Blended learning: Uncovering its transformative potential in higher education," The Internet and Higher Education, vol. 7, no. 2, pp. 95 – 105, 2004.
- [12] J. Kurhila and A. Vihavainen, "Management, structures and tools to scale up personal advising in large programming courses," in Proceedings of the 2011 conference on Information technology education, ser. SIGITE '11. New York, NY, USA: ACM, 2011, p. 38. [Online]. Available: <http://doi.acm.org/10.1145/2047594.2047596>
- [13] Silberschatz, Korth, Sudarshan: Database System Concepts (5th ed, or later) McGrawHill, 2005
- [14] A. Vihavainen, M. Paksula, and M. Luukkainen, "Extreme apprenticeship method in teaching programming for beginners," in Proceedings of the 42nd ACM technical symposium on Computer science education, ser. SIGCSE '11. New York, NY, USA: ACM, 2011, p. 9398. [Online]. Available: <http://doi.acm.org/10.1145/1953163.1953196>
- [15] A. Vihavainen, M. Paksula, M. Luukkainen, and J. Kurhila, "Extreme apprenticeship method: key practices and upward scalability," in Proceedings of the 16th annual joint conference on Innovation and technology in computer science education, ser. ITiCSE '11. New York, NY, USA: ACM, 2011, p. 273277. [Online]. Available: <http://doi.acm.org/10.1145/1999747.1999824>
- [16] L. Vygotski, Mind in society: The development of higher psychological processes. Harvard Univ Pr, 1978.

# A Smart Material Interfaces Learning Experience

Andrea Minuto  
HMI Group, University of Twente  
PO Box 217, NL-7500 AE Enschede,  
The Netherlands  
a.minuto@utwente.nl

Fabio Pittarello  
Università Ca' Foscari Venezia  
Via Torino 155  
Venezia, Italia  
pitt@unive.it

Anton Nijholt  
HMI Group, University of Twente  
PO Box 217, NL-7500 AE Enschede,  
The Netherlands  
a.nijholt@utwente.nl

**Abstract**—This paper describes a learning experience held with a class of primary school children who were introduced to a novel class of resources, named smart materials, and the interfaces built with them (Smart Material Interfaces). The pupils were guided along a multidisciplinary educational path in which traditional and innovative teaching methods were composed for educating while engaging the children. It led to the creation of 6 automated puppet plays focused on the themes of environmental awareness as a result. In this process, storytelling and visual programming acted as powerful means for merging different educational concepts and techniques. The children's engagement and the educational impact were evaluated during and after of the experience, revealing interesting results. The data collected through the direct observation and the questionnaires indicate that the experience was perceived as a positive and interesting. The post evaluation, held some months later, revealed improvements in all the areas involved by the multidisciplinary experience, from the knowledge of the properties of smart materials and the programming skills, to the increase of the environmental awareness and the skills for text analysis.

**Keywords:** Arduino, computer supported education, origami, smart material interface, Scratch, storytelling, visual programming.

## I. MOTIVATION

Educators should be able to offer up-to-date educational paths capable of integrating the novelties of science and technology with the engagement of the pupils for improved learning. Smart materials represent a novel and interesting technological topic to teach and learn. They can change their physical properties (for example color, shape, stiffness and so forth) and they can be manipulated and controlled through different hardware platforms (e.g. Arduino) for the creation of interesting and engaging interfaces (i.e. Smart Material Interfaces, SMIs). The interest of this exploratory study lays in the introduction of these complex technology topics in the Primary School and on the design of an interdisciplinary educational path supporting this goal. For reaching this goal the educational experience included scientific, technical, artistic topics and literacy skills, meant for engaging the children while educating them. It is important to underline that the topics that were introduced for stimulating the interest for the smart materials worked not just as a means but they were themselves a focus of interest. Storytelling, which has long

been recognised as a powerful means for engaging children in educational contexts, was used in this work as a glue for connecting the different educational topics. Storytelling provided an overall goal to the students' work: the creation of stories focused on environmental awareness themes. In this experience the children were challenged in the creation of origami<sup>1</sup> models as elements of a story. These elements were augmented with smart materials and programmed to act as the stories created by the children prescribed. A Scratch<sup>2</sup> based environment was used as mediating tool for translating the narrative structures into programming blocks. It connected the models to an Arduino board<sup>3</sup> for triggering the actions of the associated smart materials. The Arduino-controlled stories were finally represented in cardboard theatres.

The data collected during and after the experience indicate that the educational path was perceived as engaging and that the children improved their skills and knowledge. We collected these data during and after the experience using different means: direct observation, video recording, questionnaires and also replicating some tasks a few months after the end of the project. The engagement was positively evaluated through the analysis of various parameters [1] such as: perceived usability, felt involvement, focused attention, aesthetics, novelty and endurability. The educational impact was measured some months after the end of the pedagogical path and revealed improvements in all the areas involved by the multidisciplinary experience. In the next sections we will show an overview of related works (Sec. II), we will give an explanation of the material used (Sec. III), and the teaching process (Sec. IV). We will then describe the evaluation conducted and analyse our results (Sec. V-VIII). This will be followed by our conclusions (Sec. IX).

## II. RELATED WORKS

There are many different ways to engage and attract the interests of younger minds. Among these ways to increase motivation, one possibility is to make the task more enjoyable. It is possible for example to use interfaces made of physical objects, often belonging to the everyday experience, instead

<sup>1</sup>The Japanese art of folding paper into shapes and figures.

<sup>2</sup>A visual programming tool <http://scratch.mit.edu> from MIT

<sup>3</sup>Arduino is an open hardware platform <http://arduino.cc>

of traditional ones based on the WIMP [2] paradigm<sup>4</sup>. One of these novel interfaces is described in [3], where Sun and Han tested different kinds of input interfaces, such as: keyboards, aluminium foil pads and bananas. Even though the bananas scored as the worst in performance, they were also the best for engagement and enjoyment. Other possibilities are toys such as Makey Makey [4], that proved to be an interesting tool to create tangible interfaces with children [5]. Makey Makey allows the use of everyday (conductive) objects such as fruits to create interactive interfaces that can drive games made with Scratch. This allows the children to have fun with games, while learning and improving their personal skills. In this context of development of games for children with Scratch we need to note that other visual programming environments have contributed significantly to the field, an example is Blockly [6]. Blockly is a similar visual programming editor (usable via browser, without installation of software or plugins) that allows children to learn programming while playing. All of the above studies also tended to ask the children to make things, to produce their own object of play. As we know from [7] [8], “many studies [...] suggest that storytelling (meant as the capacity to listen, tell, and reflect on stories) is an extremely important developmental area for children, promoting a wide spectrum of cognitive functions and skills: expression, communication, recognition, recall, interpretation, analysis, and synthesis”. Some experiences related to storytelling take advantage of visual programming languages. Different researchers have designed and experimented with visual paradigms for children, with the goal of teaching them to program. Alice [9], one of the most famous languages, allows children to program a 3D environment using a drag and drop style. Looking Glass [10], a successor, introduces children to programming by coupling 3D and storytelling. Scratch is a block based graphical programming language that permits children to build 2D stories and games. Jacoby and Buechley experimented with children a different approach to storytelling with new tangible technologies such as conductive ink [11]. They taught children about circuitry and conductivity with an interesting kit (StoryClip) to produce drawings that they could bring to life with their recorded speech, by enhancing traditional paper with augmented properties.

The educational project presented in this work takes advantage both of visual programming paradigms and augmented physical objects, for building an engaging storytelling experience. In our work the plain physical objects are augmented taking advantage of the properties of a new category of materials, named smart materials<sup>5</sup>. Our approach is part of a new research area focused on the exploration of new synergies between traditional materials and smart ones. A number of researchers involved in this research used just paper, in the artistic shape of origami, to engage the users. Boden et al. [12] describe a

system designed to support augmented play and learning for children. It uses origami and augmented reality with fiducial markers. In [13], Do and Gross try to explore the possibility of creating creative environments by using interactive spaces, and using origami as a means for teaching and learning geometry and spatial reasoning. Coehlo [14] theorised about embedding materials in the paper making process to create sensors and interactive surfaces. Others in the past have tried to couple new materials with toys, for example in [15] textile is described as a user interface for an interactive toy that responds to events by changing patterns. Smart materials gave a boost for creating interfaces for learning, teaching and most of all increasing and supporting creativity with many different techniques [16]. This new kind of interfaces making use of smart materials are also called Smart Material Interfaces<sup>6</sup>. They are already embedded in electronics and products of everyday use (e.g.: sunglasses that darken only in bright environments, glasses that remember their shape even after deformation, markers that appear when the temperature reaches a certain value i.e. liquid crystal thermometers, and so forth). But only recently have they started to be used in the creation of SMIs and in do-it-yourself projects. Some of these are more expensive, others cheaper, but all of them try to interest and empower the user in making things, in participating in the creative process.

### III. MATERIALS FOR THE EXPERIENCE

The smart materials used for this experience are of two kinds, the choice was based on the most aesthetic and interesting properties: changing shapes and colours. We used shape memory alloy (SMA) wires and thermochromic paints of various colours.

The *thermochromic* paint is a paint that has a thermic threshold, once this temperature limit is reached the paint becomes transparent. We applied a serpentine of resistive wire to the back of the paper to reach the necessary temperature gradient, this allowed us to “switch the color on and off” on command.

The *SMA*s is a big family of materials: we employed NiTiNOL. This specific kind has the property to contract once the temperature threshold is reached (Flexinol<sup>7</sup>). We created several actuators with it, to be applied to the children’s creations. They were made following the implementation shown in [18] but with only one degree of freedom. To control the temperature in both cases we used Arduino and a small paper board made with conductive tape with a MOSFET<sup>8</sup> on it. Each board was used to control a single origami model and the related animations.

For the educational experience we decided to use S4A<sup>9</sup>, a modified version of Scratch. S4A allows the control of actuators through the Arduino pins using the same language as Scratch. This way the children were able to program Arduino and to create animations for the smart origami models by

<sup>4</sup>W.I.M.P paradigm: acronym for Windows Icons Mouse and Pointer interaction paradigm, coined by Merzouga Wilberts in 1980, developed at Xerox PARC in 1973.

<sup>5</sup>A smart material is a material that can change a physical property in a controlled way (for example color, shape and so forth).

<sup>6</sup>SMI are interfaces that can relay information in a material way [17].

<sup>7</sup>More technical information can be found at <http://musclewires.com>.

<sup>8</sup>A MOSFET is a specific kind of transistor.

<sup>9</sup>Realised in the context of the EU project Citilab <http://seaside.citilab.eu>.



Session	Lesson description	Survey	Main Focus	Length
1	We taught the children how to make plain origami models and create stories with them(Sec. IV-A).	First	origami	1 half-day
2	We introduced SMIs with several small examples. (Sec. IV-B).		SMIs	1 half-day
3	The children modified their stories for adding smart materials to origami (Sec. IV-C).		narration, SMIs	1 half-day
4	The children broke down their stories into narrative blocks, introducing symbols to come closer to a programming language (Sec. IV-D).		narration, programming blocks	1 half-day
5	We explained the basics of programming in S4A and taught how to create origami animations on Arduino from S4A. We asked the children to program their stories(Sec. IV-E).	Second	S4A, Arduino	1 day
6	We prepared the setup for the final cardboard theatre representations (Sec. IV-F).		S4A, Arduino	1 half-day
7	The children saw the realisation of their work and filled in the second questionnaire (Sec. IV-G).		grading	1 half-day
8	We evaluated the educational impact after the end of the experience(Sec. IV-H).		Final evaluation	1 day

TABLE I  
A SHORT DESCRIPTION, TIME LINE AND FOCUS OF ATTENTION OF EACH PHASE



Fig. 1. Children making origami models during the first day.

themselves. For allowing the children to enter, create and record the play in the proper condition, we built three, two sided, cardboard theatres: each one is about 1 m. wide and about 1.8 m. tall (Fig. 2). We also made use of a smaller cardboard theatre for testing the stories.

#### IV. TEACHING PROCESS

We experimented our learning path with a class of 19 Primary School children, all aged 9. None of them had any prior experience or knowledge about programming or visual programming. Some of them displayed partial knowledge about the use of the WIMP paradigm. The project was developed through 8 sessions held in the classroom and accompanied by additional homeworks. During the whole process two researchers were in the classroom, assisted by the children's teachers. Most of the sessions included collaborative phases where the children were organised in 6 groups: 5 of which composed of 3 children and one of 4 (3 groups of boys, 2 of girls and a mixed group), that were maintained till the end of the project. During the experience we gave 3 sets of evaluations, the first two during the experience and one after its end. In Table I we summarise the timeline of the whole experience. In preparation for the experience the teachers debated with the children about environmental behaviours.

##### A. Session 1: How to make origami models to tell a story

The children learned the basic techniques for making origami models. They were taught how to make simple forms representing animals and other shapes (Fig. 1). They worked individually for almost the whole session. In the same session we organised them into groups and asked them to create simple stories to be presented in a small cardboard theatre. Here the origami models created in the morning would play the part of characters. At the end we assigned the first homework. We asked each group to create a narration inspiring positive environmental behaviours. The story had to be presented with origami models, as for the first presentation. A unique theme was assigned to each group: *energy consumption, light management, heating, mobility, water consumption and waste management*. We also gave each group a list of positive behaviours for inspiration, coming from the activity done before the start of this experience. The week after, before starting the second session, we listened to the stories that they had created. The children presented their stories in front of their classmates, taking advantage of one of the big cardboard theatres, also used for the final representation. One of the children read the narration while the other group members moved the related origami models by hand as puppeteers.

##### B. Session 2: Explain SMIs the easy way

This session was dedicated to introduce the children to SMIs. The children were given explanations about the meaning of SMI and they were given practical demonstrations of origami models enhanced with smart materials. We showed them how the NiTiNOL wire could be used for generating transformations in the shape of an origami dog. We focused their attention on the fact that the wire, the origami model connected to it, could undergo transformations as a result of the activation of a battery. We experimented also the use of a proximity sensor for triggering the transformation. We then showed how the same NiTiNOL wire could be used to obtain simple rotations of another origami model, with a different connection. We showed the children examples of thermochromic paintings applied to origami representing a whale, a crane, a Christmas tree and a house. We focused the children's attention on the fact that the thermochromic inks disappeared when a battery was connected to a resistive wire



positioned under the painted surface, as a result of the paper heating up. In some cases the activation of the battery resulted in simple chromatic effects, while in other cases it revealed hidden drawings. The activation of the smart materials was manually triggered by the researchers or by the children. After the practical demo the children had the opportunity to examine the smart materials. They experimented by heating the colours with the natural warmth of their hands.

### C. Session 3: Let's modify the stories to make them animated

At the end of the previous session the children were given their second homework. The groups had to modify the stories they had created to take advantage of the possibilities offered by the smart materials. The idea was to animate the origami models mechanically and to create dynamic effects with the smart paint. For technical reasons, we did not allow the use of more than three smart origami models, but we gave the freedom to choose among any combination of NiTiNOL wires and thermochromic paints. The children were given a questionnaire with both closed and open questions. We used it for assessing the children's opinions after the first contact with the smart materials. A week later, we analysed the children's homework. The groups had chosen different combinations of mechanical animations and chromatic effects, sometimes modifying the narration for adapting it to the introduction of smart materials. For example a group with a story about a dog and four frogs living in a house, took advantage of smart materials for animating the most dynamic character, the dog. The children decided to use thermochromic paint for revealing the content of one of the posters written by the dog. They also added a sequence at the end of the narration where the curtains, painted in thermochromic ink, disappeared to let the sunlight in. We assisted the children in creating or modifying the existing origami models to support the edited narrations. Because of time constraints, we did not let the children fix the resistive wire. But they applied the thermochromic paints to their paper works and we positioned the wire for them after the end of the session. At the end of the session we showed the children how the puppet plays could be triggered not only manually, but also by a software program, generating complex automated shows. We gave them a test play in the cardboard theatre of a previously written story as an example of the final result. In it, the smart origami models, that we had shown them in the previous session, acted as automated characters while a recorded narration was played (Fig. 2).

### D. Session 4: Analyse the story and split in narrative blocks

At the end of the previous session, the groups were assigned their third homework: to analyse the story that they had created and identify its basic components. We asked them to identify with arrows the different sentences, distinguishing the parts played by the narrator and by the different characters. We asked the children to introduce or identify the audio effects that they wanted in the story by using additional arrows and to use a circular arrow with a number for identifying the repetitions (e.g., a triple barking should have been represented with the



Fig. 2. The demo the researchers created to show how the smart origami models work with the story. From the left: 1) in orange, the barking dog NiTiNOL actuated 2) a small crane with thermochromic wings 3) a house with flowers hidden behind the thermochromic curtains.

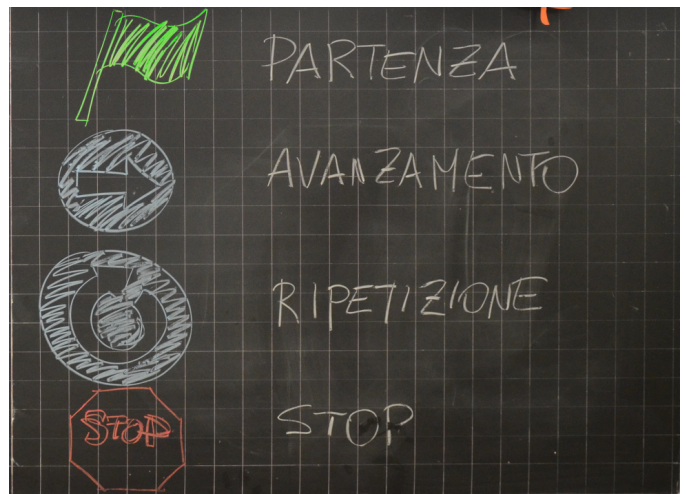


Fig. 3. Instruction marks, from the top: START, GO ON, REPEAT, STOP.

label *bark* preceded by the circular arrow and the number 3). We introduced the green flag symbol for specifying the beginning of the story and the red signal with the label *stop* for identifying the end of the story (Fig. 3). The children could use the same symbol, accompanied by a numerical label if they wanted to introduce pauses in the narration (e.g., a pause of 2 seconds between sentences spoken by a characters). For easing their work, we gave them a complete analysis' example: the story text we had just played (Fig. 2) with the symbols.

### E. Session 5: Programming (in S4A) and record the story!

This session lasted a whole day. During the first part, we showed the theatrical play of our story again, then we explained how everything worked. We showed how the smart origami models (SMIs) were connected to Arduino and how they could be controlled from S4A. We showed the children how the structures of our story had been translated into visual code that could be easily read and run by clicking

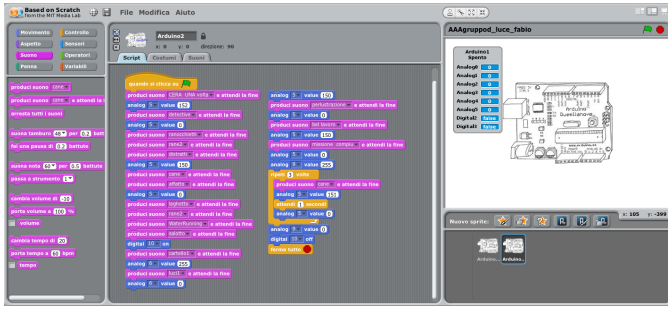


Fig. 4. A screenshot of S4A. From the left: 1) the library for changing the Arduino pins 2) a program 3) the area where the values are displayed.

a green flag placed at the beginning. We showed examples on how the narrative blocks and the other structures could be translated into visual programming entities. We focused on the programming entities that we could associate with the narrative blocks: the *play sound* block for playing the narration fragments and the special effects and the *analog* block for controlling the smart origami figures. We also showed the different control structures that we had mapped to the narration (i.e., the sequence of blocks, *start* and *stop*, *repeat* and *wait*) and parameters. After the explanation each group was supervised by a researcher while translating their stories into the programming structures with S4A. We checked the correctness of the homework. The process continued with the selection of the different blocks mapped onto narrative structures, the recording of the audio fragments of the story, the selection of the audio effects and the specification of the parameters associated to the control structures. The children tested and ran the program till they reached a satisfying result. Fig. 4 shows the visual code of the story of the dog and the four frogs. The children took advantage of all the blocks described in our tutorials: the magenta *play sound* and the blue *analog* block for animating the origami figures (i.e., the SMI connected elements). This group also introduced a lamp in their narration, so we taught them how to control it through the blue *digital* block. The children used the control structure for starting, pausing and ending the narration. The *repeat* was used to reproduce the barking sound and corresponding animation (repeated contraction and relaxation of the NiTiNOL wire).

#### F. Session 6: Fix the scene! Proceed with animations

After the story translations, we prepared for the final representations of the stories. We used 3, two sided, cardboard theatres. Each side was coloured for the environmental theme associated to the narration. After positioning the background scenarios created by the children, we positioned the smart origami models on the scene and connected them to Arduino and to a laptop running S4A. Because of their age, we did not involve the children in the electric connections. We realised all the process in the classroom, so that they could see how the results of their efforts would have been translated into an automated representation. Because of time constraint (the normal schooling hours of the class), we were only able to



Fig. 5. The story of "the four frogs and the smart dog" living in the same house. The dog, expert on energy consumption, inspires positive behaviours and gives the frogs useful advice, exposing written posters when they forget to switch off unnecessary lights. The miniatures below the main scene represent: the animation of the dog (NiTiNOL powered), the disclosure of the word written on the poster (thermochromic paint with resistive wire), the switchable light (LED) and the curtains revealing the house's interior (thermochromic paint with resistive wire).

set up and play one of the children's stories, after which the children had to go home. This was enough to let them all see the process of building the representation, but prevented them from seeing all the final stories. After the end of the session, we assembled and recorded all the theatrical representations of the stories.

#### G. Session 7: Watching and voting the stories

A week later, the children watched their stories on a large screen. One of the stories (i.e., the story of the four frogs and the smart dog) can be seen in Fig. 5. It shows how the smart origami works. After each representation each child assigned a grade to the story. A questionnaire ended the session.

#### H. Session 8: Evaluating the educational results

After the end of the experience, months later, we returned to check its educational impact.

Parameters and Questions	Smileyom. all (a)	Smileyometer value per task (b)				
		1	2	3	4	5
<b>perceived usability</b> : how easy was it for you to perform the project activities?	4.14	4.2	4.32	3.89	3.79	<b>4.47</b>
<b>felt involvement</b> : how much did you enjoy to perform the project activities?	4.33	<b>4.74</b>	4.63	3.58	4.32	4.37
<b>focused attention</b> : how interesting was for you to perform the project activities?	4.15	4.32	4.42	3.63	3.89	<b>4.47</b>
<b>novelty</b> : how new did you find the project activities?	4.26	3.74	4.79	3.11	4.79	<b>4.89</b>
<b>endurability</b> : how much would you like to perform again the project activities?	4.16	3.84	<b>4.74</b>	3.37	4.11	<b>4.74</b>
<b>aesthetics</b> : how much did you like the stories created by your fellows?	3.81 <i>jury</i>	- not applicable -				

TABLE III

SECOND QUESTIONNAIRE (SEC VII) - SCORES ASSIGNED BY THE CHILDREN TO THE 6 PARAMETERS THAT DEFINE THE ENGAGEMENT. COLUMN (A) REPRESENTS THE OVERALL MEAN SCORES, COLUMN (B) REPRESENTS THE MEAN OF EACH TASK (THE BEST TASK SCORE IS HIGHLIGHTED IN BOLD).

THE LIST OF THE TASKS: (1) MAKE ORIGAMI MODELS; (2) ENHANCE ORIGAMI MODELS WITH SMART PAINTINGS; (3) WRITE THE STORIES; (4) ENHANCE THE STORY WITH SMART MATERIALS; (5) TRANSFORM THE STORIES TO VISUAL PROGRAMS FOR AUTO-PLAY. FOR THE SIXTH PARAMETER (AESTHETICS) THE MEAN SCORE IN COLUMN (B) DERIVES FROM THE VOTES OF THE JURY OF CHILDREN.

Parameter	Mean
Overall interest for SMI	2.94
Int. for NiTiNOL wire (for moving objects)	2.61
Int. for NiTiNOL wire (for changing shapes)	2.50
Int. for thermochromic ink (for changing color)	2.67
Int. for thermochromic ink (for reveal. objects)	2.89

TABLE II

FIRST QUESTIONNAIRE (SEC. VI) - MEAN SCORES ASSIGNED BY THE CHILDREN USING A 3-POINTS SCALE

## V. RESULTS

In the next sections we present all the cumulated results during and after the experience. We tracked the educational process using several methods: direct observation, videos and questionnaires. The post evaluation, held a few months after the end of the experience, also included a set of individual and group tasks to check the educational improvements. We registered high levels of interest during all the phases, in particular for those activities that were perceived as new (i.e., origami models creation, demonstration and the making of SMI, demonstration and the making of visual programs). We measured our qualitative observations with 3 evaluations, after the first SMIs demo, at the end of the experience and a post evaluation some months later.

### VI. 1<sup>ST</sup> QUESTIONNAIRE: APPROACHING SMI

After our SMIs demo (Sec. IV-B), we captured the first reaction with the first questionnaire.

Table II shows the results of a set of closed questions targeted for the interest of the children for the SMI (overall) and for the possibility of experimenting again with the different types of materials. The mean results, measured through a 3-points scale (1=low, 3=high), display high interest for the SMI and especially for the use of thermochromic inks for revealing objects. We asked the children to imagine other uses for the SMIs and about two thirds of them expressed creative ideas beyond the simple extension of the functionalities we showed. Some of the proposals were focused on artistic uses of SMI, such as clocks and color changing shoes with the owner's preference or magical pencils capable of drawing in different colours. Many children proposed stimulating creative

functional uses, such as glowing materials illuminating the path at night, smart books capable of turning their pages, or even super-smart materials capable of self-replicating or doing housework.

### VII. 2<sup>ND</sup> QUESTIONNAIRE: THE EXPERIENCE

The second questionnaire (Sec. IV-G) was composed mainly of closed questions addressed to analyse different facets of the pupils' experience. We analysed the six different parameters that define the user engagement according to O'Brien et al. [1]: *perceived usability*, *felt involvement*, *focused attention*, *aesthetics*, *novelty* and *endurability*. This analysis is useful for measuring all those experiences that go beyond the working activity and whose success is also determined by parameters such as aesthetics or felt involvement. For this questionnaire we used a 5-point scale, with the *Smileyometer* for expressing the numeric values in a more friendly fashion. The Smileyometer [19] [20] takes advantage of pictorial representations (smileys) for eliciting children's opinions. To be sure of the children's comprehension, we distributed the questionnaire in the classroom and read the questions one by one, evidencing the focus of each one and asking the children if they had any doubt about it. Table III shows the engagement parameters, the related list of questions and the resulting mean scores. We explored the children's opinions about the different active tasks that were assigned during the activities of the project. For five of the six parameters that define the engagement, we asked the children their opinion about the following tasks:

- 1) make origami models;
- 2) enhance origami models with smart paintings;
- 3) write the stories;
- 4) enhance the story with smart materials;
- 5) transform the stories to visual programs for auto-play.

Column (a) shows the scores derived from the means of the different tasks, while column (b) displays the results for each task. The tasks are identified by a numerical label, referring to the numbered list that we have just described above in this section. For the aesthetics parameter, there are no analytical results for each task because we used the scores assigned by the children to each play when they saw them represented in the cardboard theatres. Column (a) displays positive results



Block/Structure	Mean
start block	4.53
play sound block	4.21
analog block for activating SM	3.95
wait block	4.32
sequence	4.11
repeat block	4.11

TABLE IV

SECOND QUESTIONNAIRE (SEC. VII)- EASE OF USE OF THE VISUAL BLOCKS AND CONTROL STRUCTURES (5 POINTS SCALE SMILEYOMETER).

for all the parameters that define the concept of engagement. However, the analytic scoring emphasises that the use of smart paintings (2), the story enhancement with smart materials (4) and the transformation of stories into visual programs (5) obtained higher scores for all the parameters. The making of origami models performed well but with slightly lower values, especially for the novelty and the willingness to perform the activity again. Finally the story making activity worked as a glue for the whole experience and gained positive values but lower for what concerned the novelty and the willingness to *write* stories in the future. The questionnaire also included additional analytic questions about the ease of translating the stories into visual programs. The positive answers displayed in Table IV confirm the results of the direct observation of the task execution, that was performed by all the groups nearly in autonomy after the collective demo in the classroom. While the ease of use gained high scores for all the blocks and control scores, the analog block was perceived as slightly less easy to use. This is because the meaning of the *analog* block is not intuitive. It requires one to map the SMI connected to the board with a pin number and another numerical value, the abstract duty cycle for controlling the energy fed to the SMI. A possibility for improving the situation is to change the digital (UI) and physical (Arduino) interface to a more matching and coherent meaning for the parameter choice. For example, Blockly introduced (for example on the online gaming part) a visual way to quantify the number of degrees to choose the proper value with a piechart. A complementary approach would see the Arduino interface to match colours with the S4A interface, simplifying the comprehension of the user on which is connected where.

## VIII. FINAL: LEARNING EVALUATION

After the experiment we agreed with the teachers to come back in the classroom (Sec. IV-H) to check the results of the educational experience in the middle term. We had the opportunity to come back to the school a few months after the educational experience, just before of the end of the school's year.

The evaluation was structured into 4 tasks: two questionnaires, a text analysis task and a visual programming task. We examined several facets of the experience: the awareness for environmental themes, the knowledge about smart materials, the new skills for identifying the narrative structure and translating it into visual programs.

### A. Tasks and goals descriptions

Task 1. The goal of this task was to verify the impact of the children narrations, on their personal environmental awareness after the end of the experience. We asked the children to fill in a questionnaire, a set of 6 open questions. For each question the children had to describe positive behaviours related to one of the thematic areas that were explored during the experience. The questionnaire was filled in in the classroom. The children had to complete the task individually in about 30 minutes.

Task 2. The goal of this task was to understand the level of comprehension of the properties of the different materials. The closed questions (of the second questionnaire) were focused on the properties of the smart materials, met during the educational experience. In particular we asked the children to focus on the factors determining a change of state on the smart materials. The questionnaire was filled in by the children in the classroom in about 30 minutes.

Task 3. The goal of this task was to verify whether the children had acquired the skill of performing a structured text analysis. We gave them a printed short story, asking them to do the same analysis they did for the stories during the experience (Sec. IV-D), splitting the narration in blocks and identifying them with the set of symbols that they used for their stories (Fig. 3). In addition to the symbols available during the experiment, we introduced the *Trigger* symbol, for specifying the start of the animations. Each child completed this task individually in about 30 minutes.

Task 4. The goal in this case was to check if the children had acquired the skill of manipulating the entities of a visual programming language. For this task the groups' organisation was the same used during the experience. The children were asked to map the story blocks, from the previous task, to visual programming entities, as they did during the educational experience (Sec. IV-E). Due to practical constraints the task did not involve the real activation of smart materials. However we encouraged the children to use the proper functional blocks for activating the characters of the story. Due to time constraints we limited this task to the first paragraphs of the narration. The task was preceded by a 30 seconds recap about the categories of components available in Scratch. Because the children had worked individually in task 3, we selected only the most detailed analysis. The children then integrated it with more observations. For the execution of this task, we used a room where each group completed the task in sequence, on a laptop.

### B. Results

The evaluation confirmed the positive role of the experience for teaching the children new knowledge and skills. As stated at the beginning of the section, the time for coming back to the classroom and performing a post test evaluation was determined by the availability of the teachers. The results show that even after a prolonged period of time from the end of the experience (i.e. six months) and no intermediate recap, there was a high degree of retention of the knowledge and of the skills learned during the experience.

Parameter	Initial quest.	Final quest.
Electricity	24/0	<b>41/1</b>
Heating	4/21	<b>5/26</b>
Light	12/9	<b>16/11</b>
Transportation	19/9	<i>17/14</i>
Waste	21/0	<b>29/0</b>
Water	9/23	<b>19/28</b>

TABLE V

LEARNING EVALUATION - LISTING OF POSITIVE BEHAVIOURS RELATED FOR THEMATIC AREAS. THE FIRST NUMBER SHOWS THE NUMBER OF ANSWERS RELATED TO ISSUES EVIDENCED IN THE STORIES, THE SECOND SHOWS THE NUMBER OF COMPLEMENTARY ISSUES (IN **BOLD**, POSITIVE INCREASE OF AWARENESS, IN *curse* THE NEGATIVE).

Task 1. The results show that the creation of narrations focused on environmental issues had a positive impact on the personal awareness. Table V displays the number of positive personal behaviours described by the children for each environmental theme. It compares the results between the questionnaire filled in before the educational experience (Sec. IV-A) and the final post-test questionnaire. For each questionnaire and for each theme the first number displays the number of answers directly related to issues evidenced in the stories, while the second number displays the answers related to complementary issues. In most of the cases the final questionnaire reveals an improvement in the children's awareness (in bold). Only for a single theme (transportation) did the results display a modest reduction of the awareness. We think that this result might be due to the fact that the related story provided an example that was not immediately transferrable to the everyday experience (i.e., the bird flying with its own wings instead of using a jet).

Task 2. The experience with the smart materials gave good results for the acquired knowledge, especially for those where the children had the opportunity to manipulate and not just observe them. The number of correct answers (Table VI) show that the children performed well in identifying how the two smart materials (NiTiNOL wire and thermochromic ink) changed their state and in understanding which were the factors driving the changes. About NiTiNOL wires, most children correctly identified the length changes and related these changes to the electric battery use, but some of them did not remember the associated temperature shift. We might relate this failure to the fact that it was not possible to let the children touch the wires during the demonstration. About the thermochromic ink, most children correctly identified the state changes and the causes that determined them. All the children had the possibility to test directly the influence of the temperature, by touching the painted objects and verifying the effect of the warmth of their hands.

Task 3. The results displayed relevant skills for the text analysis. Table VII shows the number of symbols used for splitting each story into logical blocks, evidencing for each type mean and standard deviation. All the children placed the *Start* symbol correctly at the beginning of the story and the majority of them used the *Go on* symbols correctly to

Topics	Correct answers
Factors changing appearance in the NiTiNOL wire:	
- ambiental noise (no)	17/18
- wire temperature (yes)	6/18
- room illumination (no)	17/18
- electricity (yes)	17/18
- proximity (no)	11/18
The NiTiNOL wire changed its appearance:	
- changing its length (yes)	16/18
- changing its visibility (no)	17/18
- changing its color (no)	14/18
Factors changing the thermochromic ink:	
- wind (no)	18/18
- ink temperature (yes)	16/18
- room illumination (no)	16/18
- electricity (yes)	15/18
- ambiental noise (no)	18/18
The thermochromic ink changed its appearance:	
- detaching itself from the sheet of paper (no)	18/18
- shrinking the underlying sheet of paper (no)	18/18
- changing its visibility (yes)	18/18

TABLE VI

LEARNING EVALUATION - PROPERTIES OF SMART MATERIALS

Block	Mean	St. dev.
start symbol	1.0	0.0
go on symbol	8.3	4.9
trigger symbol	4.9	2.3
repeat symbol	3.1	0.8
stop symbol	7.4	4.5

TABLE VII

LEARNING EVALUATION - STORY ANALYSIS: USE OF SYMBOLS FOR GIVING A STRUCTURE TO THE WHOLE STORY

split the text of the story into fragments. Only in 5 of the 19 the children used a very low number of *Go on* symbols (from 0 to 3 symbols). Most children identified the situations of the story that could be mapped to a cycle (e.g., repeated dog barks, noises and actions of animals involved in the story). The children appreciated the possibility to specify the animation of the characters with the *Trigger* symbol, recalling the experiments done with the smart origami. Most of them specified a high number of animations (mean 4.9, standard deviation 2.3, Table VII), related to the appearance of a character in the scene. The children also used the *Stop* symbol in different fashions. Some of them interpreted it as an entity that required an explicit restart and therefore placed a *Go on* block after each use of the *Stop* symbol, while others considered it as a temporary stop that did not require an explicit restart. In most cases we identified a precise logic underlying the association of the symbols to the text. Only in some of the cases (4 out of 19) were symbols not placed in a coherent fashion. The majority of the children did a good job for the text analysis.

Task 4. The results show that all the groups succeeded in creating a Scratch program of similar complexity (Table VIII). Most groups did not have problems for mapping the Scratch components to the analysis. A single group, composed of

Group	time	start	play	repeat	wait	analog
electricity	19.0	1	4	1	2	1
heating	16.0	1	5	1	0	1
light	14.0	1	5	1	2	1
transportation	16.0	1	5	1	1	1
waste	21.0	1	4	1	1	1
water	18.0	1	4	1	0	1

TABLE VIII  
LEARNING EVALUATION - TIME AND USE OF PROGRAMMING ENTITIES  
FOR TRANSLATING THE STORY INTO A VISUAL PROGRAM

children that did not give a detailed description of the structure in the previous task, needed additional support for improving the text analysis and the mapping work. The children had no problem in using the Scratch, even for the most sophisticated (i.e., the cycle that requires nested components), but still they needed time to adapt to the interface and mouse. They needed only simple verbal support for accomplishing complementary operational sequences, such as the creation and the use of audio tracks. All the groups succeeded in completing the visual programming activity, in 16 to 21 minutes (Table VIII).

## IX. CONCLUSION

The final evaluation demonstrated that the design of the educational experience was successful in many respects. The children acquired new knowledge in relation to new technological topics, such as the properties of smart materials, and acquired new skills for programming interfaces based on them. We noticed improvements in all the areas involved by the multidisciplinary experience, from the increase of the environmental awareness to the skills for the text analysis. At the end of the educational process the result of the children's efforts was both a working mechanism and a cultural artefact that was evaluated even for its aesthetic qualities.

The results of the direct observation and the questionnaires show that the children learned new concepts, acquired different skills and were engaged both in cognitive and emotional terms throughout the experience. The children learned new methods of expression, they were very interested in origami and visual programming, and declared their willingness to try again the different facets of the whole multidisciplinary experience. About storytelling: we had a confirmation of its positive role for educational paths. This goes beyond the simple teaching of literacy skills and their use for connecting different educational topics and techniques. Storytelling itself received a positive boost from the definition of the innovative educational path. As can be seen from the questionnaires, the children were not very interested in creating stories in the traditional fashion, but they were happy to create them with the SMIs. The fact that the stories were played on a screen instead of the physical theatre, had probably some minor influence on the evaluation, but we expect that this would be worsening the results and not improving them. The results are instead all very encouraging. We designed this path focusing on the experience of creation.

With this S4A experimentation, we followed the tradition of Logo and visual programming, but added the role of children

as makers. As can be seen in Table III column (b), rightmost task (5), transforming the stories into visual programs for the play, 4 out of 5 parameters gained the maximum scores among the other tasks. The introduction of a visual programming paradigm brought the possibility of automatically controlling the materials' transformation. In this educational experience, the shift from smart materials to SMIs allowed the move from the simple knowledge to the experimental activation of these materials. It is important to underline that the tools that were introduced for stimulating the interest for the smart materials, worked not only as a means but they were a focus of interest themselves. Reciprocally, we can observe that smart materials were not only the main focus of this experiment, but also a useful means for explaining technology to children.

In conclusion, all the results and the children's responses indicate that the experience was perceived as a positive and interesting activity. This shows how innovative research topics, such as SMIs, can be integrated into a pedagogical path for primary schools, merging the traditional learning and other techniques (assisted by suitable visual programming tools and physical technologies). We believe that this kind of application should be further explored and that the experience presented can be an interesting future path to look forward to.

## ACKNOWLEDGMENTS

We would like to thank all the people that supported this project: Matteo Fumagalli and Raffaella Carloni from the RAM Group of University of Twente, Lorenzo Moroni from MERLN Institute for Technology-Inspired Regenerative Medicine Maastricht University, the teachers Patrizia and Tecla Pasqualon from the primary school and a special thanks goes to Roberta Saccilotto and Giuliana Bordin. This publication was supported by a CTIT-ITC collaboration project.

## REFERENCES

- [1] 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.
- [2] M. G. Helander, T. K. Landauer, and P. V. Prabhu, Eds., *Handbook of Human-Computer Interaction*, 2nd ed. New York, NY, USA: Elsevier Science Inc., 1997.
- [3] E. Sun and S. Han, "Fun with bananas: Novel inputs on enjoyment and task performance," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2013, pp. 1275–1280.
- [4] B. M. Collective and D. Shaw, "Makekey makekey: Improvising tangible and nature-based user interfaces," in *Proc. of TEI '12*. New York, NY, USA: ACM, 2012, pp. 367–370.
- [5] R. Davis, Y. Kafai, V. Vasudevan, and E. Lee, "The education arcade: Crafting, remixing, and playing with controllers for scratch games," in *Proc. of IDC '13*. New York, NY, USA: ACM, 2013, pp. 439–442.
- [6] "Blockly," <https://developers.google.com/blockly/>, [Online; acc. July-2015].
- [7] A. Druin, Ed., *The Design of Children's Technology*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998.
- [8] F. Garzotto and M. Forfori, "Fate2: Storytelling edutainment experiences in 2d and 3d collaborative spaces," in *Proc. of IDC '06*. New York, NY, USA: ACM, 2006, pp. 113–116.
- [9] "Alice," <http://www.alice.org>, [Online; acc. May-2014].
- [10] "Looking glass," <http://lookingglass.wustl.edu>, [Online; acc. May-2014].
- [11] S. Jacoby and L. Buechley, "Drawing the electric: Storytelling with conductive ink," in *Proc. of IDC '13*. New York, NY, USA: ACM, 2013, pp. 265–268.



- [12] M. Bodén, A. Dekker, S. Viller, and B. Matthews, "Augmenting play and learning in the primary classroom," in *Proc. of IDC '13*. New York, NY, USA: ACM, 2013, pp. 228–236.
- [13] E. Y.-L. Do and M. D. Gross, "Environments for creativity: A lab for making things," in *Proc. of C&C '07*. New York, NY, USA: ACM, 2007, pp. 27–36.
- [14] M. Coelho, L. Hall, J. Berzowska, and P. Maes, "Pulp-based computing: a framework for building computers out of paper," in *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2009, pp. 3527–3528.
- [15] L. Berglin, "Spookies: Combining smart materials and information technology in an interactive toy," in *Proc. of IDC '05*. New York, NY, USA: ACM, 2005, pp. 17–23.
- [16] A. Minuto and A. Nijholt, "Smart material interfaces as a methodology for interaction: A survey of smis' state of the art and development," in *Proc. of SMI '13*. New York, NY, USA: ACM, 2013, pp. 1–6.
- [17] A. Minuto, D. Vyas, W. Poelman, and A. Nijholt, "Smart material interfaces: A vision," in *Proc. of INTETAIN 2011*. Springer, 2011, pp. 57–62.
- [18] A. Minuto and A. Nijholt, "Growing grass: A smart material interactive display, design and construction history," in *Proc. of SMI '12*. New York, NY, USA: ACM, 2012, pp. 7:1–7:5.
- [19] J. Read and S. Macfarlane, "Endurability, engagement and expectations: Measuring children's fun," in *Interaction Design and Children*, Shaker Publishing. Shaker Publishing, 2002, pp. 1–23.
- [20] J. C. Read, "Validating the fun toolkit: An instrument for measuring children's opinions of technology," *Cogn. Technol. Work*, vol. 10, no. 2, pp. 119–128, Mar. 2008.

# Semantic video annotation for accessible resources in flipped classrooms

Ilaria Torre

Department of Computer Science, Bioengineering,  
Robotics and Systems Engineering  
University of Genoa, Italy  
ilaria.torre@unige.it

Gianni Vercelli

Department of Computer Science, Bioengineering,  
Robotics and Systems Engineering  
University of Genoa, Italy  
gianni.vercelli@unige.it

**Abstract** — In this paper we present a project for applying a framework for accessibility to educational activities in flipped classrooms by exploiting semantic video annotation. The framework exploits semantic and adaptive technologies and is aimed to support accessibility to physical real-world things.

**Keywords:** Internet of Things, accessibility, semantic video annotation, user adapted interaction, technology-enhanced learning, Liked Data.

## I. INTRODUCTION

The advent of the Internet of Things (IoT) and the Web of Things (WoT) has opened new opportunities for people with special needs [1]. The Perception Layer of the IoT is able to identify objects and gather information from the environment, the Network layer transmits information obtained from the Perception layer and the Application layer is a set of services that take as input the data gathered from the Perception layer to satisfy the needs of the users.

In this paper we present a framework for accessibility in the WoT and a project for applying it within educational environments. The proposal is inspired on Linked Data principles, and uses Semantic Video Annotation to support learners with special needs in flipped classrooms.

Recent literature on flipped or inverted classrooms propose to use/realize videos and to record lectures so students can view them out of class when they prefer. This asynchronous approach frees up in class time for learning activities [2], including exercises, laboratory experiments, document analysis or speech presentation. Video annotation in online and flipped classroom is proposed as a mean to improve learner engagement [3, 4], critical reflection [5] and learning performance [6]. Moreover it has been used to support visually impaired people [7]. In this paper, we describe how the WoT and the Semantic Web technologies can be fruitfully adopted – together with video annotation tools – to support students with special needs.

The first part of the paper (Section 2) describes the framework for accessibility, while the second part (Section 3) presents the use of semantic video annotation in flipped classrooms.

## II. ACCESSIBILITY FRAMEWORK

### A. Goals and approach

Accessibility is the term used to indicate whether an object can be used by people of all abilities and disabilities [8]. It is a multi-faceted concept since accessibility may concern real world things and environments as well as web pages, software applications and ICT devices [9]. Accessibility may concern physical and cognitive disability, but also may include logical barriers. For example, an object is not accessible for a user that does not know how to use it but will become accessible after instructing her; or an application form that is not accessible to a user who does not comprehend its language, will become accessible if a proper translation is provided.

The WoT enriches everyday physical things by linking them to their digital counterpart using HTTP standards. Thus, we obtain augmented cyber-physical things [10, 11] that can be accessed and used in different ways, by *exploiting their digital or their physical side*. As a consequence, even if physical objects are not natively accessible, they can become accessible by means of a software layer, or an application, that adapts access and methods of interaction with them. While physical objects cannot be made accessible for everyone, virtual objects connected to physical objects can [12]. For instance, a physical dictionary or a calculator can be not accessible to visually impaired students. However, the accessibility to these real-world objects can be achieved by equipping the dictionary or calculator with a software layer, such as a smartphone application, that provides the information using the adequate modality: audio in case of blindness or simply larger text in case of impaired sight. Hence, if the application is accessible, the dictionary or calculator will become accessible as well. Similarly, if a scene in the environment cannot be perceived by a visually impaired person, a video recording annotated with scene descriptions may enable the subject to understand it.

Notice that the vast consumer electronics market is already filled with smart devices that can be accessed and controlled remotely via apps (smart fridges, thermostats, heart rate monitors, etc.). The approach that we present here, goes in this direction, but enhances accessibility as its main objective and exploits semantic and adaptation technologies, as discussed in

related works [13, 14, 15, 16]. The accessibility framework is funded on three main blocks that can be seen as *actions*:

1. *Enriching physical objects* with semantic annotations (the application scenario that will be presented in the following is focused on semantic video annotations);
2. *Matching data* about the real-world object features against user needs, preferences and current environment (this may require to catch, profile and annotate the user and the usage context);
3. *Exploiting adaptive techniques* in order to adapt the web-based counterpart of the physical object to make it accessible to different kinds of users in different conditions.

We use Linked Data (LD) as semantic annotation paradigm because of its suitability to foster the integration of heterogeneous data and their connection, sharing and reuse. This is a great improvement with respect to current smart IoT devices in the market since it allows to exploit the available data on the Web about the physical object, the user needs/preferences and especially it allows to share growing knowledge about accessibility requirements. Thus, the framework combines the new approaches based on WoT and LD and exploits adaptation techniques to adapt the interaction of the virtual side of the augmented physical object to make the physical object accessible.

## B. Background

The integration of computational and physical elements, especially when complemented with intelligent mechanisms, has broadened the potential of cyber-physical systems in several areas. WoT cyber-physical objects have been developed to cope with problems including intervention, coordination and augmentation of human capabilities (e.g., healthcare monitoring and delivery). WoT cyber-physical objects can be physical objects with embedded sensors and processing capabilities, but they can also be everyday artifacts (e.g. books, goods, shop shelves, desks) with attached tokens linked to a virtual counterpart on the Web; users access this virtual counterpart by scanning the attached token (e.g., QR code, RFID tags) and by getting information about its location (URI). Smart objects can also embed tiny Web servers which make it possible to communicate to such objects using HTTP standards and also to invoke services provided by these objects.

While previous research studies dealing with accessibility and ICT were most focused on making devices, software and platforms accessible, current projects exploit ICT to offer augmented services. Several FP7 European Projects have worked on providing frameworks and ontologies for accessibility – e.g., Open Accessibility Everywhere ([www.aegis-project.eu](http://www.aegis-project.eu)), Accessibility Assessment Simulation Environment ([www.accessible-eu.org](http://www.accessible-eu.org)), OASIS ([oasis-project.eu](http://oasis-project.eu)) and Cloud 4all ([cloud4all.info](http://cloud4all.info)).

Our approach follows a similar direction, but slightly differs from them since it is specifically focused on using ICT to make current physical objects accessible, while the mentioned projects are more focused on enhancing

accessibility of digital devices and web applications. Thus our model can exploit the results of these projects (including the ontologies they have defined, in particular ACCESSIBLE and AEGIS) but emphasizes the use of WoT and the use of LD to connect heterogeneous data about accessibility features.

Several applications have been designed to solve accessibility problems. The latest approaches, methods and tools are collected in [17]. However, the real problem when dealing with assistive technologies is that impairments are heterogeneous and often a subject has more disabilities together. Also limiting the analysis to people, and in particular students, with visual impairments, researchers describe the complexity of exploiting assistive technologies since the combination of different impairments and the different levels of impairment influence the way each technological support is experienced [18].

In this scenario, the possibility of dynamically identifying the kind of disability and adapting the virtual side of the physical object becomes a critical challenge. Adaptation is the core of the GPII ([gpii.net](http://gpii.net)) project for a global infrastructure for accessibility and the focus of a semantic framework for assistive technology within the Cloud4all FP7 project [19]. This framework is designed to support user interface adaptation to different assistive technologies and configurations. However, also these projects are focused to make digital devices more accessible, while our objective is to exploit WoT and semantic adaptive technologies to make physical objects more accessible.

## C. WoT object annotation

Building a digital representation of a physical object is the first step to virtualize it. This process associates each object to its digital representation, described in [20] as Digital Object Memory (DOM), which can be passive or active and can be used for different purposes, ranging from storing temporary data obtained by sensors to storing and representing complex information. Following LD principles, objects are identified with a URI and this URI can be associated to other resources about the object. For example, Al-Khalifa and Hend [21] associate physical objects to their audio description and tag the object with a QR code that contains the URI of its audio description on the Web. This simple cyber-physical object may be used to face accessibility problems of physical objects. It can support visually impaired and blind people to identify objects in the environment.

Our approach uses the architecture for object annotations described in [13]. It is a three-layered architecture consisting of a *physical* layer, the layer of the *digital memory* DOM, and a *Linked Data* layer (see the grey layered boxes in Fig. 1).

(i) The *physical layer* includes: the physical object, the specific modality of interaction with the object, such as pointing, scanning, touching or using a mediation device (e.g., a smartphone), and the modality of identification of the object, such as RFID, QR-Code, Semacode and techniques for visual object recognition.

(ii) The layer of the *digital memory* DOM contains the description of the object and the way to access it, according to

the Object Memory Model (OMM) [20]. Basically it is a repository of digital data that is linked through URIs with a physical artefact.

(iii) The *Linked Data* layer describes the physical object according to the LD principles and should preferably link the related dataset on the Linked Open Data Cloud (LOD). This layer enables the object to link other objects and to be linked by related objects in the Web of Data (WoD), thus offering the possibility to be shared, extended and reused.

A layered architecture for annotations enables to collect all data concerning an object within a unique logical repository and to expose data about it in a flexible fashion. This architecture is made possible by the new scenario of augmented physical objects, where for example, a Linked Data wrapper can be in charge to publish just a subset of the information stored in the object memory of the object. Ding et al. propose the use of LD to link the user preference and subsets of real-time environment data [22].

The adoption of LD provides several advantages: 1) LD practices are designed to foster the possibility of integrating heterogeneous data and reuse them in different ways [17], 2) the LOD contains increasingly mass information that will complement and enrich the smart object annotation, 3) the object description could be enriched by the users, by creating new triple and thus adding knew knowledge that other users may find precious; and 4) Semantic Web technologies allow for a number of reasoning mechanisms that can sustain the adaptation process. Most of this process can be handled by the state of the art OWL ontology reasoners [23].

Based on the framework, a physical object, such as a laboratory tool, can be provided as a smart IoT tool accessible at different levels. A simple and easy-to-be deployed level is providing the instructions according to the framework principles: text instructions (physical layer) are labeled with a QR code or an RF-ID tag associated to an online DOM which contains several kinds of resources, fitting different types of needs. The LD description allows to link this data to related data on LOD: given that the model of this tool may have features that make it similar to other tools, this connection may allow that all objects of a certain model of the tool may have the same instructions, recommendation and adaptation strategies to be downloaded from the web.

It is worth noting that people, as well as artefacts, being real-world “objects”, can have their own digital representation, as showed in Fig. 1 (Person). The layered representation of the user is particularly relevant to our approach, since it enables to store knowledge about the user features and needs concerning accessibility at different levels of visibility, and this data can be used to adapt the interaction with the object (described in the next section). Of course, a vital requirement is the respect of the end-users’ privacy when accessing their data. Data about accessibility may be extremely sensitive. Hence, if the adaptation module resides in the object, the user should be allowed to share just the minimal part of her/his profile that is necessary for optimizing the interaction. If the adaptation module resides on the user side (e.g., on personal mobile device) these problems are easier to address, since there is no need to directly inform the smart object about the person’s

needs or disabilities. Privacy requirements depend on a variety of contextual socio-cultural factors and can be analyzed using privacy requirements distillation approach, such as [24].

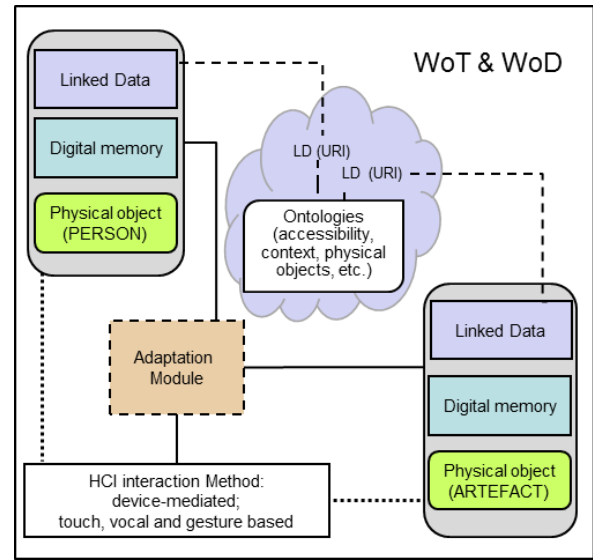


Figure 1. Interaction model of the accessibility framework.

Specific applications implementing the architecture can define how and where to store the digital representation of the objects: locally in the object, or on a server, which can be accessed via HTTP. In such a scenario, virtual representations of objects can serve as central hubs of object information [25] that may combine and continuously update data from a wide range of sources. Part of them can be slowly changing data (e.g. descriptive features of the object), other data may come from sensors and be continuously updated.

Ma [26] proposes a classification that emphasizes this point, proposing a four-layered architecture, including: object sensing layer, data exchange layer, information integration layer, and application service layer. In all the models, the basic idea is that the digital representation of physical objects should harmonize the access to the heterogeneous set of underlying objects with a common language and procedures. This enables applications to get the information they need about the object, based on their specific purpose and users.

#### D. Interaction with annotated objects

Fig. 1 displays the model for the interaction with the augmented physical objects. The object (Artefact) and the user (Person) who wants to access it are represented according to the layered architecture described above. The Linked Data Layer is associated (through a dotted arrow) to the URI in the Linked Data Cloud. This Cloud includes the ontologies that enable the semantic description of the artefacts and of the user features. They provide the vocabularies to represent the RDF statements about the real-world things and accessibility requirements.

Notice that, even though user’s data are annotated according to LD principles, they do not necessarily have to be open to public access. Their access can be managed using different policies, depending on the specific implementation of

the model. Furthermore, the object descriptions can be enriched by the users themselves or supervisors by producing new triples that may be useful to address the need of other users. A number of semantic annotation tools can be used by common users to annotate items (e.g. most CMS include semantic annotation support). This can be quite useful to address the accessibility problem. For example, a user with certain impairment may leave advices of how to best approach a certain item for users with similar disabilities. In the same way, a user speaking in a language that is not supported by the object may decide to help subsequent users by providing a translation of some useful information. Of course, information inserted by users should be handled carefully, since they may include incorrect information. Different policies may thus be implemented to address the trust problem of the different sources. Past experiences, such as semantic wikis [27, 28], proved the potentiality of allowing common users to build their own knowledge.

The bottom left side of the figure shows that the interaction with the object can be mediated by another device, such as a smartphone, or can exploit natural interaction modalities. Mediated access typically uses native applications that get the location of the virtual counterpart of the physical object and access them. Typically, they get the URI of the object and access the Web server at the specified URI. As mentioned above, the server can be embedded within the physical smart object, or on the Web. Natural interaction modalities require specific equipment of the cyber-physical object. For example, the provision of speech modality requires the adoption of ASR (Automatic Speech Recognition) technology and may further implement NLP (Natural Language Processing) techniques; the provision of tactile modality of interaction requires the adoption of multi-touch surface technology. E.g., Microsoft Surface is a development platform that enables to create applications that hide the computer logic below the surface and allow users to interact with a high-end graphics display similar to a coffee table. Thus, users get the service (provided by an application) by using only their fingers.

Finally, Fig. 1 shows the Adaptation Module (AM) that is in charge of adapting the user interface and the interaction modality of the virtual side of the physical objects. It is defined according to the definition of adaptive software [29] and of context-aware adaptive system [30], namely a type of specialized software that uses information from the environment (user needs and context features) to improve its behavior over time. Its function, here, is to use the available information about the object (Artefact) and about the user and its environment (Person) to make the object accessible.

Different configurations are possible for the AM: inside the physical objects, as in distributed models, or externally as in centralized models (on the user's mediation agent or on a web server). A vast literature on adaptive and user modeling systems has been produced in the past [6] about advantages and limits of client side, server side and distributed solutions, however the new scenario of a sensorized society demands for new models.

In general, independently of the architecture, the AM should be in charge of collecting accessibility requirements

from different sources, matching them with configuration options of the user interface and user capabilities and adapting the interface accordingly.

In a distributed model, the adaptation is carried out by each physical object, while in a centralized model, a unique module should manage the adaptation for different objects, acting as an agent of the user. In the latter case, given the semantic representation of the user requirements, for example hearing troubles, it should coordinate the different objects so that all of them convert audio notifications into a common haptic feedback, such as a vibration, modulated according to a shared scale of intensity to signify the type of notification, or use tickers or subtitles on displays.

### III. SEMANTIC VIDEO ANNOTATION

In this section we present a prototypical demonstrator that implements the semantic annotation component described in the framework. The idea is to exploit video annotation in a flipped classroom to support three kinds of impaired students: visually impaired students, hearing impaired students and students affected by learning disorders.

We have identified these types of impaired students as target categories, however it is important to underline that the current demonstrator is not focused on adaptation for different kinds and levels of disability. This will be done in the next step of the design. The current stage just implements the architecture of the framework, that will be used as the basic infrastructure to develop the adaptation service.

This modular approach is made possible by the use of explicit representation of knowledge concerning each component of the framework: the learner model, the real-world objects and the adaptation techniques.

As we mentioned in the Introduction, flipped classroom concerns a teaching paradigm where traditional in-class lessons are replaced by video recorded lectures and in-class activities concern practice exercises, laboratory experiments, document analysis, debate or speech presentation.

Our **objective** is to exploit the accessibility framework to support the three categories of impaired students mentioned above, both in online activity and in-class activity.

The **approach** is the following:

- improving the accessibility of online video lectures by exploiting semantic annotations and adapting them to the type of impairment,
- improving the accessibility of in-class activity by:
  - ✓ video recording in-class activities
  - ✓ annotating them with semantic annotations and adapting them to the type of impairment.

Video lecturers are digital objects therefore they can implement only the second and third layer of the layered architecture for object annotation. Differently, in-class activities are real-world scenes made of real-world things. Therefore we are able to implement all of the theoretical layers of the accessibility framework.

In order to manage the semantic annotation we exploit Apache Marmotta platform.

#### A. Platform for video annotation

The current implementation is the integration of a responsive interface and an annotation tool, both realized using HTML5-CSS3-JS technologies like the Foundation framework from Zurb<sup>1</sup>, built on top of a Linked Data Endpoint service using Apache Marmotta 3.3<sup>2</sup>.

Apache Marmotta is a top level project of Apache Software Foundation, and provides an open source Linked Data Platform for the interlinking of data repositories with the *Web of Data* according to the principles of the World Wide Web Consortium (W3C). It has been developed as a continuation and simplification of the Linked Media Framework (LMF) project<sup>3</sup>. With LMF framework, the research group in Salzburg extended the LD principles to multimedia content (videos, photos, graphics, etc.), while at the same time realizing W3C's vision of the Read-Write-Web. Marmotta is highly modular and extensible to build custom LD applications, like our prototype. Its core components are the Linked Data Server with SPARQL 1.1. and the LD Cache.

A useful option in Marmotta is the extension of SPARQL with specific multimedia functions and relations (e.g., *rightBeside*, *spatialOverlaps*, *after*, etc. ).

For the annotation of video contents, we followed the principles in [30]. Video resources are URI, and locally we add annotations (tags) to denote video *fragments* (parts of the original URI) creating RDF triples which encode temporal and spatial subsequences building semantic relationships between an object/human/artefact and the video fragment.

The semantic annotation tool, displayed in Fig. 2, allows teachers and authors of educational resources to search videos on large social repositories (Youtube and Vimeo), as well as on social networks (Facebook), and connecting social tagging with formal/informal classifications (Wikipedia, WikiCommons) by means of Linked Data. The semantic Linked Data Endpoint is managed as an instance of Apache Marmotta.

#### B. Layered semantic video annotation

In this section we analyze in-class activities given that they fully fit the scope and requirements of the accessibility framework. The goal is to make “in-class scenes” understandable to students with different types of disability.

Based on the accessibility framework, the first *action* to be performed is *enriching physical objects with semantic annotations*. In this instantiation of the framework, the real-world objects are the in-class activities. This represents a very complex object since an in-class activity is composed of nested real-world objects that have to be managed independently. According to the layered architecture for object annotations, the *digital memory* of the object (DOM) is a repository of digital data that is linked with a physical object, and may be

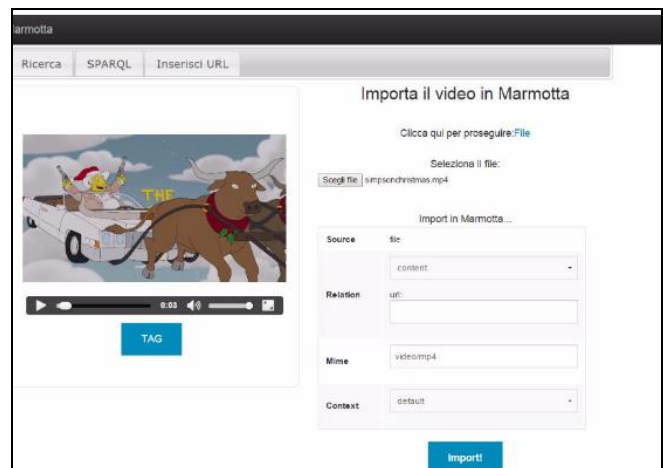


Figure 2. The semantic annotation tool interface for importing a video.

populated with static or dynamic data from entities that interact virtually or physically with the artefact.

Since we have scenes composed of things, we need first to create a DOM of each object. Video recording and pictures of the object will form the set of materials associated to the virtual representation of the physical object.

The approach we adopted is described below:

- to record the scenes of the in-class activity,
- to capture the real-world things in the class (in order to populate a database for pattern recognition and annotation); this includes capturing both animated things (persons) and unanimated things (artefacts),
- to identify automatically or semi-automatically things in the scenes, creating and annotating fragments (on the timeline ),
- manually and collaboratively annotating actions concerning the in-class activity.

According to the architecture for object annotation, the third layer is the Linked Data layer, defined as a subset of DOM. Each object and video is annotated using LD and stored in Marmotta as RDF Triples. In the previous sections we extensively discussed the flexibility of a layered architecture and the reasons for using the LD approach.

In order to enhance interoperability and interlinking in the LOD, it is useful to annotate objects by using popular and effective ontologies such as the following ones:

- FOAF<sup>4</sup> is an ontology for person annotation;
- ACCESSIBLE<sup>5</sup> and AEGIS<sup>6</sup> are ontologies for accessibility related features (their integration provides a multilayer ontology which includes standards, guidelines, techniques and the description of device features, functional limitations of users with disabilities and impairments. Moreover ACCESSIBLE includes verification rules for describing accessibility requirements and constraints);

<sup>1</sup> <http://foundation.zurb.com/>

<sup>2</sup> <http://marmotta.apache.org>

<sup>3</sup> <http://www.w3.org/2001/sw/wiki/LMF>

<sup>4</sup> <http://www.foaf-project.org/>

<sup>5</sup> [http://160.40.50.89/Accessible\\_Ontology](http://160.40.50.89/Accessible_Ontology)

<sup>6</sup> [http://160.40.50.89/AEGIS\\_Ontology](http://160.40.50.89/AEGIS_Ontology)



- Ontology for Media Resources<sup>7</sup> is both a core vocabulary (a set of properties describing media resources) and its mapping to a set of metadata formats currently describing media resources published on the Web. It is a W3C Recommendation.
- Other domain-specific or multi-domain ontologies may be necessary to annotate things. In our demonstrator we have exploited DBpedia to add tags to objects in recorded in-class scenes and also to add tags to objects and persons in video lectures (see Fig. 3).



Figure 3. The tool interface for adding a person linked to a DBpedia resource.

Depending on the kind of disability, *different types of annotations have to be included*. For visually impaired students, audio recording has to be synchronized with precise descriptions of the objects and of the actions. For hearing impaired students, annotations have to transcript audio recording, managing overlapping voices and noise. Finally, for students with learning disorders it could be necessary to provide a different type of annotation, which enables students to understand the scene. For example by using analogy-based methods or by retrieving linked concepts<sup>8</sup>.

To manage the last case, we have performed some experimental tests by retrieving DBpedia data linked to the objects in the scenes stored in Marmotta. The result is that the annotation is driven by some recommendations that are automatically generated and saved in the LD Cache of Marmotta.

Thus, new knowledge is created from both pattern recognition on captured video fragments (keyframe and objects identification) and knowledge retrieval from the LOD cloud. The former is referred to as *intensional knowledge* and the latter as *extensional knowledge*.

Notice that the tool for semantic video annotation is part of the accessibility framework but it can be usefully exploited with non-impaired students as well. In fact, despite a huge amount of recorded material is available in video repositories, a detailed description of their contents is lacking. Such videos could include meaningful notions and show examples as well as standard situations, use cases or case histories. To make

them searchable and usable, the relevant frames should be identified and tagged. The platform for video annotation that we presented above is an easy-to use and powerful tool that can be used by experts but also by students as we experimentally tested in our setup.

About this, it is worth observing that a way to include students with different needs is to provide them a role which enhances their abilities [32]. Collaborative tagging could be an effective instructional activity that follows the principles of knowledge building and constructivism, but could be also an effective strategy to produce different types of annotations for students with different needs.

#### IV. CONCLUSION

In this paper we have presented a framework for exploiting Web of Things, semantic annotation and adaptation techniques to support impaired people to access real-world things. Subsequently, based on this framework, we have discussed the implementation of a video annotation set up that is the basic infrastructure for increasing accessibility to in-class activities and online lectures. The prototypical implementation presented in the paper addresses the knowledge base requirements for video annotation.

The contribution of the paper is twofold: (i) from a theoretical point of view it presents a *general-purpose innovative model* for enhancing the accessibility to things that are in the physical world (e.g., books, calculators, laboratory tools, etc.) or that happen in the physical world (e.g., in-class activities), showing the potential power of LD-based annotation; (ii) from an educational point of view, the approach of *video recording and semantic annotating* in-class activities fits the scenario of flipped classrooms, and proposes a novel approach to face accessibility problems of impaired students.

As a future work a deep study on sets of disabilities and of associated requirements and adaptation techniques is planned.

#### ACKNOWLEDGMENTS

This research has been funded by the University of Genoa, within PRA 2013 projects, Prot. N. 9563.

#### REFERENCES

1. Domingo, M., C., 2012. An overview of the Internet of Things for people with disabilities, *Journal of Network and Computer Applications*, Vol. 35, pp. 584–596.
2. O'Flaherty, J., Phillips, P., 2015 The use of flipped classrooms in higher education: A scoping review, *The Internet and Higher Education*, Vol. 25, April 2015, pp. 85-95.
3. Aubert, O. and Jaeger, J. Annotating Video with Open Educational Resources in a Flipped Classroom Scenario. *CoRR* abs/1412.1780, 2014.
4. Guo, P. J., Kim, J., & Rubin, R., How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the first ACM conference on Learning@ scale conference*, 2014, pp. 41-50.

<sup>7</sup> <http://www.w3.org/TR/mediaont-10/>

<sup>8</sup> Several researches on learning disorders are available such as [32]. Our demonstrator is inspired to this work but does not implement its methods strictly since our current objective is to set up the knowledge base infrastructure and investigate the validity of the approach.

5. Risko, E. F., T. Foulsham, S. Dawson, A. Kingstone, The Collaborative Lecture Annotation System (CLAS): A New TOOL for Distributed Learning", *IEEE Transactions on Learning Technologies*, vol.6, no. 1, 2013 pp. 4-13.
6. Pardo, A., Mirriahi, N., Dawson, S., Zhao, Y., Zhao, A., & Gašević, D. Identifying learning strategies associated with active use of video annotation software. In *Proc. of the Conference on Learning Analytics And Knowledge*, 2015, pp. 255-259.
7. Encelle, B., Ollagnier-Beldame, M., Pouchot, S., & Prié, Y. Annotation-based video enrichment for blind people: A pilot study on the use of earcons and speech synthesis. In *Proc. of the ACM Conference on Computers and accessibility*, 2011, pp.123-130.
8. Clarkson, P.J., Coleman, R., Keates, S., Lebbon, C. *Inclusive design: Design for the whole population*. Springer Science & Business Media, 2003.
9. Stephanidis, C. and Antona, M. (eds), *Universal Access in Human-Computer Interaction*. LNCS, Vol. 8513, 2014.
10. Fortino, G., Russo, W., Rovella, A., Savaglio, C. On the Classification of Cyberphysical Smart Objects in Internet of Things. *Proc. of the Int. Workshop on Networks of Cooperating Objects for Smart Cities (UBICITEC)*. Vol. 1156, 2014, pp. 76-84.
11. Lee, A. E. *Cyber Physical Systems: Design Challenges*. In *Proc. of the 11th IEEE Symposium on Object Oriented Real-Time Distributed Computing* IEEE Computer Society, Washington, DC, USA, 2008, pp. 363-369.
12. Featherstone, D. *An Inclusive Internet of Things: Accessibility in the Palm of Your Hand*, available at <http://simplyaccessible.com/article/things/>, accessed 18 April, 2014.
13. Torre, I. Interaction with Linked Digital Memories. *Proc. of the Int. Workshop on Personalized Access to Cultural Heritage within UMAP*, Rome, July 14, *CEUR Proc*, vol. 997, 2013, pp. 80-87.
14. Torre, I. Virtualization of Objects and Adaptive Interaction in an Inclusive Web of Things, *Proc. of the Int. Conference on Interfaces and Human Computer Interaction*, pp. 15 – 17 July 2014, Lisbon, Portugal.
15. Coccoli, M., Torre, I., Interacting with annotated objects in a semantic web of things application, *Journal of Visual Languages and Computing*, Volume 25, Issue 6, 2014, pp. 1012-1020.
16. Adorni, G., Coccoli, M., Torre, I. *Journal of E-Learning and Knowledge Society*, Volume 8, Issue 2, May 2012, pp. 23-32.
17. Heath, T. and Bizer, C., *Linked Data: Evolving the Web into a Global Data Space*. *Lectures on the Semantic Web*, Morgan & Claypool, 2011, pp. 1-136.
18. Zhou, L., Parker, A. T., Smith, D. W., & Griffin-Shirley, N. *Assistive technology for students with visual impairments: Challenges and needs in teachers' preparation programs and practice*. *Journal of Visual Impairment & Blindness*, 105(4), 2011, pp. 197-210.
19. Kaklanis, N., Votis, K., Giannoutakis, K., and Tzovaras, D. A Semantic Framework for Assistive Technologies Description to Strengthen UI Adaptation, *Universal Access in HCI*, LNCS Vol. 8513, 2014, pp 236-245.
20. Barthel, R., Kröner, A., Hauptert, J. Mobile interactions with digital object memories, *Pervasive and Mobile Computing*, Vol. 9(2), 2013. pp. 281–294.
21. Al-Khalifa, Hend S. Utilizing QR code and mobile phones for blinds and visually impaired people. *LNCS Vol. 5105*, 2008, pp 1065-1069.
22. Ding, C., Wald, M. and Wills, G. Linked data for accessibility: from techniques to users. *Proc. Int. Conf. on e-Society*, Lisbon, PT, 13 - 16 Mar 2013, pp. 514-516.
23. Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., & Katz, Y. Pellet: A practical owl-dl reasoner. *Web Semantics: science, services and agents on the World Wide Web*, 5(2), 2007, pp. 51-53.
24. Thomas, K., Bandara, A. K., Price, B. A., & Nuseibeh, B. Distilling privacy requirements for mobile applications. In *Proc. of the ACM Conference on Software Engineering*, 2014, pp. 871-882.
25. Verdouw, C. N., Beulens, A. J. M and Vorst van der J. G. A. *Virtual Logistic Networks in Dutch Horticulture*. *Proc. of the 4th Production and Operations Management World Conference*, Amsterdam, 2-4 July, 2012.
26. Ma, H.D., *Internet of Things: Objectives and Scientific Challenges*. *Journal of Computer Science and Technology*. Vol. 26, No 6, 2011, pp. 919-924.
27. Coccoli, M., Vercelli, G., Vivanet, G., *Semantic Wiki: A collaborative tool for instructional content design*, *Journal of E-Learning and Knowledge Society*, Volume 8, Issue 2, May 2012, pp. 113-122.
28. Krötzsch, M., Vrandečić, D., & Völkel, M. *Semantic mediawiki*. In *The Semantic Web-ISW*, Springer Berlin Heidelberg, 2006, pp. 935-942.
29. Norvig, P. and Cohn, D., *Adaptive software*. *PC AI Magazine*, Vol. 11, No. 1, 1997, pp. 27-30.
30. Dey, A.K., Abowd, G.D. and Salber, D. A conceptual framework and a toolkit for supporting rapid prototyping of context-aware applications, *Human-Computer Interactions Journal*, Vol. 16(2), 2001, pp.97–166.
31. Kurz, T., Güntner, G., Damjanovic, V., Schaffert S. and Fernandez, M. *Semantic enhancement for media asset management systems*. *Multimedia Tools Appl.* 70(2), 2014, pp. 949-975.
32. Reid, Robert, Torri Ortiz Lienemann, and Jessica L. Hagaman. *Strategy instruction for students with learning disabilities*. Guilford Publications, 2013.

# Digitally Enhanced Assessment in Virtual Learning Environments

Pierpaolo Di Bitonto, Enrica Pesare, Teresa Roselli, Veronica Rossano

Department of Computer Science

University of Bari

Via Orabona, 4 – Bari, Italy

{enrica.pesare, teresa.roselli, veronica.rossano}@uniba.it

**Abstract**—One of the main challenges in teaching and learning activities is the assessment: it allows teachers and learners to improve the future activities on the basis of the previous ones. It allows a deep analysis and understanding of the whole learning process. This is particularly difficult in virtual learning environments where a general overview is not always available. In the latest years, Learning Analytics are becoming the most popular methods to analyze the data collected in the learning environments in order to support teachers and learners in the complex process of learning. If they are properly integrated in learning activities, indeed, they can supply useful information to adapt the activities on the basis of student's needs. In this context, the paper presents a solution for the digitally enhanced assessment. Two different Learning Dashboards have been designed in order to represent the most interesting Learning Analytics aiming at providing teachers and learners with easy understandable view of learning data in virtual learning environments.

**Keywords**- *Learning Analytics; Virtual Learning Environments; Assessment; Learning Dashboard*

## I. INTRODUCTION

In educational processes the formative evaluation plays a key role in effectiveness of learning since it allows the learning path to be adapted to actual student's abilities [1]–[3]. It differs from the summative assessment that aims at evaluating the educational outcomes of a specific learning path. In order to apply the formative evaluation, virtual environments supply different tools, such as quizzes, online exercises, and so on.

They are important both for students, that can self-assess the acquired knowledge, and for the teachers, that could verify if her/his educational strategies are adequate to the classroom measuring how much of the topics have been assimilated by the students. But, in e-learning contexts in order to make the formative evaluation informative it could not be limited to results of quizzes and tests. It is important to enrich those results with data about the interactions between the users (students and teachers) and the system. For example, interesting data can be the level of participation to the different activities, the quality of interaction and communication among peers, and so forth. This perspective was also the focus of the Working Group at EDUsummIT 2011 [4], [5]. The group stated that digitally-enhanced assessment requires: 1) an authentic learning experience involving digital media with 2)

embedded continuous unobtrusive measures of performance, learning and knowledge, which 3) creates a highly detailed (high resolution) data records which can be computationally analyzed and displayed so that 4) learners and teachers can immediately utilize the information to improve learning.

In this context the paper presents a solution for improving the formative assessment in e-learning platforms. In particular, the learning analytics will be studied in order to be integrated in an e-learning platform to manage the available data. Finally, two different dashboards were designed and built to facilitate the interpretation of data using a graphical representation.

## II. MOTIVATION AND PROBLEM DEFINITION

The analysis of the state of the art about the assessment allows different approaches to be classified in quantitative and qualitative methods. The quantitative approaches usually are focused on analytic measures and quantification of the student performances to make them understandable and comparable. Often the quantitative assessment is used for the summative evaluation to measure the knowledge and skills acquired at the end of a learning path. The qualitative approaches, instead, aim at improving the learning process, while it happens, giving continuous feedbacks to promote actions and interventions to reduce the gap between the performance actually achieved by the learner and the expected performance. These types of assessment are used for the formative evaluation. The two approaches have different goals, methods and consequences but they are not necessarily at odds. Recently, however, the need to prove the effectiveness of educational institutions at different levels with evidence of the success of the educational activities has pushed the quantitative approach more than the qualitative one.

Beyond this, the assessment is a complex process: the traditional “face to face” education relies on the role of the evaluator, like a teacher or a team of teachers, who is required to carefully consider and weigh all the criteria involved in the final evaluation. In distance learning environments, the evaluator rarely has the overall picture of the learning process. Often, in fact, only quantitative evaluations, such as multiple-choice tests, are used. These are unreliable and not always significant [6], [7]. But the assessment in virtual environments presents new opportunities and challenges that should be investigated.

Research on digitally enhanced assessment is still at early stage [8]: it is necessary to understand if and how technology can support both the quantitative and the qualitative assessment. Moreover, new models of students' evaluation and assessment are requested to take full advantage of technologies [9]. As pointed out by Pachler et al. [1], indeed, the technology for assessment are not educational itself, but they can empower the educational effectiveness of assessment processes. Among the different emergent technologies the Learning Analytics have a high potential in it.

#### A. *Learning Analytics*

The Learning Analytics (LAs) represent the “measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and environments in which it occurs” [10].

Research in this field is becoming very popular because the digitally enhanced assessment is very pressing in virtual learning and LA can supply the perfect tool to this end. The LA, in fact, gives methods to interpret data collected in LMSs to understand which activities involve learners and to customize the learning processes. Moreover, these data are useful also to the learners to become aware of their own knowledge and abilities in specific contexts. Thus, higher results can be achieved if students and institutions would be involved as stakeholders in the definition of learning analytics [11]. To this aim, some researches distinguish Learning Analytics (LA), Academic Analytics (AA) and the Educational Data Mining (EDM). Each of them involves different stakeholders, with distinct purposes at various levels of abstraction [12], [13]. Their common goal is to process data to find out problems and plan solutions in order to enrich the learning paths and to ensure educational success.

In particular, the EDM are useful to get value from large sets of data using data mining and machine learning methods, AA are useful to evaluate and analyze university and educational institutions from an organizational point of view [14], while LA are addressed to analyze data in order to model social connection and learning preferences in educational settings. In this perspective, the LA are the most suited to support the digitally enhanced assessment.

As said before, the LA analyze mainly the user generated data, one of the main problems dealing it is the privacy. To this end, Slade and Prinsloo [11] propose to distinguish two levels of LA data usage: the educational level and the no educational ones. The first one aims to facilitate the evaluation, reflection and personalization of curricula and it is mainly addressed to students and teachers; the second one is addressed to business analysis of the educational institutions.

Another risk in using LA is to exceed in the quantification of activities. The LA can be used “to track learner progress, to assist in developing and maintaining motivation, to help the definition of realistic goals and to develop plans to achieve them” [15]. But performance measurements may not be enough if they are not enriched by appropriate reflections on the learning itself.

### III. LEARNING ANALYTICS PROCESS

Given the complexity of the assessment process and the inadequacy of fully automated evaluations to take into account many factors, a digitally enhanced assessment proposal has been defined. The work uses the LA to provide teachers and learners a set of tools to simplify the assessment process and to make more significant the assessment results.

First of all, we need to identify and collect the interactions: as a matter of fact, during learning activities, students interact through the system with other people and resources. The type and intensity of the interactions vary depending on both the learning environment and the educational resources.

To this end it is important to classify the resources on the basis of their interactivity type: it is active if the content is mainly practical, such as exercise, experiment, and so on; it is explicative if the content is expositive, such as text, slides, and so forth. Moreover, it is important to classify the interactions with people (other students, teachers, tutors) that can be synchronous (through video conferences, chat, etc.) or asynchronous (through forums, wikis, mailing lists, etc.). In both cases, interactions with people and resources supplied data trails that can be analyzed and used in order to improve the learning process.

Once these data are collected and classified, Chatti et al. [16] propose a cyclic interaction where they are processed, analyzed and presented. For each new visualization, in fact, data need to be pre-processed, starting from new queries to original data, and then presented. In order to make the data significant for the formative assessment it is important to use different information visualization techniques. Visual representation of learner data, in fact, allows teacher to monitor the learner's progress and to provide her/him more effective feedbacks than those based on quizzes and tests results. Furthermore the visual representation of data can be shared between teacher and learner, allowing discussion and reflection on actual data in order to improve student learning awareness.

Finally, this reflection can have impact on both the teaching/learning strategies to be adopted and the type of content to be supplied. In details, the teacher will be able to enhance the educational paths, using new teaching strategies, tools and/or teaching materials, and to adapt the training/learning process to the learners; on the other, the learner will be aware of her/his achieved and not achieved learning goals and will be able to change her/his learning strategies according to the outcomes.

### IV. DASHBOARD DESIGN

To implement the described process, two dashboards have been designed for the teacher and the student.

The research on the dashboard design is still emerging but the properties proposed by Few [17] are interesting. The information in the dashboard: (1) has to support situational awareness and to promote rapid perception through the use of different visualization technologies; (2) should be presented in a way that would facilitate the decision-making process; (3) has to present, preferably in one view, the most important data that must be emphasized more than the rest.

To achieve these goals we analyzed the predictors and indicators, the learning analytics techniques and the actions and responses proposed in the literature in order to properly design our dashboards.

#### A. Predictors and Indicators

According to EDUCAUSE [18], three types of predictors and indicators has been identified: Dispositional Indicators, Activities and Performance Indicators, and Student's Artifacts.

Dispositional Indicators come from the information available on the student's background when s/he faces a new learning context. They are used before the beginning of the course, providing some information about his/her predisposition to learn. Many indicators are impartial and easily quantifiable (e.g. age, gender, ethnicity, grade point average, etc.); some of them are powerful predictors (e.g. the grade point average) but some research works have also included psychological measures of aptitude. Shum & Crick [19], for instance, propose the use of learning analytics to make visible the learning aptitudes and the transferable skills associated with learning in different contexts, measured on 7 dimensions (Changing and learning, Critical curiosity, Meaning Making, Creativity, Learning Relationships, Strategic Awareness, Resilience) through the questionnaire Effective Lifelong Learning Inventory (ELLI).

Activities and Performance Indicators come from the digital "breadcrumbs" left by learners during their learning activities [18]. Some of them are quantitative and are collected using monitoring systems such as LMS logs (number and frequency of logins, number of posts in a discussion forum, grades and results of quizzes). These data are relatively simple to collect and can be easily analyzed showing the results in visualization tools.

Student's Artifacts are the results of students work: essays, blog and discussion forum posts, etc. The analysis of these artifacts can provide information about the achievement of required level of experience and reasoning skills but, unlike the activities and performance indicators, they are difficult to be automatically quantified.

#### B. Learning Analytics Techniques

For what concerning the Learning Analytics numerous techniques have been proposed to identify meaningful patterns from the data set of the educational field. Chatti et al. [16] distinguishes them in statistics, information visualization, data mining and social network analysis.

Several Learning Management Systems (LMS) implement simple reporting tools that provide basic statistics on the interactions of the student with the system, such as total number of access, number of access per page, the distribution of access over time, posting and answering rate, percentage of read materials, etc. However, these techniques are often difficult to interpret for the LMS users.

Displaying this data in visual form can simplify the interpretation and analysis of data. Thanks to human visual

perception skills, visual representation is often more effective than a simple flat text or data table. Various techniques for displaying information, such as graphs, scatter plot, 3D representations and maps, can be used to display information in a clear and understandable format. The most difficult task in this case is to define what representation is the most effective to the proposed target. In the learning context traditional data tables more often are replaced by graphs in order to better represent the learner performances.

Data mining techniques can be used to generate prediction and classification models from collected data, to organize them in cluster according to their similarity in order to discover association rules and interesting relationship among data.

Finally, social network analysis allows connections among users in a learning environment to be discovered.

#### C. Actions and Responses

Based on these data and techniques it is necessary to determine which actions and responses are the most effective for both students and teachers. Indeed, students often pay very little attention (or sometimes none at all) to the tools and resources that they perceive as mechanical, impersonal and superficial; Actions and Responses generated by the LAs must be therefore properly designed.

They can be presented as fully automated responses that do not require any action from the user (such as a green/yellow/red alert), or semi-automated responses, that show significant paths in learners activities, often focusing on decreasing paths, and suggesting possible action to intervene.

### V. TEACHER AND STUDENT DASHBOARDS

The model and the dashboards, were validated using data collected in Moodle LMS during a postgraduate master in "Research Manager" in the context of the National Operative Program 2007-2013 Training Plan Project of Strengthening of structures and facilities of science and technology of the Scientific and Technological Site "Magna Grecia".

The master was organized using blended settings, thus some lecturers were supplied in e-learning and some collaborative activities (using chat and forum) were organized.

#### A. Teacher

The teacher needs to monitor the performance of all the activities and all the students in each course. Among the leading indicators, the number of accesses has been chosen because it provides adequate information with few processing.

An overview of the access is presented to the teacher when s/he enters the dashboard, as shown in Fig.1. A linear visualization has been chosen because of time series data.

Using the mouse on the graph other details are visualized using the rollover technique, a tooltip shows the exact number of accesses in each course on a specific date, to allow comparison of numerical data.

Prof. Teresa Roselli

Visione Generale

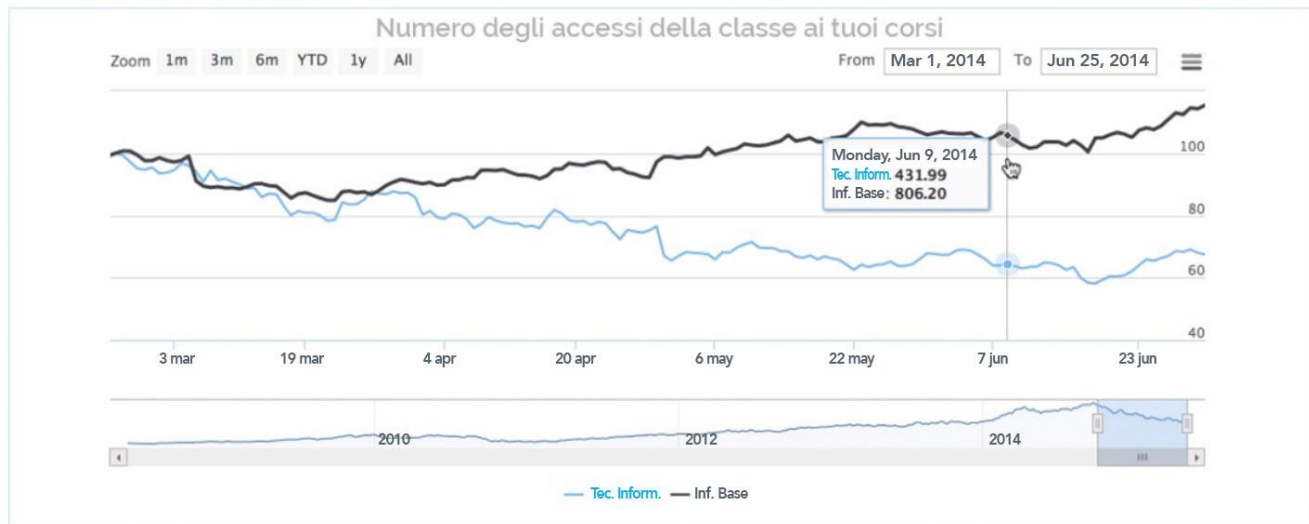


Figure 1. Number of student's access to all courses in the LMS. Different lines indicate different courses.

Also zoom and filter operations are allowed on the same graph: the teacher can zoom to preset time intervals (in the top left corner) to get a periodic overview, which is more detailed than the general one. Moreover, the teacher can select a specific interval by selecting starting and ending date (in the top right corner) to identify critical events or periods.

No explicit response is presented in this case because the evaluation of the trend is up to the teacher. To provide her/him with detailed information about the kind of activities performed during her/his courses, s/he can also have access to an overview of the activities performed in the classes organized according to the kind of activity.

Also in this case s/he can filter the data, to exclude or include specific activities in the graph: for example, if the teacher is already aware of the access to educational resources of her/his class, s/he may decide to exclude this data from the graphs (Fig. 2) and to analyze the percentage of the distribution of the collaborative activities performed (chat and forum), to discover how many collaborative activities have been performed for each course.

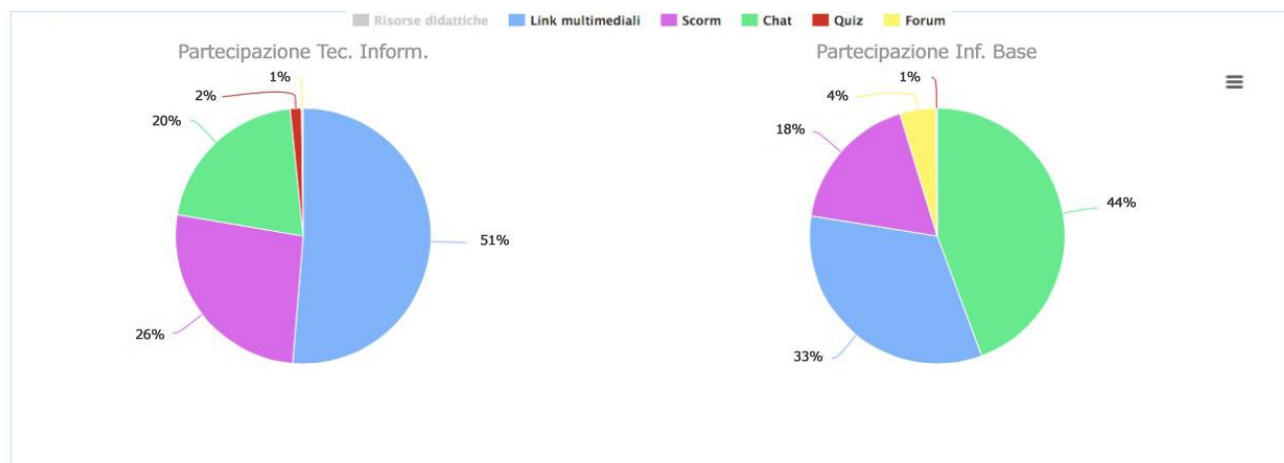


Figure 2. The percentage of student's usage of activities in the teacher courses, once the teacher has excluded the educational resources data.





Figure 3. Average students access of a specific course.

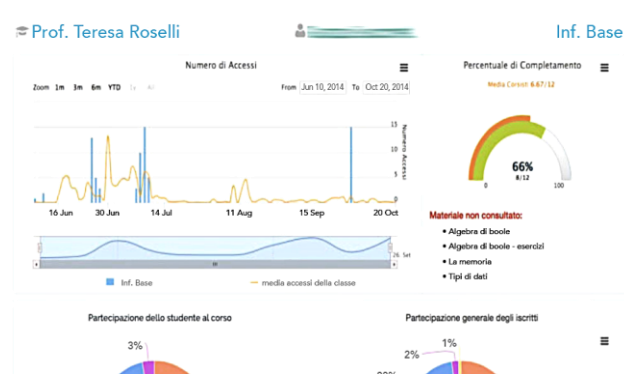


Figure 4. The student dashboard seen by the teacher (number of access, the percentage of course completion, list of resources not seen yet).

For each course, an overall dashboard is supplied in order to find out the average of access, as shown in Fig. 3.

In this case too, it will be possible to view the graph of access of a specific student and the percentages of the performed activities, such as access to learning resources, multimedia resources, chats, quizzes etc. In addition, s/he will see the average grades and results achieved by the students.

In order to compare the class average data with each student's data, the teacher can select the student dashboard (Fig. 4) and visualize the data about student's activities, such as number of access, viewed resources and performed activities, and can visualize the resources not yet seen and missing activities. This kind of analysis allows the teacher to personalize the support of the student in the rest of the course.

## B. Student

In the learning process it is important that also the student can monitor her/his own performances in the courses in which s/he is enrolled. The main dashboard will show student's access and activities graphs. In addition, s/he will also see the completion percentage of the courses, as shown in Fig. 5.

The student, as the teacher, can zooms graphs, filter data and have more details on specific features. In addition, s/he can compare her/his own performed activities with the average activities of other members of the class. Also in this case the selection of activities will draw two pie charts that allow a comparison in real time between the student's data and the average data of the class.

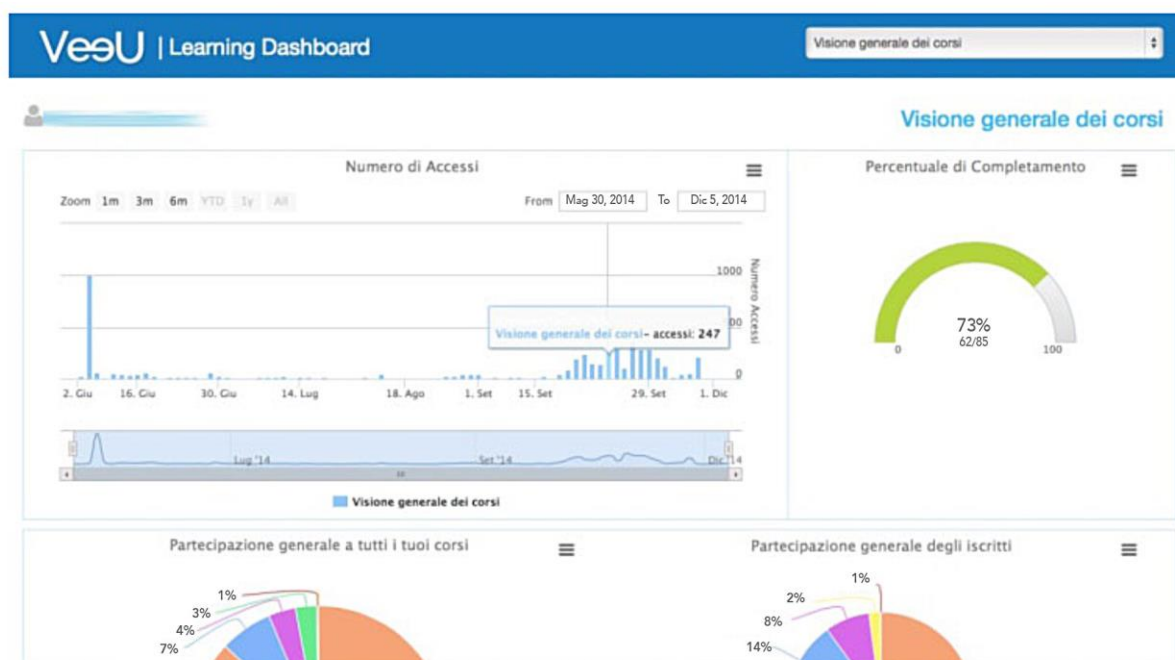


Figure 5. Main Student's Dashboard (number of access, percentage of course completion, activities graphs).

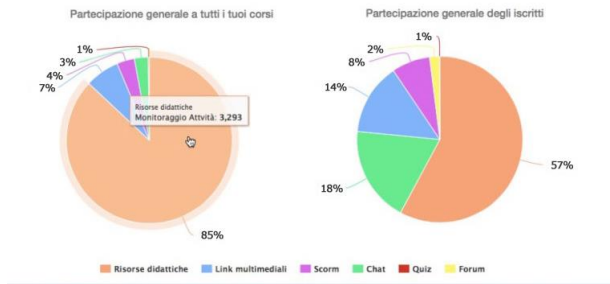


Figure 6. Student activities graph in the left side and Class activities graph in the right side (data concerns: didactic resources, multimedial links, scorm, chat, quiz, forum).

For example in Fig. 6 the student can easily notice that s/he has spent a lot of effort using learning resources but s/he has not participated in the forum. Excluding the learning resources from the pie chart s/he can also notice that also her/his participation in chat activities was lower than the average of the class (Fig. 7). This kind of comparison can suggest how the student can improve her/his participation in the course to improve learning outcomes.

Note that other students' data are here presented in aggregated form to prevent privacy issues. Numerical details are available using the rollover technique.

Moreover the student will visualize her/his own test results. The spider chart has been used to allow the comparison among the results in all courses in which s/he is enrolled (Fig. 8). The graph shows a general overview of data, while data about each single test will be visible with rollover operations on it. This allows the student to have a clear vision of the learning gap (if any) s/he has and to plan where and how to spend her/his learning time making the learning more effective.

To get a deeper view, the student can access the dashboard of each course through the appropriate drop-down menu in the top-right, as shown in Fig. 9.

In addition to the graph of access and the activities graph, this dashboard will provide student with the completion percentage of the course and the list of material and activities s/he has not completed yet, to promote an easy access to those resources in order to overcome her/his learning gaps.

## VI. CONCLUSIONS AND FUTURE WORKS

The dashboards have been tested in an off-line pilot study using data collected by Moodle LMS during a master in "Research Manager" in the context of the National Operative Program 2007-2013 Training Plan Project of Strengthening of structures and facilities of science and technology of the Scientific and Technological Site "Magna Grecia". The learning activities of the master started in May 2014 and ended in September 2014.

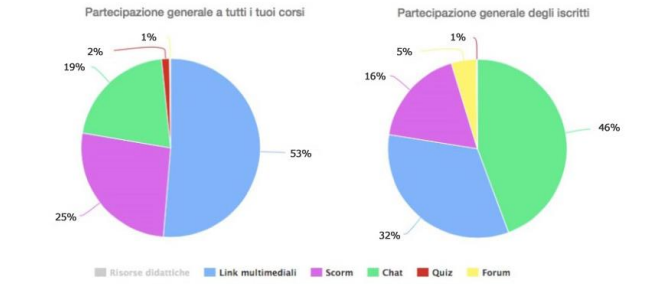


Figure 7. Student activities graph in the left side and Class activities graph in the right side where data about didactic resources has been excluded.

For each course the students were required to access to the didactic material published in the LMS, and to participate to collaborative activities through forums and chats. The data collected during the e-learning activities allows to test that all graphs and data were well visualized in the dashboards. A deep analysis of data could be done in order to discover relationships between the quality of student-system interaction and the students' outcomes.

Currently, a plugin for Moodle is being developed in order to integrate the dashboards in the e-learning environment. Then a pilot study, to measure if the visualization of data enhances the student learning, will shortly be conducted.



Figure 8. Spider chart displaying the average score obtained in the tests of all courses in which the student is enrolled.

## ACKNOWLEDGMENT

We would thank the PON - R&C 2007-2013 Project - Piano di Formazione del Progetto di Potenziamento delle strutture e delle dotazioni scientifiche e tecnologiche del Polo Scientifico Tecnologico "Magna Grecia" and the students Di Mitri Daniele, Sacchetti Grazia and Romanelli Vincenzo who build the dashboards in their bachelor thesis.

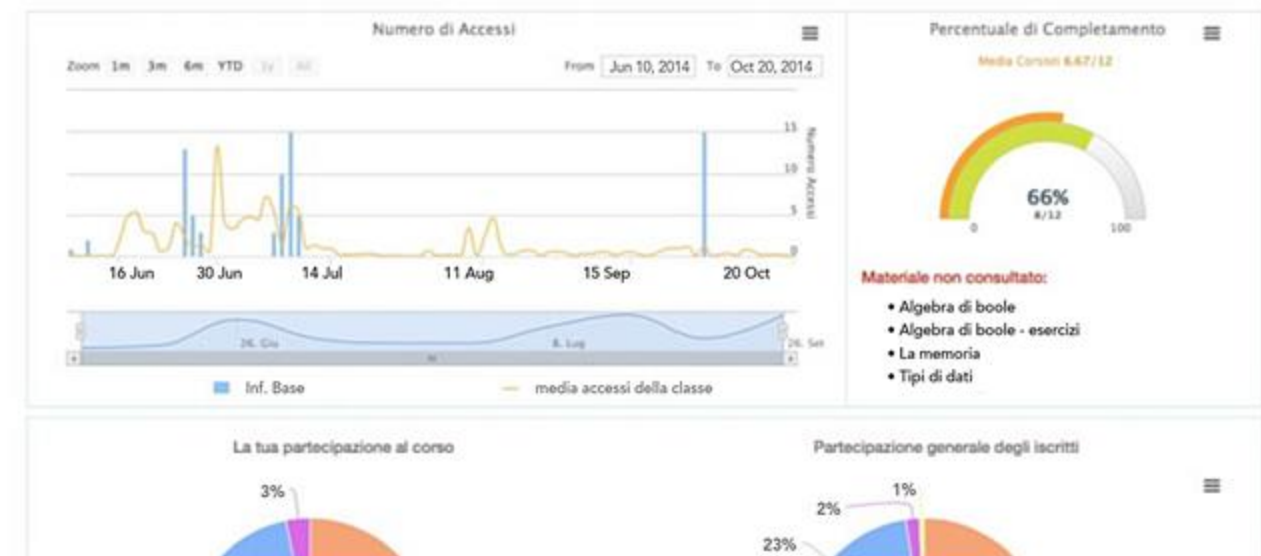


Figure 9. The student dashboard (number of access, the percentage of course completion, list of resources not seen yet).

## REFERENCES

- [1] N. Pachler, H. Mellar, C. Daly, Y. Mor, D. Wiliam, and D. Laurillard, "Scoping a vision for formative e-assessment: a project report for JISC," no. April, 2009.
- [2] P. Black and D. Wiliam, *Inside the black box: Raising standards through classroom assessment*. Granada Learning, 1998.
- [3] M. Booth, "Learning Analytics: The New Black," *Educ. Rev.*, vol. 47, pp. 52–53, 2012.
- [4] EDUsummIT, "Assessment To Move Education into the Digital Age," *EDUsummIT 2011 Build. a Glob. community policy-makers, Educ. Res. to move Educ. into Digit. age.*, 2011.
- [5] M. Webb and D. Gibson, "Challenges for information technology supporting educational assessment," *J. Comput. Assist. Learn.*, vol. 29, no. 5, pp. 451–462, 2013.
- [6] R. S. J. Baker, A. T. Corbett, K. R. Koedinger, S. Evenson, I. Roll, A. Z. Wagner, M. Naim, J. Raspat, D. J. Baker, and J. E. Beck, "Adapting to when students game an intelligent tutoring system," in *Intelligent tutoring systems*, 2006, pp. 392–401.
- [7] S. Davies, "Effective Assessment in a Digital Age," *Jisc*, vol. 2009, no. 30th July, pp. 1–64, 2010.
- [8] G. Conole and B. Warburton, "A review of computer-assisted assessment," *Res. Learn. Technol.*, vol. 13, no. 1, pp. 17–31, 2005.
- [9] M. Coccoli, A. Guercio, P. Maresca, and L. Stanganelli, "Smarter universities: A vision for the fast changing digital era," *J. Vis. Lang. Comput.*, vol. 25, no. 6, pp. 1003–1011, 2014.
- [10] G. Siemens and P. Long, "Penetrating the Fog: Analytics in Learning and Education," *Educ. Rev.*, vol. 46, pp. 30–32, 2011.
- [11] S. Slade and P. Prinsloo, "Learning Analytics: Ethical Issues and Dilemmas," *Am. Behav. Sci.*, vol. 57, no. March, pp. 1–20, 2013.
- [12] R. Ferguson, "Learning analytics: drivers, developments and challenges," *Int. J. Technol. Enhanc. Learn.*, vol. 4, no. 5/6, p. 304, 2012.
- [13] G. Siemens and R. S. J. Baker, "Learning Analytics and Educational Data Mining: Towards Communication and Collaboration," *Proc. 2nd Int. Conf. Learn. Anal. Knowl.*, pp. 252–254, 2012.
- [14] J. P. Campbell, P. B. DeBlois, and D. G. Oblinger, "Academic Analytics: A New Tool for a New Era," *Educ. Rev.*, vol. 42, no. October, pp. 40–57, 2007.
- [15] E. Duval, "Attention please!: Learning analytics for visualization and recommendation," *LAK '11 Proc. 1st Int. Conf. Learn. Anal. Knowl.*, pp. 9–17, 2011.
- [16] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, "A Reference Model for Learning Analytics," *Int. J. Technol. Enhanc. Learn.*, vol. 4, no. 5, pp. 318–331, 2012.
- [17] S. Few, *Information Dashboard Design: Displaying data for at-a-glance monitoring*. Analytics Press, 2013.
- [18] M. Brown, "Learning Analytics: Moving from Concept to Practice," *Educ. Learn. Initiat. Br.*, no. July, pp. 1–5, 2012.
- [19] S. Buckingham Shum and R. Deakin Crick, "Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics," *LAK '12 Proc. 2nd Int. Conf. Learn. Anal. Knowl.*, no. May, pp. 92–101, 2012.

# A Distributed Framework for NLP-Based Keyword and Keyphrase Extraction From Web Pages and Documents

P. Nesi, G. Pantaleo and G. Sanesi

Distributed Systems and Internet Technology Lab, DISIT Lab, <http://www.disit.dinfo.unifi.it>

Department of Information Engineering (DINFO),

University of Florence – Firenze, Italy

[paolo.nesi@unifi.it](mailto:paolo.nesi@unifi.it), [gianni.pantaleo@unifi.it](mailto:gianni.pantaleo@unifi.it)

**Abstract**—The recent growth of the World Wide Web at increasing rate and speed and the number of online available resources populating Internet represent a massive source of knowledge for various research and business interests. Such knowledge is, for the most part, embedded in the textual content of web pages and documents, which is largely represented as unstructured natural language formats. In order to automatically ingest and process such huge amounts of data, single-machine, non-distributed architectures are proving to be inefficient for tasks like Big Data mining and intensive text processing and analysis. Current Natural Language Processing (NLP) systems are growing in complexity, and computational power needs have been significantly increased, requiring solutions such as distributed frameworks and parallel computing programming paradigms. This paper presents a distributed framework for executing NLP related tasks in a parallel environment. This has been achieved by integrating the APIs of the widespread GATE open source NLP platform in a multi-node cluster, built upon the open source Apache Hadoop file system. The proposed framework has been evaluated against a real corpus of web pages and documents.

**Keywords** – *Natural Language Processing, Part-of-Speech Tagging, Parallel Computing, Distributed Systems.*

## I. INTRODUCTION

Nowadays we are living in a hyper-connected digital world, dealing with computational solutions progressively more oriented to data-driven approaches, due to the increasing population of web resources and availability of extremely vast amounts of data [1]. This fact has opened new attractive opportunities in several application areas such as e-commerce, business web services, Smart City platforms, e-Healthcare, scientific research, ICT technologies and many others. Modern information societies are characterized by handling vast data repositories (both public and private), according to which Petabyte datasets are rapidly becoming the norm. For instance, in 2014 Google has been estimated to process over 100 Petabytes per day on 3 million servers<sup>1</sup>; in the first half of 2014, Facebook warehouse has been reported to store about

300 PB of Hive data (with an incoming daily rate of about 600 TB)<sup>2</sup>. These numbers reveal how the process of storing data is rapidly overcoming our ability to process them in an automatic and efficient way. Such a huge pool of available information has attracted a lot of interest within several research and commercial scenarios, ranging from automatic comprehension of natural language text documents, supervised document classification [2], content extraction and summarization [3], design of expert systems, recommendation tools, Question-Answer systems, query expansion [4], up to social media mining and analysis, in order to collect and assess users' trends and habits for target-marketing, customized services. Therefore, applications and services must be able to scale up to items, domains and data subset of interest. Approaches based on parallel and distributed architectures are spreading also in search engines and indexing systems; for instance, the ElasticSearch<sup>3</sup> engine, which is an open source distributed search engine, designed to be scalable, near real-time capable and providing full-text search capabilities [5].

NLP systems are commonly executed in a pipeline where different plugins and tools are responsible for a specific task. One of the most intensive tasks is annotation, defined as the process of adding linguistic information to language data [6], usually related to their grammatical, morphological and syntactical role within the given context. Text annotation is at the basis of higher level tasks such as extraction of keywords and keyphrases (which deals with the identification of a set of single words or small phrases representing key segments that can describe the meaning of a document) [7], up to the design of Semantic Computing frameworks (involving activities such as supervised classification of entities, attributes and relations), that lead to the production of structured data forms, typically in the form of taxonomies, thesauri and ontologies. Current sequentially integrated NLP architectures, where each pipeline step uses intermediate results and outcomes produced by previous processing stages, are prone to problems related with information flow, from congestion to information losses [8]. It

<sup>1</sup> <http://www.slideshare.net/kmstechnology/big-data-overview-2013-2014>.

<sup>2</sup> <https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/>.

<sup>3</sup> <https://www.elastic.co/products/elasticsearch>



has been often addressed in literature how existing NLP tools and frameworks or are not well suited to process very large corpora, since their primary design focus was not oriented to scalability [9]. The only reasonable approach to handle Big Data problems seems to be the well-known computer science concept of “*Divide and Conquer*” [10]. The basic idea is to partition a large problem into smaller sub-problems; to the extent that these sub-problems are independent, they can be parallelly processed by different threads. This aspect, concurrently with the evolution of distributed systems multi-processor computer architectures and network speeds, is leading to the application of distributed architectures and parallel-computing paradigms to Big Data mining and processing activities.

Keywords and keyphrases annotation can be useful for content extraction and summarization, in order to produce machine-readable corpora, as well as building content-based multi-faceted search queries. For instance, scientific articles are often annotated with keywords, in a way similar as it happens with metadata annotation of multimedia resources. Conversely, expanding our view to the available online resources, currently a large portion of web documents still does not have any keywords or keyphrases assigned. Besides, since manual annotations results to be an extremely time-consuming and inefficient process, the necessity arises to design efficient and scalable automated solutions to extract keywords and keyphrases from massive amounts of unstructured text documents.

The present work describes a novel framework which allows the execution of general NLP tasks, through the use of the open source GATE<sup>4</sup> tool [11], on a multi-node cluster based on the open source Apache Hadoop<sup>5</sup> Distributed File system (HDFS). The paper is organized as follows: Section II illustrates related work, in terms of state of the art and open issues for both commercial and research literatures; in Section III, an architectural overview of the proposed system is presented; in Section IV, a validation of the system, performed against a real corpus of web resources, is reported; finally, Section V is left for conclusions and future perspectives.

## II. RELATED WORK

The task of Automatic keyword extraction has been extensively studied in literature. Existing methods are typically divided into four categories: simply statistic, linguistic, machine learning and other mixed approaches [12]. Statistical methods usually relies on term position, term frequency, co-occurrence and related relevance metrics, such as TF-IDF [13]. The linguistic approach is based on NLP techniques, such as Part-of-Speech (POS) tagging, lexical analysis, syntactic analysis, possibly exploitation of semantic features [14]. Machine learning methods treat the keyword extraction task as supervised learning problem using a training dataset, using different techniques such as naive Bayes algorithms [15], least square support vector machines (LS-SVM) [16], etc. Other mixed approaches mainly use a combination of the previously

described techniques, possibly adding some heuristic knowledge, for instance the use of annotated lists, gazetteers, blacklists to remove stop words [17], selecting only certain part-of-speech tags as candidate keywords [18]. Moreover, external resources (in addition to training corpora) can be used as lexical knowledge bases, such as Wikipedia [19] or DBpedia. Graph-based approaches typically extract a graph from each input documents and use a graph-based ranking function to determine the relevance of the nodes, which yields the importance of key terms [20]. Topic-based clustering is used in content extraction and text summarization methods; this usually involves grouping the candidate keyphrases into topics or domains [21], [22].

We find first attempts of employing parallel computing frameworks for NLP tasks in the middle 90s: Chung and Moldovan [23] proposed a parallel memory-based parser called PARALLEL, implemented on a parallel computer, the Semantic Network Array Processor (SNAP). Later, Van Lohuizen [24] proposed a parallel parsing method relying on a work stealing multi-thread strategy in a shared memory multi-processor environment. Hamon et al. realized Ogmios [9], a platform for annotation enrichment of specialized domain documents within a distributed corpus. The system provide NLP functionalities such as word and sentence segmentation, named entity tagging, POS-tagging and syntactic parsing. Jindal et al. [25] developed a parallel NLP system based on LBJ [26], a platform for developing natural language applications, and Charm++ [27] as a parallel programming paradigm. Exner and Hugues recently presented Koshik [28], a multi-language NLP processing framework for large scale-processing and querying of unstructured natural language documents distributed upon a Hadoop-based cluster. It supports several types of algorithms, such as text tokenization, dependency parsers, coreference solver. The advantage of using the Hadoop distributed architecture and its programming model (known as MapReduce), is the capability to efficiently and easily scale by adding inexpensive commodity hardware to the cluster.

We find also commercial tools aiming at solving the addressed problem. Beemoth<sup>6</sup>, produced by Digital Pebble, is an open source platform for large scale document processing based on Apache Hadoop, employing third party NLP tools, including GATE, Tika and UIMA. InfoTech RADAR<sup>7</sup> is a software solution implementing NLP and Sentiment Analysis on the Hortonworks Sandbox Hadoop environment. Some existing NLP tools have proposed improvements to realize large-scale processing solutions; this is the case of the GATEcloud project [29], which is an adaptation of the GATE software suite to a cloud computing environment (using the PaaS paradigm), although it is delivered under a fee payment, unlike the original platform.

Recently, other parallel computing frameworks have been proposed in addition to the Hadoop MapReduce. For instance, Spark framework [30] has been originally developed at Berkeley UC to support a wider class of applications,

<sup>4</sup> <https://gate.ac.uk/>

<sup>5</sup> <http://hadoop.apache.org/>

<sup>6</sup> <https://github.com/DigitalPebble/beemoth/wiki/tutorial>

<sup>7</sup> <http://www.itcinfotech.com/software-product-engineering/solutions/RADAR.aspx>

especially those implementing acyclic data flow models, such as iterative algorithms and applications requiring low-latency data sharing processes. Spark introduces programming transformations on Resilient Distributed Datasets (RDD) [31], which are read-only collections of objects distributed over a cluster of machines providing fault tolerance by rebuilding lost data by using lineage information (without requiring data replication). RDD allows to store data on memory, as well as to define the persistence strategy. Gopalani and Arora [32] have compared Spark and Hadoop performances on clustering K-means algorithms, showing that Spark outperforms Hadoop on different cluster configurations.

### III. SYSTEM ARCHITECTURE

As introduced earlier, the proposed system aims at extracting keywords and keyphrases from web resources (retrieved by crawling online web pages and documents of business entities and research institutes) in a distributed architecture. The necessity of realizing a more efficient and scalable solution, in order to improve performances, as well as data integrity and failures handling, led us to the choice of the open source Apache Hadoop framework, which have been installed on a multi-node cluster. The Hadoop ecosystem is implemented in Java, and it is capable of supporting distributed

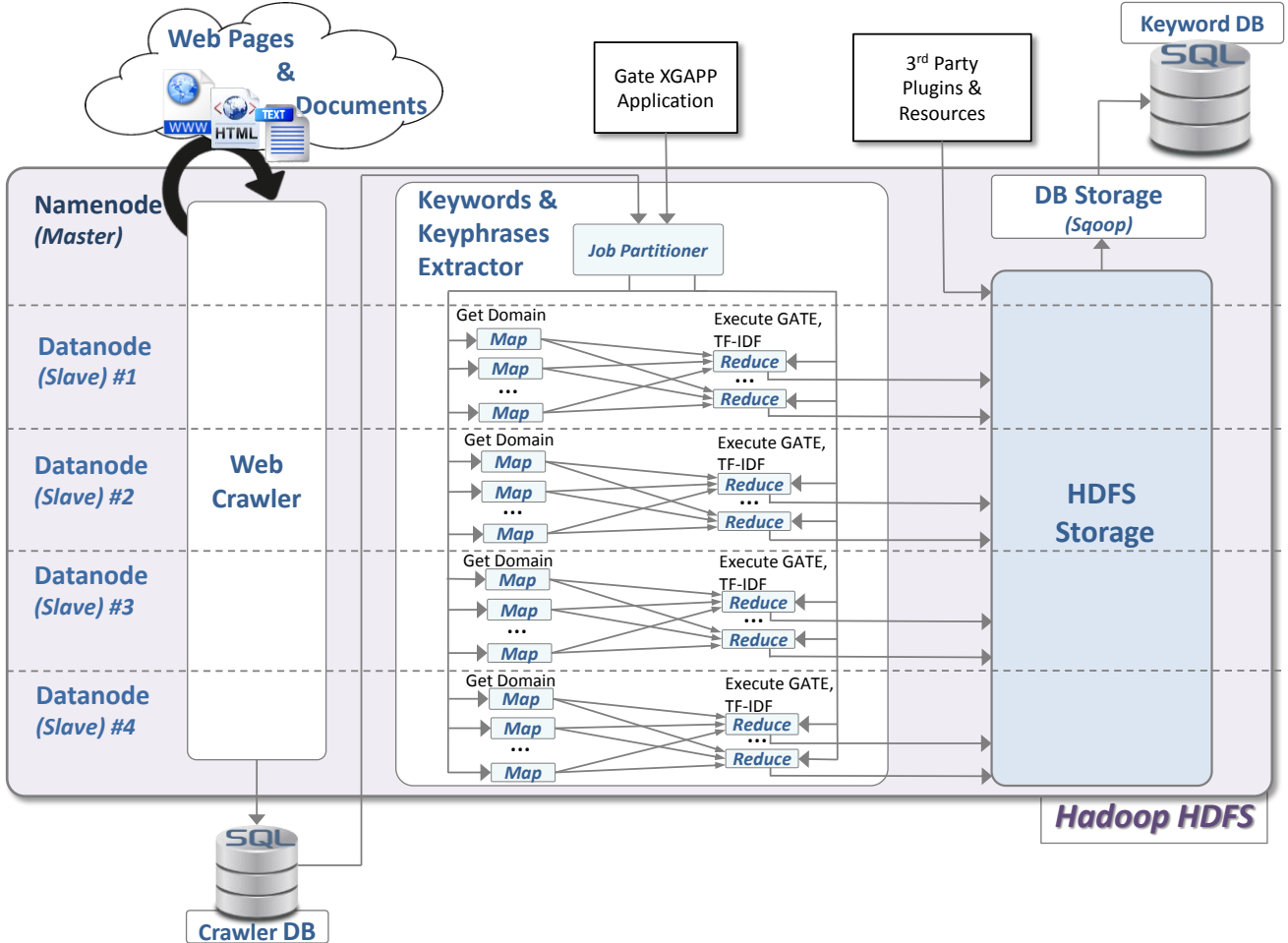


Figure 1. Overview of the proposed system architecture.

applications, large-scale data processing and storage, providing high scalability. Data access and management relies on the Hadoop Distributed File System (HDFS), modeled upon the Google File System – GFS (Ghemawat et al., 2003). Typically, a cluster is composed by a master Namenode, which assigns tasks and tracks their execution to the different clients (Datanodes). Datanodes are responsible also for data storage through its MapReduce programming model.

In our context, MapReduce is used to parallelize the crawler work, the execution of NLP tasks and final SQL database

storage. In our case, NLP tasks are performed by using the GATE Embedded Java APIs, and they are related to the extraction of keywords and keyphrases from online parsed unstructured text, although our approach is general enough to potentially execute any generic GATE-based application. At the end of the distributed process (involving web crawling, text parsing, annotation, keywords/keyphrase extraction and pruning, based on their relevance), the produced outcome is stored in the HDFS file system. Finally, a dedicated procedure stores designed keywords and keyphrases in an external SQL



database. The whole system architecture is implemented in Java. In the next subsections, a description will follow of the main modules constituting the proposed framework, which are listed as following:

- The *Web Crawler* module.
- The *Keywords/Keyphrases MapReduce Extractor* module, which is responsible of two main operations:
  - The execution of the GATE application via MapReduce and the subsequent storage of extracted keywords/keyphrases in the HDFS.
  - The keywords and keyphrases relevance estimation, obtained by computing the TF-IDF function for each extracted keywords/keyphrases, performed to assess their relevance with respect to their whole corpus (for filtering purposes).
- The *DB Storage* module which finally stores designed keywords and keyphrases into an external SQL database.

An overview of the proposed architecture is depicted in Figure 1. The next subsection are committed to describe in further details the above listed modules.

#### A. Web Crawler

The crawling engine of the proposed system is based on the open source Apache Nutch<sup>8</sup> tool. It has been initialized with a set of seed URLs of commercial companies, services and research institutes. Actually, as later addressed as an open issue for future work, a future goal will be to exploit the capabilities of the proposed system in annotating text relevant features for content extraction and summarization (in order to estimate, as an example, the domain of interest of analyzed web domains, as well as the main key topics).

The Nutch based crawler workflow is divided into the following phases: The first is the *Inject* phase, in which some seed URLs are injected for initial bootstrapping. Then, in the *Generate* phase, a generating algorithm produces a set of URLs that are going to be fetched. The *Fetch* phase deals with fetching the previously generated set of URLs into segments. Subsequently, the *Parse* phase is dedicated to the parsing of fetched segments content; the *Update* phase populates an external SQL Database with parsed segment contents. Finally, the Apache Solr<sup>9</sup> technology is used to index all the collected documents, providing also a search interface. The Solr index is ultimately present only onto the master Namenode.

#### B. Keywords/Keyphrases MapReduce Extractor

This module reads and takes as input the URLs (stored in the SQL database). The content of the corresponding parsed

web page is obtained through a query to the Solr index. In this way, documents are created for each corpus represented by a single web domain. The crawler and keywords extractor modules work asynchronously, actually the crawling task and the extraction process are scheduled and executed independently. Moreover, the present module is in charge of defining the MapReduce model. Typically, Map functions divide the work into smaller jobs or file blocks, which are subsequently mapped among the different Datanodes by the Namenode. Specific Map functions are defined to generate key/value pairs representing logical records from the input data source. Subsequently, Reduce tasks merge all intermediate values associated with the same key. The Hadoop services in charge of managing and assigning tasks and data blocks to the different nodes are the *JobTracker* and the *TaskTracker*. The former communicates with the Namenode to retrieve data locations and the directory tree of all files in the file system, in order to find available nodes and set specific tasks to assign. The latter represents a node in the cluster that is able to receive MapReduce tasks from a JobTracker; each TaskTracker is configured with a set of slots, indicating the number of tasks that can be accepted). Here, Map and Reduce functions are defined in a proper way for our goals: since we are interested in annotating keywords and keyphrases at single web-domain level, we designed a Map function that associates key/value record pairs where the *key* is the URL of the single web page, and the *value* is the corresponding web domain. The Reduce function, in turn, fulfills the setup, launch and execution of a multi-corpora GATE application (each corpus containing text documents and pages belonging to a single web domain), as well the subsequent estimation of extracted keywords/keyphrases relevance at web domain level (as later described).

#### C. GATE Application

This functional block of the *Keywords/Keyphrases MapReduce Extractor* module integrates and executes the GATE application in the MapReduce environment, according to the input configuration parameters and the pipeline, defined in an external configuration .xgapp file. This XML-based file is defined, by extension, as the effective GATE application, containing file paths and references to all the Processing Resources and plugins used. In this specific case, the ANNIE (*A Nearly-New Information Extraction System*) plugin, more specifically the *Tokenizer* and the *Sentence Splitter* tools, have been used to parse and segment the text content of crawled documents, while the *TreeTagger* plugin has been used for POS-tagging. Finally, the Java written *JAPE* (*Java Annotation Pattern Engine*) plugin syntax has been employed to define custom rules for filtering undesired, noisy parts of speech (such as conjunctions, adverbs, prepositions etc.). These rules are contained in a dedicated .jape file. Common nouns and adjectives are then annotated as potential keywords candidates. Candidate keyphrases are then identified as contiguous phraseological combinations and patterns of candidate keywords.

<sup>8</sup> <http://nutch.apache.org/>

<sup>9</sup> <http://lucene.apache.org/solr/>

Regarding the execution of the GATE application in the Hadoop distributed environment, the following strategy has been followed: the Namenode loads, at run time, a zip archive containing all the needed GATE APIs, configuration and application files, libraries and plugins, in the HDFS Distributed Cache. This is one of the possible solutions adopted for handling read and write operations on files in the HDFS. This has been considered an efficient solution since, by this way, our application will copy in memory and extract the necessary files only once, and the allocated content will be accessible by all the Datanodes, without the need of installing required plugins and third party tools on each single cluster node. An additional advantage of this approach stems by the fact that any generic GATE application can be potentially executed in our Hadoop-based architecture (taking benefits of all the NLP features and capabilities offered), providing to embed the .xgapp application file, the .jape file for annotation rules and patterns, as well as all the required resources and plugins in the input zip file.

#### D. TF-IDF Relevance Estimation

Relevance estimation for candidate keywords and keyphrases is performed by computing the TF-IDF (*Term Frequency – Inverse Document Frequency*) function. This metric is widely adopted in Information Retrieval to assess how significant is a given term not really within the single document in which it has been retrieved, but rather with respect to the whole documents collection (corpus). Actually, a lot of common words like articles or conjunctions may appear several times in a document but they are not relevant as key-concepts to be indexed or retrieved. As a matter of fact, TF-IDF is given by the combination of two functions: TF (*Term Frequency*) which provides a measure about how frequently a term occurs in a certain document, and IDF (*Inverse Document Frequency*) which measures how important is a certain term with respect to the whole corpus. The final TF-IDF value for a candidate keyword  $k$  in a document  $d$  within a corpus  $D$  is calculated as:

$$(TF - IDF)_k = TF_k \cdot IDF_k ,$$

where:

$$TF_k = \frac{f_k}{n_d} , \quad IDF_k = \log \frac{N_D}{N_k}$$

being  $f_k$  the number of occurrences of the candidate keyword  $k$  in the document  $d$ ,  $n_d$  the total number of terms contained in document  $d$ ,  $N_D$  the total number of documents in the corpus  $D$  and  $N_k$  the total number of documents within the corpus in which the candidate key  $k$  appears.

After calculating TF-IDF for each potential keyword and keyphrase, candidates with a TF-IDF value under a defined threshold are pruned, while those with a value above the threshold are designed as definitive keywords/keyphrases.

These ones are finally stored in the Hadoop HDFS file system, annotated together with their corresponding TF-IDF values and source web domain URL.

#### E. DB Storage

This module accomplishes the storage of the final system output (temporarily stored in the HDFS by the Extractor Module) in an external SQL database. This operation is carried out by means of the Apache Sqoop<sup>10</sup> open source tool, specifically designed for data transfer between Hadoop HDFS and structured datastores. The Sqoop tool has been installed on the master only and the export feature has been used to insert data stored in the HDFS into the database. In order to successfully accomplish the data export, full read/write privileges on the database have been granted to all the machines on the cluster. Each database record is populated with an extracted keyword or keyphrase, its corresponding POS-tag (or a different custom tag if it is a keyphrase), TF-IDF value and the source web domain.

### IV. EVALUATION

The performances of the proposed system have been evaluated against a dataset composed of 10000 web page and documents, which is a subset of the resources ingested by our Distributed Crawler module (which has currently been gathering more than 6 million web URLs). The Hadoop cluster architecture used for tests has been assessed on different configurations, ranging from 2 to 5 nodes. Each node is a Linux 8-cores workstation with Hadoop HDFS installed. The master Namenode, besides, required also the Apache Sqoop software installation to manage SQL write operations for final output storage. In order to avoid data integrity errors and failures due to decommission and recommission of cluster nodes, Hadoop allows to perform a rebalance of stored blocks among the active nodes of the cluster, if necessary.

For each cluster configuration, a fixed number of keywords/keyphrase extraction tests has been performed, on the same defined dataset. The MapReduce model supplies speculative execution of tasks, and it is designed to provide redundancy in order to handle fault tolerance. By this way, it may happen that the Namenode JobTracker has to reschedule failed or killed tasks, and this can affect the execution time of the whole process. Therefore, for performance comparison, the best processing times have been selected among all the tests performed for each node configuration. By this way, the number of attempts for re-executing failed or killed tasks is supposed to be minimized. For the whole test dataset, containing about 10000 documents, a total of nearly 3.5 million keywords and keyphrases have been extracted. Time processing results for the different tested node configurations are shown in Table 1. The single-node configuration has been actually implemented as a two-nodes cluster, with a master Namenode and one slave node running only as a Datanode (in order to avoid HDFS space and blocks balancing problems on

<sup>10</sup> <http://sqoop.apache.org/>

decommissioning too many nodes). That is, the TaskTracker daemon stopped (so that it does not take part to the MapReduce process). As a term of comparison, running the same GATE application on the same corpora dataset on a single non-Hadoop workstation took approximately 60 hours. A possible explanation to this significant performance gap can be the fact that the Java code of our standalone GATE application is not optimized for multi-thread, while the MapReduce adaptation executed in Hadoop can benefit of MapReduce configuration parameters, which define the maximum number of map and reduce task slots to run simultaneously (exploiting multi-core technology even on a single-node cluster). The resulting speed-up curve for our test data is shown in Figure 2.

As it can be noticed, the scaling capabilities of the proposed system confirm the nearly linear growth trend of the Hadoop architecture. However, the low curve slope suggests that significant performance improvements can be achieved only for larger number of nodes in the cluster.

TABLE I. EVALUATION RESULTS: TIME PERFORMANCES ASSESSED FOR DIFFERENT CLUSTER CONFIGURATIONS.

<i>Configuration</i>	<i>Processing Time (hh:mm:ss)</i>	<i>Speed - Up</i>
HDFS - single node	07:17:01	-
HDFS - 2 nodes	05:21:53	1.36
HDFS - 3 nodes	04:08:00	1.76
HDFS - 4 nodes	03:39:42	1.99
HDFS - 5 nodes	03:20:09	2.18

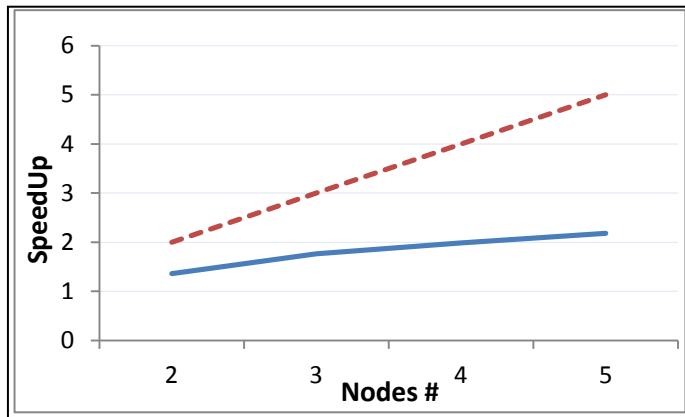


Figure 2. Processing time performances depicted for distributed extraction of Keywords and Keyphrases (solid curve), performed with different cluster configurations, against the ideal linear trend (dotted curve).

## V. CONCLUSIONS AND FUTURE WORK

In this paper, a distributed system for keywords and keyphrases extraction from text content of web pages and documents has been presented. The parallel architecture is provided through the implementation of the open source Apache Hadoop framework, while text annotation and key features extraction rely on the NLP opens source GATE

platform. The main advantages and contributions brought by the proposed system are two: the first is the integration of a web crawler which allow to use our system as a standalone application, avoiding interfacing issues with external tools for documents ingestion and indexing (actually, the most part of current NLP tools assumes that input documents are already collected, properly cleaned and formatted). In addition, the proposed system provides the capability of potentially executing any generic GATE application (thus allowing to perform a wide range of NLP activities) in a distributed design, without the need for programmers to modify and update every time the code, in order to follow the steps usually required for parallel computing development (such as task decomposition, mapping and synchronization issues). Open issues for future work are, in addition to test a wider range of GATE based applications, the evaluation of processing times on larger cluster configurations and larger document corpora, with the goal of better assessing the trend of performance improvements. Moreover, it could be interesting to adapt the Hadoop implementation of the keywords/keyphrases extraction on other parallel computing, distributed architecture, such as Spark. Regarding possible quality improvements to the currently adopted NLP solutions used for key features extraction, the annotation of proper nouns (for instance, VIP person names and toponyms) could also be provided, by the use of lists, gazetteers and other external knowledge resources. This might be useful, as well as enhancing the relevance of produced output, for content annotation, summarization and domain characterization purposes in higher expressive contexts, such as Semantic Computing frameworks.

## REFERENCES

- [1] J. Lin, and C. Dyer, "Data-Intensive Text Processing with MapReduce", Morgan & Claypool Publishers, 2010.
- [2] F. Colace, M. De Santo, L. Greco and P. Napoletano, "Text classification using a few labeled examples. Computers in Human Behavior", Vol. 30, pp. 689-697, 2014.
- [3] R. Al-Hashemi, "Text Summarization Extraction System (TSSES) Using Extracted Keywords", International Arab Journal of e-Technology, Vol. 1(4), pp. 164-168, 2010.
- [4] F. Colace, M. De Santo, L. Greco and P. Napoletano: Weighted Word Pairs for query expansion. Inf. Process. Manage. 51(1): 179-193 (2015).
- [5] O. Kononenko, O. Baysal, R. Holmes and M. W. Godfrey, "Mining Modern Repositories with Elasticsearch", in Proc. of the 11<sup>th</sup> Working Conference on Mining Software Repositories, pp. 328-331, 2014.
- [6] N. Ide. and L. Romary, "International standard for a linguistic annotation framework", Natural Language Engineering, Vol. 10(3-4), pp. 211-225, 2004.
- [7] A. Hulth, "Improved automatic keyword extraction given more linguistic knowledge", in Proc. of the 2003 Conference on Empirical Methods in Natural Language Processing, Sapporo, Japan, 2003.
- [8] T. Luis, "Parallelization of Natural Language Processing Algorithms on Distributed Systems", Master Thesis, Information Systems and Computer Engineering, Instituto Superior Técnico, Univ. Técnica de Lisboa, 2008.
- [9] T. Hamon, J. Deriviere and Nazarenko, "Ogmios: a scalable NLP platform for annotating large web document collections", in Proc. of Corpus Linguistics, Birmingham, United Kingdom, 2007.
- [10] J. Lin and C. Dyer, "Data-Intensive Text Processing with MapReduce", Morgan & Claypool Publishers, 2010.
- [11] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "GATE: A Framework and Graphical Development Environment for Robust NLP

- Tools and Applications,” in Proc. of the 40<sup>th</sup> Anniversary Meeting of the Association for Computational Linguistics, ACL ‘02, Philadelphia, 2002.
- [12] C. Zhang, H. Wang, Y. Liu, D. Wu, Yi Liao and Bo Wang, “Automatic Keyword Extraction from Documents Using Conditional Random Fields”, *Journal of Computational Information Systems*, 2008.
  - [13] Y. Matsuo and M. Ishizuka, “Keyword extraction from a single document using word co-occurrence statistical information”, *International Journal on Artificial Intelligence Tools*, 2004.
  - [14] J. Kaur, V. Gupta, “Effective Approaches For Extraction Of Keywords”, *IJCSI International Journal of Computer Science Issues*, Vol. 7(6), pp. 144-148, November 2010.
  - [15] I. Witten, G. Paynte, E. Frank, C. Gutwin and C. Nevill-Manning, “KEA: practical automatic keyphrase extraction”, in Proc. of the 4<sup>th</sup> ACM Conference on Digital Library, 1999.
  - [16] C. Wu, M. Marches, J. Jiang, A. Ivanyukovich and Y. Liang, “Machine Learning-Based Keywords Extraction for Scientific Literature”, *Journal of Universal Computer Science*, Vol. 13(10), pp. 1471-1483, 2007.
  - [17] Z. Liu, P. Li, Y. Zheng and M. Sun, “Clustering to find exemplar terms for keyphrase extraction”, in Proc. of the 2009 Conf. on Empirical Methods in Natural Language Processing, pp. 257-266, 2009.
  - [18] F. Liu, D. Pennell, F. Liu and Yang Liu, “Unsupervised approaches for automatic keyword extraction using meeting transcripts”, in Proc. of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 620-628, 2009.
  - [19] O. Medelyan, E. Frank and I. H. Witten, “Human-competitive tagging using automatic keyphrase extraction” in Proc. of the 2009 Conf. on Empirical Methods in Natural Language Processing, pp. 1318-1327, 2009.
  - [20] K. S. Hasan and V. Ng, “Automatic Keyphrase Extraction: A Survey of the State of the Art”, in Proc. of the 52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics, Vol. 1, pp. 1262-1273, 2014.
  - [21] M. Grineva, M. Grinev and D. Lizorkin, “Extracting key terms from noisy and multitheme documents”, in Proc. of the 18<sup>th</sup> Int. Conf. on World Wide Web, pp. 661-670, 2009.
  - [22] Z. Liu, W. Huang, Y. Zheng and M. Sun, “Automatic keyphrase extraction via topic decomposition”, in Proc. of the 2010 Conf. on Empirical Methods in Natural Language Processing, pp. 366-376, 2010.
  - [23] M. Chung and D. I. Moldovan, “Parallel Natural Language Processing on a Semantic Network Array Processor”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 7(3), pp. 391-404, 1995.
  - [24] M. P. van Lohuizen, “Parallel Processing of Natural Language Parsers”, in Proc. of the 15<sup>th</sup> Conf. of Parallel Computing, pages 17-20, 2000.
  - [25] P. Jindal, D. Roth and L.V. Kale, “Efficient Development of Parallel NLP Applications”, Tech. Report of IDEALS (Illinois Digital Environment for Access to Learning and Scholarship), 2013.
  - [26] N. Rizzolo and D. Roth, “Learning Based Java for Rapid Development of NLP Systems”. In Proc. of the International Conference on Language Resources and Evaluation (LREC), 2010.
  - [27] L. V. Kale and G. Zheng, “Charm++ and AMPI: Adaptive Runtime Strategies via Migratable Objects”, in *Advanced Computational Infrastructures for Parallel and Distributed Applications*, pp. 265-282, Wiley Interscience, 2009.
  - [28] Exner, P. and Nugues, P., “KOSHIK - A Large-scale Distributed Computing Framework for NLP”, in Proc. of the International Conference on Pattern Recognition Applications and Methods (ICPRAM 2014), pp. 463-470, 2014.
  - [29] V. Tablan, R. I. Cunningham and K. Bontcheva, “GATECloud.net: a platform for large-scale, open-source text processing on the cloud”, *Philosophical Transactions of the Royal Society*, 2013.
  - [30] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker and I. Stoica, “Spark: Cluster Computing with Working Sets”, Technology report of UC Berkeley, 2011.
  - [31] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma; M. McCauly; M. J. Franklin, S. Shenker and I. Stoica, “Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing”, in Proc. of the 9<sup>th</sup> USENIX Symposium on Networked Systems Design and Implementation (NSDI 12), pp. 15-28, 2012.
  - [32] S. Gopalani and R. Arora, “Comparing Apache Spark and Map Reduce with Performance Analysis using K-Means”, *International Journal of Computer Applications*, Vol. 113(1), pp. 8-11, 2015.

# Discovery and registration of components in multimodal systems distributed on the IoT

B. Helena RODRIGUEZ  
W3C's MMI Working Group Editor  
Paris, FRANCE  
helena.rodriguez@shopedia.fr

Jean-Claude MOISSINAC  
Institut Mines-Telecom  
Telecom ParisTech CNRS LTCI  
46, rue Barrault 75634 Paris CEDEX 13  
moissinac@telecom-paristech.fr

**Abstract**— One of the major gaps in the current HTML5 web platform is the lack of interoperable means for an application to discover services and applications available in a given space and network. This problem is shared by the multimodal applications developed with web technologies, for example, in smart houses or applications for the Internet of Things. To address this gap, we produced a SOA approach for the W3C's Multimodal Working Group that aims to allow the discovery and registration of components used in multimodal interaction systems in the web of things. In this approach, the components are described and virtualized in a dedicated module communicating with two dedicated events, and registering components in a Resources Manager to facilitate the fine management of concurrent multimodal interactions, and the interoperable discovery, registration and filtering of features provided by heterogeneous and dynamic components in the web of things.

**Keywords**—component; Multimodal Interaction, Semantic Interaction, Interface Services, MMI Architecture and Interfaces, Pervasive computing, Context awareness

## I. INTRODUCTION

The Multimodal Architecture and Interfaces (MMI-Arch) is a current Recommendation of the World Wide Consortium [1] introducing a generic structure and a communication protocol to allow the components in a multimodal system to communicate with each other. It proposes also a generic event-driven architecture and a general frame of reference focused exclusively in the control of the flow of data messages. This frame of reference has been proposed due to a lack of distributed approaches for multimodal systems “in the cloud”. These approaches are mostly produced in ad-hoc solutions, as shown by a state of the art of 100 relevant multimodal systems where it was observed that more than 97% of the systems had little or no discovery and registration support [2].

At the time, even the W3C's MMI Architecture and Interfaces (MMI-Arch) and its runtime framework [3] failed to address: 1) the component's discovery and registration to support fusion (integration) and fission (composition) mechanisms, 2) the modality component's data model needed by this registry and 3) the modality component's annotation to facilitate the orchestration (and even the turn-taking) mechanism.

These three issues are addressed and to some extent resolved by our SOA proposal, which is now adopted as a W3C's recommendation.

Thus, our proposal has become an interoperable extension for the MMI-Arch's model, designed to support the automation of the discovery, registration and composition of multimodal semantic services. It is also designed to fulfill the requirements of high-level Quality of Service (QoS) like: the accurate selection of components when these are not available anymore, do not meet the expected functionality or disrupt the context of use.

With these goals in mind, our contribution was structured on three parts: 1) a new addressing method needed for the component's announcement at bootstrapping; 2) an architectural extension in order to support the handling of the state of the multimodal system using a virtual component approach for registration and 3) two new events for the messaging mechanism, to address the requirements of discovery and registration on distributed systems.

These three parts are currently completed by the creation of a common and interoperable vocabulary of states and generic features to allow the gross

discovery of modalities in large-networks over a concrete networking layer [4]

In the following sections we will present our contribution as follows: In Sec. 2 we will give an overview of the problem, followed by a study of the related work in Sec 3. In Sec. 4 we describe our work on Discovery and Registration and finally, we present a conclusion and some perspectives to continue this work.

## II. PROBLEM STATEMENT

Historically, multimodal systems were implemented in stable and well-known environments. Its complexity demanded laboratory-like implementation and very few experiences were developed for real-time contexts or component distribution. But this situation has evolved. The web developer's community is progressively confronted with the problem of modality integration in large-scale networks which is expected to be huge in the years to come, when the Web of Things will attain a state of maturity.

The increasing amount of user-produced and collected data will also require a more dynamic software behavior with a more adequate approach 1) to handle the user's technical environment where the demand for energy supply is getting higher and higher, and 2) to encourage and improve the efficiency in consumption boosting the creation of systems compatible with smart-grid technologies.

For example, in Japan [5] (as in European countries) the distributed applications will play a very strategic role in the reduction of energy consumption, helping to evolve to an on-demand model. With this goal, the sustainable consumption in houses must be handled and analyzed distantly, using data collected by multimodal applications installed on multiple kinds of devices of the Internet of Things.

These applications must interact in a coordinated manner in order to improve the energetic efficiency of the application behavior, to collaborate in the home automation management and in some cases, even the user profiling; the whole with a distributed platform.

Thus, we face the raising of multiple issues, concerning the multimodal user interaction with very heterogeneous types of devices (some of them

with low resources), protocols and messaging mechanisms to be synchronized in an interoperable way.

In this context, on one side, modality discovery and selection for distributed applications becomes a new working horizon giving new challenges for multimodal systems, user-centric design and the web research. And in the other side, generic and interoperable web approaches, using web technologies but capable of going beyond the browser model, will be unavoidable.

### A. Multimodal Discovery and Registration

Multimodal systems are computer systems endowed with rich capabilities for human-machine interaction and able to interpret information from various communication modes. According to [6] the three principal features of multimodal systems are: 1) the fusion of different types of data; 2) real-time processing and temporal constraints imposed on information processing; 3) the fission of restituted data: a process for realizing an abstract message through output on some combination of the available channels.

On these systems, modality management is mostly of the time hard-coded, leaving aside the problem of a generic architecture respond to extensibility issues and the need of discovery, monitoring and coordination of modalities in real-time with context-awareness. Consequently, multimodal applications were manually composed by developers and shared via web APIs and embedded web technologies, in an ad-hoc and proprietary way.

To address this lack of a generic approach, the MMI Architecture proposes an architectural pattern for any system communicating with the user through different modalities simultaneously instantiated in the same interaction cycle. In this unique context of interaction the final user can dynamically switch modalities. This kind of bi-directional system combines inputs and outputs in multiple sensorial modes and modalities (e.g. voice, gesture, handwriting, biometrics capture, temperature sensing, etc) and can be used to identify the meaning of the user's behavior or to compose intelligently a more adapted, relevant and pertinent message.

Other important characteristic of the Multimodal Architecture and Interfaces specification is that it



uses the MVC design pattern generalizing the View to the broader context of the multimodal interaction presentation, where the information can be rendered in a combination of various modalities. Thus, the MMI recommendation distinguishes (Fig. 1) three types of components: the Interaction Manager, the Data Component and the Modality Components.

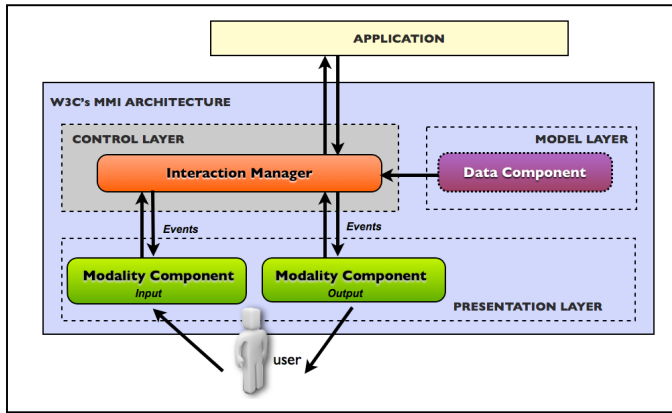


Figure 1. The W3C's Multimodal Architecture

The Interaction Manager is a logical component responsible of the integration and composition of the interaction cycles following multimodal rules. It handles all the exchanges between the components of the Multimodal System and the hosting runtime framework. To ensure some of these tasks, the Interaction Manager must access the data stored in a second component representing the Model, the Data Component, which is a logic entity that stores the public and private data of any module or the global data of the System.

Finally, in the MMI Architecture, the term Modality covers the forms of representing information in a known and recognizable rendered structure. For example, acoustic data can be expressed as a musical sound modality (e.g. a human singing) or as a speech modality (e.g. a human talking).

The component representing the presentation layer in the MMI Architecture is indeed, a Modality Component. This is a logic entity that handles the input and output of different hardware devices (e.g. microphone, graphic tablet, keyboard) or software services (e.g. motion detection, biometrics sensing).

Modality Components are also loosely coupled software modules that may be either co-resident on a device or distributed across a network. This aspect

promotes low dependence between Modality Components, reducing the impact of changes and facilitating their reuse. In result, these components have little or no knowledge of the functioning of any other modules and the communication between modules is done through the exchange of events following a protocol provided by the MMI architecture.

Nevertheless, the architecture focuses only on the interaction cycle, leaving aside 1) the support of the lifecycle of Modality Components -from a system perspective-, 2) a more dynamic application behavior, and 3) the non- functional goals of some features to adapt the application to a particular space, device family and interaction type, using context-aware techniques.

As a result, we decided to extend the architecture's model in our work with the MMI Working Group in order to address the lack of an interoperable frame of reference to handle the runtime system's lifecycle that includes the dynamical discovery and registration of Modality Components, as we will see on Sect. 4.

### B. Beyond the Browser

Today, there is an enormous variety and quantity of devices interacting with each other and with services in the cloud: 7 trillion wireless devices will be serving 7 billion people in the 5 years to come [7]. Web technologies are expected to be at the center of the Internet of Things (IoT), thanks to their universal adoption and huge scalability. Nevertheless the definition of a standardized programming model for objects beyond the page-browser mechanism has not been established yet, and the classical internet of documents or the internet of knowledge has being built with a series of architectural premises that could be inadequate and even a foundational obstacle to this new challenge.

To address it, device-centric technologies, proposals and protocols are spreading all over the current discussion around the Web of Things, assuming that the "infinite things problem" will be resolved by creating "virtual images" of this reality on IoT systems. But this solution will just transfer a real-world problem to a virtualized one, with the concurrency of policies, architectures, platforms, protocols and standards that such a transfer implies.

On one hand, browser vendors are advocating for browser-based solutions assuming that a model that works well for web pages (based on the document model) and web apps (mostly based on client-server models) in computers and mobiles, can be easily extended to any other kind of objects.

But, how can we model a rice cooker as a document? Is it really logical to communicate with an air conditioner as a “data resource”? How to apply a client-server model to reflexive objects in the network, acting at the same time as server and clients of their own services? Are addresses registers of devices as stable as the addresses directories of web pages or web apps? How to express on, off or stand-by states with web technologies? And what will be the environmental and energetical price to this choices?

On the other hand device vendors are advocating for energy efficient and lightweight protocols fine-tailored for constrained devices; and willing to provide a web gateway to allow the communication between these devices and the web. After near 20 years of research, some of the industrial consortiums leaded by energy providers and device vendors built a series of low-level protocols and technologies supported by national policies: KNX [8], ZigBee SEP 2.0 [9], Z-Wave [10], Echonet [11], ETSI M2M [12], DLNA [13], UPnP [14], ZeroConf [15], etc.

As the above list shows, these concurrent protocols and technologies have to be evaluated and selected by a developer or a new device producer. If this panorama of exploded technologies continues, the situation that mobile developers endured during years will reappear in the web of objects: heterogeneous operating system, SDK's, app distribution circuits and developing models for an infinity of objects.

To sum up, there is a real and urgent need of a vendor-agnostic model of components and communication, to encompass the diversity of proposals and technologies in the Internet of Things, and the need of generic devices to reduce the effort of implementation for developers and app vendors.

Our effort in the MMI Working Group, has been always focused to evolve Web technologies from device-centric applications, to natural interaction experience and user-centered models that will extend the definition of an application to seamlessly

encompass multiple heterogeneous devices collaborating and sharing resources and computational capabilities, both locally and across the web.

As an illustration of the problem, in a multimodal system devices may contain nested logical devices, as well as functional units, or services. A functional categorization of devices is currently defined by the UPnP protocol with 59 standardized device templates and a generic template profile, the Basic Device. With the same spirit, the Echonet consortium defines a number of 7 Device Groups for multiple Classes of Devices while Zigbee SEP 2.0 defines 20 device categories. In the three cases, the device specification defines explicitly the device's properties and access methods. In contrast, more generic protocols, like Z-wave use the Generic device approach and 3 abstract classes of Devices. Device classifications are provided also by The Composite Capability Preference Profiles Specification of the W3C or even with the User Agent Profile Specification extension of CC/PP [16] maintained by the Open Mobile Alliance (formerly the WAP Forum) with the Specification's Part 7: Digital Item Adaptation, in which Terminals and Terminals capabilities are described.

It is also possible to leverage the current work of the W3C's Device APIs Working Group [17], which is working on a set of heterogeneous deliverables going from the device object level to very specific features, browser extensions, HTML5 extensions and event networking issues: Vibration API, Battery Status API, HTML Media Capture, Proximity Events, Ambient Light Events, Media Capture and Streams, MediaStream Image Capture, Media Capture Depth Stream Extensions, Network Service Discovery (HTTP-based services advertised via common discovery protocols within the current network), Wake Lock API, Menu API and the sensor API to come.

This example showing the device description proposals, illustrates the concurrency of concerns, approaches and proprietary interests around the “thing” indexing and registration problem. This work can be made more extensible and less driven by the specific capabilities of today's mobile devices by aligning it with the generic, device-independent Multimodal Interfaces API. It would also be very useful to integrate these proposals with the

taxonomic efforts already made by consortia like Echonet during the last 20 years in a common and standardized vocabulary and generic API.

We can imagine that the horizon opened by the web of things is as exponential as the technical solutions currently available. This situation explains and supports the MMI Working Group generic approach and, as we will present on the following sections, defines our proposal for discovery and registration of Modality Components.

### III. OVERVIEW OF RELATED WORK

In a previous work [2] we studied a sample of 16 multimodal architectures that were selected from a previous analysis of a larger set (100) of multimodal implementations. The selection criteria has been the amount of information provided by the authors about the architectural facets of the implementation, its completeness and its representativeness of three domains of research: distribution, the modality description and the use of semantic technologies.

In the following section we will present a first group of emerging trends directly related to the criteria of discovery and registration, and later, a second group transversal to the same criteria.

#### A. Emerging trends related to the criteria

##### 1) EVENT HANDLING.

The first recurrent topic is event handling. Seven architectures tried to address the management of events, which is normal in the human computer interaction research because user interfaces are highly event-oriented.

The event management concerns are resolved with seven different techniques. In OAA [18], triggers provide a general mechanism to express conditional requests. Each agent in the architecture can install triggers either locally, on itself, or remotely on its facilitator or peer agents. There are four types of triggers: communication triggers; data triggers; task triggers; and time triggers.

GALATEA [19] uses macro-commands while an Agent Manager that possesses a macro-command interpreter expands each received macro-command in a sequence of commands and sending them sequentially to the designated modules. With task control layers in OPENINTERFACE [20], communication paradigms (event-based, remote procedure call, pipe, etc) are implemented with

adapters/connectors using rules for instantaneous events and persistent events. In MEDITOR [21], events are handled with three specialized managers: the input messages queue, the input messages generator and the output messages generator. The temporal order is ensured and disambiguation is handled with a routing table and predefined rules. Hardwired Reactions are the tool in REA [22], for quick reactions to stimuli. These stimuli then produce a modification of the agent's behavior without much delay, as predefined events.

In DIRECTOR [23] events are handled at the level of pipeline execution –continuous- and at the level of scripting –discrete-. In HEPHAISTK [24], events are handled by the Event Manager, which ensures the temporal order of events. The client application is a client, but also is another input source, and consequently the Event Manager is needed also as a recognition agent, which communicates through a set of predefined messages.

In contrast, the MMI Architecture responds to the same concern with the Interaction Life-Cycle Events, and the proposal of a dedicated component: the Interaction Manager. This solution provides a clear separation between the interaction control and the interaction content data, but hardwired mechanisms are not envisioned, neither the transport queue mechanism implemented in MEDITOR, GPAC [25] and HEPHAISTK that can be an important support for the fusion / fission of modalities. In consequence, these mechanisms were detected as possible extensions to the W3C's Architecture to provide some complementary resources to handle multimodal events in an interoperable way.

##### 2) STATE MANAGEMENT

The second key topic, recovered from 5 of the sampled architectures is the state management. It corresponds also to the session management. This feature is oriented to register the evolution of the interaction cycle and provides the information about any modification of the state of the system and the components. It is designed as a monitoring process in support of the decision layer (SMARTKOM [26], HEPHAISTK), as a display list manager in support of the fusion and fission mechanisms (DIRECTOR), as a blackboard (OAA, HEPHAISTK), a central place where all data coming from the different

sources are standardized, and other interested agents can dig them at will.

Finally, the states are handled by an object manager -for decoding and rendering purposes- (GPAC), and even as a routing table (MEDITOR). Concerning this subject the MMI Framework recommends a specific component to handle the multimodal session and the state of components; yet, it does not give details about the interfaces needed to use this component or about its role in the management of the interaction cycles. As a result, an extension to the MMI Architecture can be conceived to complete this generic description with specific details about the eventual implementation, behavior and responsibilities of this state manager.

## *B. Emerging trends transversal to the criteria*

### *1) GENERIC MODELS*

The first transversal key topic is the definition of models: 12 of our architectures proposed interesting approaches concerning the modeling of the entities that participate in the multimodal interaction. However, only SMARTKOM addresses the modeling task with a proposal coming from web semantic technologies.

In addition, depending on the modeled entity, the models are more or less expressive or homogeneous, and consequently, usable. The modeling of the multimodal interaction phenomenon (SMARTKOM, HEPHAISTK, MEDITOR), the task (GALATEA, OPENINTERFACE, SQUIDY [27]), the dialog interaction (REA, GALATEA, SMARTKOM), and the devices (SMARTKOM) is more extensive, tested and advanced than the modeling of the user (REA), the application (OAA, SMARTKOM, ELOQUENCE, GPAC, HEPHAISTK) or the environment & context of usage (SMARTKOM) conceived to support and enrich the multimodal interaction.

This growing and common interest on models -expressed in SMARTKOM as a foundational principle, opens the way to reinforce the MMI recommendation with an effort to address this issue and to see how the MMI Framework & Architecture can respond to data modeling needs.

### *2) DISTRIBUTED ARCHITECTURES*

The second transversal topic is distribution. It is tackled with solutions like the remote installation of triggers (OAA), the distribution of the fusion-fission

mechanisms into nodes and components (OPENINTERFACE, SQUIDY) that can even be external to the multimodal system, the management of inputs as “sensed” data (input sensors) or as broadcasted media containing behavior (and interaction) information in the distributed streams (GPAC); and finally, the distribution of application services (SMARTKOM, HEPHAISTK). This topic is also reflected on the service-oriented proposals of application services and services advertisement (OAA, SMARTKOM) and the networking services layer to manage the broadcasted input and output data of a rich application (GPAC). The MMI Framework & Architecture reflects this topic in its distributed nature based on web standards. Nevertheless, there are few current implementations using the web services or a service-oriented approach from a distributed perspective.

The current implementations are oriented to prototype mobile interfaces (Orange Labs), to provide a multimodal mobile client for health monitoring (Openstream), to test an authoring tool (Deutsche Telekom R&D) and to complete JVoiceXML, an open source platform for voice interpretation (TU Darmstadt). We believe that it is possible that interesting extensions arise from a fully SOA implementation of the MMI Framework & Architecture standard according with its distributed nature and the needs that are appearing with the Internet of Things.

### *3) CONTROL DELEGATION*

A final transversal topic is the delegation of the interaction management by a client application. It is present in the form of application agents (CICERO, OAA) or application services (SMARTKOM, HEPHAISTK). The MMI Framework & Architecture does not deal with this subject because the application is meant to be the concrete implementation of the architecture. A delegation approach supposes that an external functional core can delegate the management of the interaction to a multimodal system built in accordance with the standard, and providing multimodal functionalities to the client application installed on devices with low processing capabilities.

This approach is not currently addressed, even if it could be the type of requirement of a multimodal browser, a home gateway virtualization or an IoT web application. Our current work on the W3C

MMI Working Group addresses the possible extensions that such approach could bring and how the MMI Architecture standard can support this type of future implementation.

In short, the study of the related work allow us to structure and define our collaboration in the W3C Multimodal Working Group, to extend the MMI Architecture with a proposal oriented to facilitate the distributed implementations coming from the Internet of Things.

#### IV. DISCOVERY & REGISTRATION FOR MMI SYSTEMS

To the best of our knowledge, there is no standardized way to build a web multimodal application that can dynamically combine and control discovered components by querying a registry build based on the modality states. At the same time -as we showed in Sect. 3 - research efforts also lack of this distributed perspective. Based on this previous analysis, we decide to focus on three complementary extensions to the MMI Architecture: 1) we propose to complete the current addressing method in order to evolve from a client-server model to an anycast model. 2) We propose to reinforce the management of the “multimodal session”, and more precisely, a dedicated component to handle the system’s state and support the system’s virtualization of components. And 3) we propose to extend the transport layer with two new events designed to complete and reinforce the interaction Lifecycle Events.

##### A. Extending the MMI addressing methods

To inform the system about the changes in the state of the Modality Components, an adaptive addressing mechanism is needed. We consider that the combination of push/pull mechanisms is crucial to extend the MMI Architecture to the Web of Things. For example, in the case of the unavailability of a given Modality Component, it needs to communicate with the control layer. This situation is not necessarily related to the interaction context itself, but it can affect it, because the interaction cycle can be stopped or updated according to this change on the global state of the system.

In the current state of the Multimodal Architecture Specification [1], interaction events like Prepare or Start, must be triggered only by the

Interaction Manager and sent to the Modality Components. In result, a Modality Component cannot send messages to the Interaction Manager other than the message beginning the interaction cycle: the newContext event. Any other event originated by an internal command or like in our example, by a change on the component’s state cannot be raised. Nevertheless, to start an interaction cycle the Modality Component needs to be already part of the system and to be registered. The registration process is part of a previous phase, when even the presence of the user is not mandatory and the communication must be bidirectional.

As Modality Components are reflexive objects in the network acting at the same time as server and clients, they need to communicate and to receive messages as well. The flow of messages always initiated by the Interaction Manager is not sufficient to address use cases evolving in dynamic environments, like personal externalized interfaces, smart cars, home gateways, interactive spaces or in-office assistance applications. In all these cases, Modality Components enter and quit the multimodal system dynamically, and they must declare their existence, availability and capabilities to the system in some way.

To address this need, we proposed our first extension, which is a bidirectional flow of messages to support a complete number of addressing methods and to preserve a register of the system’s global state. One of the results of this new flow of messages is the capability to produce the advertisement of Modality Components. It allows the Multimodal System to reach correctness in the Modality Components retrieval and also affects the completeness in the Modality Component retrieval. To return all matching instances corresponding to the user's request, the request criteria must match some information previously registered before the interaction cycle starts.

For this reason, the MMI Architecture should provide a means for multimodal applications to announce the Modality Component’s presence and state. This was the first step to address the distribution requirement: Modality Components can be distributed in a centralized way, a hybrid way or a fully decentralized way.

For the Discovery & Registration purposes the distribution of the Modality Components influences how many requests the Multimodal System can handle in a given time interval, and how efficiently it can execute these requests. Even if the MMI Architecture Specification is distribution-agnostic, with this extension Modality Components can be located anywhere and communicate their state and their availability to new a dedicated component: the Resources Manager.

#### *B. Extending the MMI Architecture's modules*

The new flow of messages between the Modality Components and the control layer needed a mechanism tracing the relevant data about the session and the system state. This is the first of the responsibilities for the second extension, the Resources Manager. This manager is responsible for handling the evolution of the “multimodal session” and the modifications in any of the participants of the system that could affect its global state. It is also aware of the system's capabilities, the address and features of modalities, their availability and their processing state. Thus, the Resources Manager is nested in the control layer of the multimodal system and keeps the control of the global state and resources of the system. And the extended control layer encompasses the handling of the multimodal interaction and the management of the resources on the multimodal system. In this way, with our extension, the architecture preserves its compliance with the MVC design pattern.

The data handled by the Resources Manager can be structured and stored in a virtualized manner. In this way, the Resources Manager can be calibrated for mediated discovery -and federated registering-. The Resources Manager uses the scanning features provided by the underlying network, looking for components tagged in their descriptions with a specific group label. If the discovered component is not tagged with a group label, the Resources Manager can use some mechanism provided to allow subscriptions to a generic group. In this case, the Modality Component should send a request using the new flow of messages and using one of the new discovery events to the Resources Manager, subscribing to the register of the generic group.

In this way, the Resources Manager translates the Modality Component's messages into method calls on the Data Component, like the MVC pattern

proposes but also, the Resources Manager broadcasts to the Modality Component the changes on the system's state or notifies it following a subscription mechanism. Upon reception of the notification, the Modality Component updates the user interface according to the information received.

The Resources Manager supports the coordination between virtualized distributed agents and their communication through the control layer. This enables to synchronize the input constraints across modalities and also enhances the resolution of input conflicts from distributed modalities. It is also the starting point to declare and process the advertised announcements and to keep them up to date and the core support for mediated and passive discovery and it can also be used to trigger active discovery using the push mechanism or to execute some of the tasks on fixed discovery [4]. The Resources Manager is also the interface that can be requested to register the Modality Component's information. It handles all the communication between them and the registry. The flow of discovery queries transit through it, which dispatches the requests to the Data Component and notifies the Interaction Manager if needed. These queries must be produced using the state handling events presented on the next Section.

To summarize, the Resources Manager delivers information about the state and resources of the multimodal system during and outside the interaction cycle.

#### *C. Extending the MMI Event model*

With a new flow of messages and a new component handling the state of the system, a Modality Component can register its services for a specific period of time. This is the basis for the handling of the Modality Component's state. Every Modality Component can have a lifetime, which begins at discovery and ends at a date provided at registration. If the Modality Component does not re-register the service before its lifetime expires, the Modality Component's index is purged. This depends on the parameters given by the Application logic, the distribution of the Modality Components or the context of interaction.

When the lifetime has no end, the Modality Component is part of the multimodal system indefinitely. In contrast, in more dynamic



environments, a limited lifetime can be associated with it, and if it is not renewed before expiration, the Modality Component will be assumed to no longer be part of the multimodal system. Thus, by the use of this kind of registering, the multimodal system can implement a procedure to confirm its global state and update the «inventory» of the components that could eventually participate in the interaction cycle. Therefore, registering involves some Modality Components' timeout information, which can be always exchanged between components and, in the case of a dynamic environment, can be updated from time to time. For this reason, a registration renewal mechanism is needed. We proposed a registration mechanism based on the use of a timeout attribute and two new events: the checkUpdate Event and the UpdateNotification, used in conjunction with an automatic process that ensures periodical requests.

The checkUpdate Event is provided a) to verify if there are any changes in the system side; b) to recover the eventual message; c) to adapt the request timeout if needed and d) to trigger automatic notifications about the state of the Modality Component, if the automaticUpdate field in the response is true. If a Modality Component is waiting for some processing provided by other distributed component, the checkUpdate Event allows the recovery of progressive information and the fine-tuning of requests by changing the timeout attribute. This enhances input/output synchronization in distributed environments.

On the other hand, the Update Notification is proposed a) to periodically inform the Resources Manager about the state of the Modality Component; b) to help in the decision making process (on the server side, for example).

For notification of failures, progress or delays in distributed processing the Update Notification (Fig. 4) ensures periodical requests informing other components if any important change occurs in the Modality Component's state. This can support, for example, grammar updates or image recognition updates for a subset of differential data (the general recognized image is the same but one little part of the image has changed, e. g. the face is the same but there is a smile)

The use of the timeout attribute helps in the management of the validity of the advertised data. If

a Modality Component's communication is out-of-date, the system can infer that the data has the risk of being inaccurate or invalid. The checkUpdate Event allows the recovery of small subsets of the information provided by the interaction manager, to maintain up to date the data in the Modality Components as in the Resources Manager.

## V. CONCLUSION

The work on standardization produced by the MMI Working Group in the last two years and its focus on distribution has been fruitful.

Today, the first step needed to allow the component's discovery and registration helping for a more adaptive fusion and fission mechanisms is done. The Components have now a means to announce its capabilities and states through different addressing modes. Second, the modality component's data model needed as a building block for a multimodal registry is founded, starting by a common taxonomy of generic states (out of the scope of this document, but available in [5]) (Fig.8) and the construction of a generic classification system for devices and groups of modes and modalities. This premises of classification will allow facilitate the orchestration mechanism with the Modality Component's annotation. A mechanism that is now possible, thanks to the extension of the MMI Architecture's event model with two events specified for discovery and registration needs.

These three issues are covered by our results and now are entering on the W3C's recommendation processes to be available to the community of web developers. From the requirements extracted from the analysis of the state of the art and a series of use cases provided by the industry [4], we produce three pertinent extensions.

To handle multimodal events (See Sect. 3-A-1) in an interoperable way, we extend the MMI Architecture by completing the current addressing method in order to evolve from a client-server model to an anycast model using bidirectional communication.

To ensure the handling of states (See Sect. 3-A-2), we proposed to support the management of the «multimodal session» by a dedicated component using a virtualization of components to reflect the current state of the system.

To allow distribution (See Sect. 3-B-2), we proposed to extend the transport layer with two new events completing and reinforcing the interaction Lifecycle Events. An finally, to support the delegation of control (See Sect. 3-B-3) and to use generic models (See Sect. 3-B-1), we proposed a virtualization mechanism used to create and store the registry, based on generic multimodal properties, a generic model of states needed for keeping the registry up-to-date and a generic vocabulary for the description of Modality Components.

The MMI's Modality Component is an abstraction flexible enough for any implementation of the Internet of Things and networking model, while keeping an interoperable structure. The MMI Architecture is built around the management of continuous media and their states not only as outputs (presentations) but also as inputs. This means that the architecture is fine-tuned to handle issues derived from very dynamic environments needing session control and recovering with all kinds of medias and interaction modes.

In this paper we presented our current work on Discovery and Registration of Modality Components from a generic and interoperable technology that will allow us to face the infinity created by the web of things. From an extensive study of the state of the art, we produced a series of requirements and evaluation criteria that founds the proposal presented on Sect. 4, which is now a W3C's Recommendation.

In a future activity the W3C working group will produce an annotation vocabulary and the support of the semantic annotation in the "info" dedicated attribute on the new discovery and registration events. This vocabulary is a first step on the direction of a more expressive annotation of the interaction with Modality Components using ontologies and a more intelligent composition of semantic web services for multimodal applications with rich interaction features.

- [1] MMI-Arch: <http://www.w3.org/TR/mmi-arch/> Visited at :01/04/2015
- [2] B. H. Rodriguez. "A SOA model, semantic and multimodal, and its support for the discovery and registration of assistance services". PhD Thesis, Institut Mines-Télécom, Telecom ParisTech, Paris, 2013.
- [3] Multimodal Interaction Framework <http://www.w3.org/TR/mmi-framework/> Visited at :01/04/2015
- [4] B.H. Rodriguez (Ed)., D.Dahl, R. Tumuluri, P. Wiechno and K. Ashimura. Registration & Discovery of Multimodal Modality Components in Multimodal Systems: Use Cases and Requirements. W3C Working Group Note 5 July 2012. Available at: <http://www.w3.org/TR/mmi-discovery/> Visited at :01/04/2015
- [5] International Symposium on Home Energy Management System -Joint discussion with the W3C MMI WG -, Keio University Shonan Fujisawa Research Institute, 25-26 Feb., 2015
- [6] J. Coutaz, L. Nigay, D. Salber, A. Blandford, J. May and R. Young "Four easy pieces for assessing the usability of multimodal interaction: the CARE properties" In: Proceedings of INTERACT'95Lillehammer, June 1995
- [7] W3C Workshop on the Web of Things Enablers and services for an open Web of Devices, 25-26 June 2014, Berlin, Germany <http://www.w3.org/2014/02/wot/>
- [8] KNX network communications protocol for intelligent buildings (EN 50090, ISO/IEC 14543) <http://www.knx.org/knx-en/index.php>
- [9] The Smart Energy Profile 2 <http://www.zigbee.org/zigbee-fordevelopers/applicationstandards/zigbeesmartenergy/>
- [10] Z-Wave <http://www.z-wave.com>
- [11] ECHONET Energy Conservation and HomecareNetwork. <http://www.echonet.gr.jp/>
- [12] ETSI Machine to machine communication. <http://www.etsi.org/technologies-clusters/technologies/m2m>
- [13] Digital Living Network Alliance. <http://www.dlna.org>
- [14] Universal Plug and Play <http://www.upnp.org>
- [15] Zero Configuration Networking <https://developer.apple.com/bonjour/index.html>
- [16] Composite Capabilities/Preference Profiles <http://www.w3.org/Mobile/CCPP/>
- [17] Device APIs Working Group. <http://www.w3.org/2009/dap/#roadmap>
- [18] D. Martin, A. Cheyer, and D. Moran, "The Open Agent Architecture: A Framework for Building Distributed Software Systems," Applied Artificial Intelligence, Volume 13, Number 1-2, January-March 1999, pp. 91-128.
- [19] NITTA. Activities of Interactive Speech Technology Consortium (ISTC) Targeting Open Software Development for MMI Systems
- [20] M. Serrano, L. Nigay, J-Y. Lawson, L. Ramsay, and S. Denef "The openInterface framework: a tool for multimodal interaction" In: CHI '08 extended abstracts on Human factors in computing systems -CHI EA08. ACM, New York, NY, USA. p.p.3501-3506.
- [21] Y. Bellik Interfaces Multimodales : Concepts, Modèles et Architectures. PhD Thesis, University Paris-South 11, Orsay, 1995.
- [22] J. Cassell "Embodied Conversational Agents. Representation and Intelligence in User Interfaces" In: AI Magazine, 2001. Vol. 22. No.4.
- [23] [http://en.wikipedia.org/wiki/Adobe\\_Director](http://en.wikipedia.org/wiki/Adobe_Director) Visited at :01/04/2015
- [24] B. Dumas, D.Lalanne, and R. Ingol. Démonstration: Hephaïstos, une boîte à outils pour le prototypage d'interfaces multimodales, 2008.
- [25] J. Le Feuvre et al., Experimenting with Multimedia Advances using GPAC, ACM Multimedia, Scottsdale, USA, November 2011 <http://dl.acm.org/citation.cfm?doid=2072298.2072427>
- [26] Herzog, Gerd and Reithinger, Norbert. "The SmartKom Architecture: A Framework for Multimodal Dialogue Systems" In: SmartKom: Foundations of Multimodal Dialogue Systems, 2006. Springer Berlin Heidelberg, p.p. 55-70. [http://dx.doi.org/10.1007/3-540-36678-4\\_4](http://dx.doi.org/10.1007/3-540-36678-4_4)
- [27] Werner A. König, Roman Rädle, and Harald Reiterer. 2009. Squidy: a zoomable design environment for natural user interfaces. In Proceedings of the 27th international conference extended abstracts on Human factors in computing systems (CHI EA '09). ACM, New York, NY, USA, 4561-4566. DOI=10.1145/1520340.1520700

# Vehicle Type Identification Based on Car Tail Text Information

Ruixue Yin, Weibin Liu  
Institute of Information Science  
Beijing Jiaotong University  
Beijing 100044, China  
e-mail: wblu@bjtu.edu.cn

Weiwei Xing  
School of Software Engineering  
Beijing Jiaotong University  
Beijing 100044, China

**Abstract**—Based on feature matching of car tail text information, a novel approach for vehicle identification is proposed in this paper. Our method creatively implements Scale-Invariant Feature Transform (SIFT) to extract distinctive invariant features from car tail text sub-images, which is a new application of SIFT. In our approach, firstly, the coordinate and the content of a vehicle's license plate are presented during the process of plate localization and recognition. Secondly, based on the plate location information, we apply a text information localization procedure which could be divided into two processes, robust localization and accurate localization. Then, a SIFT-based template matching method is provided to recognize the text information. Finally, we are able to determine whether the result conforms to the known vehicle type captured according to the plate license contents in the vehicles information file. The experimental results show a high recognition rate in acceptable time and prove the availability of vehicle type identification.

**Keywords**—Scale-Invariant Feature Transform; vehicle type identification; car tail text information; image matching

## I. INTRODUCTION

The appearance of crimes that involve vehicles disrupts the usual traffic, public security and brought great potential safety hazard to the traffic, such as vehicle theft, and fake plate vehicles. Therefore, research on intelligent traffic management has both social and economic value. The Intelligent Transport System (ITS) based on computer vision has been referred in the literature quite frequently in recent years. Many subsystems of ITS have been developed and applied all over the world. Among those various subsystems, Vehicle Manufacturer Recognition (VMR) is a crucial subject, but also rather difficult task. Currently, license plate recognition technology is widely applied in ITSs and quite matures in some ways. What's more, systems which are exclusively based on automatic license plate or logo detection and recognition lost sight of the car tail text information detection and recognition. However, it is still an open field to work due to complexity of vehicle information

and imaging conditions.

As a symbol of the car manufacturer, vehicle's logo has attracted lots of attention, and methods have already presented in existing literatures [1, 2]. In addition, there's more information which could be elicited at the rear of the car, such as the specific model and displacement of a car. There are a lot of research about the vehicle license plate and logo, whereas very few people do enough work on the detail information mentioned above.

The principal work for car tail text identification is feature matching, which is also a difficult task. To identify correspondences between two images is an important task in the computer vision; these applications include object recognition, gesture recognition, image stitching, object tracking and industrial inspection. Choosing efficient features is the initial step to reliably perform that task, even under complicated shooting situations and geometric transformations. One of the most commonly used methods to extract and represent features is SIFT, which is invariant to image rotation and scale and robust across a substantial range of affine distortion, addition of noise, and partially invariant to changes in illumination.

In this paper, a new method of vehicle type validation based on text information of car tail is proposed. The proposed method uses Scale-Invariant Feature Transforms to represent the text and is more robust than other image features, such as Hu Invariant Moment proposed by Hu [3] in 1962 and Corner Detector by Harris and Stephens [4] in 1988, when treating the complex characters, like Chinese characters. The proposed algorithm is shown to be effective and efficient, which demonstrates excellent performance on a representative and typical database collected from the real world.

The rest of this paper is organized as follows. Section II sets out the latest development and research on vehicle identification and feature matching. In Section III, we present the overall framework for vehicle type identification in our work. Section IV shows the experimental results of the system process. By testing hundreds of texts captured from rear-view images, the results show that the proposed method has a good recognition effect. Finally, the paper is concluded in Section V.

---

This research is partially supported by National Natural Science Foundation of China (No. 61370127, No.61100143, No.61473031, No.61472030), 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). The opinions expressed are solely those of the authors and not the sponsors.  
{ Corresponding author: Weibin Liu, wblu@bjtu.edu.cn }

## II. RELATED WORK

Many vehicle type identification methods have been applied nowadays. Among those frequently used tasks, various technologies have been applied. These methods have been proposed to recognize the vehicle manufacturer and model from frontal or rear views of vehicles. Vehicles are identified by extracting features, and then matching these features as templates or as a machine learning problem. The detection and matching of interest points serve as the basis for many computer vision applications; including image/video retrieval, object categorization and recognition, and 3-D scene reconstruction. A wide variety of detectors and descriptors have already been proposed in the existing literature.

The development of image matching by using local features can be traced back to the work of Moravec [5] on stereo matching using a corner detector in 1981. In 1988, Harris and Stephens improved the Moravec detector to make it more repeatable under small image variations and near edges. Harris also showed its value for efficient motion tracking and 3D structure from motion recovery [6], and the Harris corner detector has since been widely used for many other image matching tasks. These feature detectors called corner detectors are not selecting just corners, but rather an image localization that has large gradients in all directions on a predetermined scale. The Harris corner detector is very sensitive to changes in image scale, so it does not offer a good basis for matching images of different sizes.

The ground-breaking work of Lowe and Brown [7-9] showed that SIFT could extend the local feature approach to achieve scale invariance. They also described a local descriptor with more distinctive features while being less sensitive to image distortion such as translation, rotation, scaling or any combination of these. They proposed a method to use groups of interest points to compute local 2D transformation parameters. By using these different points they form the feature descriptors which are invariant to any 2D projective transformation. As to feature matching, they identify matching key from the new images, Beis and Lowe used a modification of the k-d tree algorithm called the Best-bin-first search method [10] that can identify the nearest neighbors with high probability using only a limited amount of computation. In addition to enabling robust matching, they present a scheme that each match represents a hypothesis of the local 2D transformation. Then Ballard use broad-bin Hough transform clustering [11] to select matches that could also reject outliers.

Because of the variations in the visual appearances vehicle types, developing a highly accurate vehicle recognition system in applications is still very challenging. Most recently published papers have primarily classified vehicles into different types for control of traffic flow. For example, Dlagnekov [12] used a license plate detector to define a vehicle ROI and then a SIFT matching scheme to retrieve desired vehicle recognition system. Zafar et al. [13] used a contourlet transform to extract vehicle features and then applied a 2-D linear discriminant analysis for dimensionality reduction, achieving a VMR system. Hsieh et al. [14] proposed a Symmetrical SURF (Speeded-Up Robust Features) and its application to vehicle make and model recognition based on the

front view of vehicles. The method based on vehicle type identification was proposed by [15], they use SIFT to extract global and local features from the front-view or rear-view images of vehicles. However, this method does not have good recognition accuracy between similar types of vehicles.

The above methods are all related to the vehicle logo or the vehicle face, but the vehicles with the same logo or face shape perhaps have different manufacturers, models or swept volume, most of these information though are distributed on the car tail, due to the lack of the uniform standard, it becomes quite difficult to identify these text information. The traditional character recognition methods have been presented in the past years [16-19]. Nevertheless, these methods should have the separate characters first, in a practical application, because of the complex environment, most characters with deformation, conglutination and blurring affect character segmentation directly, and what's more, the text information captured from vehicle images usually within even small area, all these causes led to traditional character recognition methods' limitation in vehicle text information recognition.

The most related work to ours is SIFT proposed by David Lowe. We described a SIFT-based vehicle car tail text information feature matching schema, whose aim is to obtain vehicle type from car tail text information captured in the rear-view vehicle images. The SIFT matching module detects and extracts keypoints in the text sub-images that are located and segmented from the source images, describes them and matches them with keypoints stored in a sample image database. Then the process is optimized by clustering the matched keypoints. The proposed method is assessed on a training set(database) containing 200 text sub-images, using a testing set containing 1200 query images.

## III. PROPOSED APPROACH

The proposed vehicle type identification (VTI) method mainly has two stages: training and classification. In the training stage, car tail text samples are collected from vehicle rear-view images of common models. Then, SIFT features are extracted from all samples. Next, feature representation is built for each sample, and is organized with the image's path into the training set. The classification stage consists of three main parts: (1) license plate recognition, (2) text area localization, (3) text information matching, and the part (2) could be divided into two sub-tasks: robust text area localization and accurate text area localization.

In this section, we discuss parts (2) and (3) in details, and for the completeness, we briefly describe part (1), which mainly follows previous work in [3, 12, 16]. In the first part, we locate the license plate area and then recognize the text content of the license plate to obtain the information of the vehicle, which has been collected in the known vehicle information file. Next, text area sub-images are intercepted from the query image by a priori knowledge. After that, SIFT features are extracted from each text area sub-image and match with the samples in the training set. The result shows whether the vehicle's model conforms to the known vehicle type that captured according to the plate license contents in the vehicles information file. The process flow diagram is shown in Fig. 1.

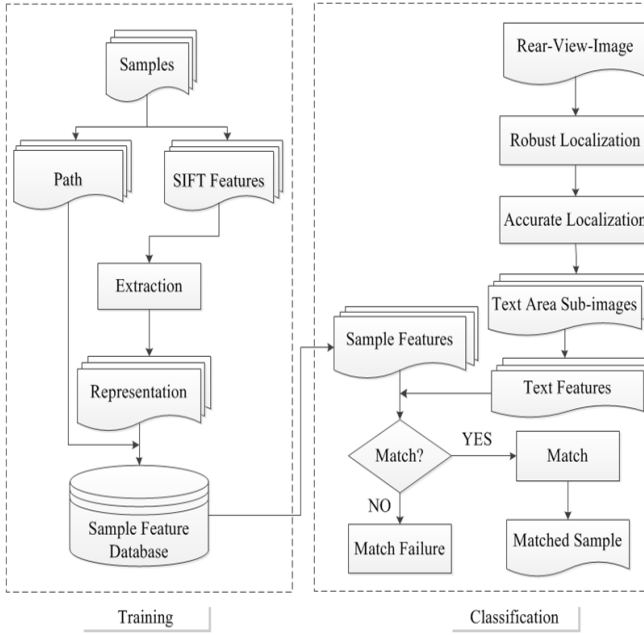


Figure 1. VTI system flowchart.

#### A. SIFT Feature Representation

Feature representation is the key issue of building a sample feature database, which distinguishes one query image from other samples. SIFT is the state-of-the-art in the field of image recognition and is used in a wide range of image retrieval applications. It exploits the idea of representing images by a set of scales and invariant keypoint descriptors using histograms of locally orientated gradients. The invariant features, or just keypoints, are detected and extracted, exploring the scale space of the image. Then the features are localized and filtered to preserve only those that are likely to remain stable over affine transformations. The process of extracting features for a SIFT keypoint descriptor encompasses four stages.

1) Detection of scale-space extrema: in this stage, Lowe [8] using the extrema in the Difference-of-Gaussian (DoG) function convolved with the image. Then all the images are detected by comparing a pixel (marked with X) to its 26 neighbors in  $3 \times 3$  regions at the current and adjacent scales. Pixel is chosen as a keypoint if its value is either minimum or maximum.

2) Accurate keypoint localization: rejecting keypoints with low contrast (sensitive to noise) or are poorly localized along an edge (DoG operator has high edge responses) to provide an improvement to matching and stability.

3) Orientation assignment: assigning a consistent orientation to each keypoint to make sure the property of rotation invariance.

4) Descriptor assignment: parameters (image location, scale and orientation to each keypoint) that have been assigned by previous operations impose a repeatable local 2D coordinate system in which to describe the local image region. Therefore, the next step is to compute a descriptor for the local

image region. Finally, a rotation-invariant descriptor called SIFT, which computes a histogram of locally oriented gradients around the interest point and stores the bins in a 128-D vector.

#### B. Feature Matching Scheme

For each feature  $i$  in the query images, SIFT matching scheme is applied on finding its Nearest-Neighbor (NN) matches among all the features stored from images  $j$  in a database. The nearest neighbor is defined as the keypoint with minimum Euclidean distance for the invariant descriptor vector:

$$NN_i = \text{cardinality}\{\arg[\|Q_i - D_i(j)\| < \gamma]\} \quad (1)$$

Where  $Q_i$  is the  $i^{th}$  descriptor for the query image,  $D_i(j)$  is the  $i^{th}$  descriptor for the  $j^{th}$  image in the database and  $\gamma$  is an appropriate threshold value. The selection of threshold ( $\gamma$ ) directly impacts the numbers of NNs in the database for each feature (keypoint). In Fig. 2, the parameter ( $\delta$ ) is plotted versus threshold ( $\gamma$ ), where  $\delta$  is described by (2):

$$\delta = \frac{NN_i}{KP_Q * KP_D} \quad (2)$$

Where  $NN_i$  is calculated from (1) for each feature  $i$  in the query image.  $KP_Q$  is the number of keypoints detected from the query image, and  $KP_D$  is the number of the keypoints in the database.

The keypoint descriptor has a 128-dimensional feature vector, therefore, an algorithm [10] similar to k-d tree (Friedman et al. [20]) called the Best-Bin-First (BBF) are used in the matching scheme to speed up search. In the implementation, we cut off search after checking the first 200 NN candidates. It returns the closest neighbor with high probability over a reasonable time frame.

There are 200 samples in the database, and each sample has about 500 or more keypoints, numbers of which may not important to the match stage. To discard more false matches arising from the background, Nearest Neighbor features are clustered using the Hough Transform [21, 22], and in order to

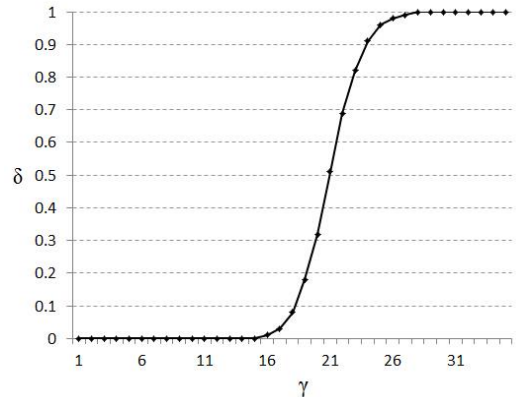


Figure 2. Reduced Nearest Neighbor parameter ( $\delta$ ) plotted versus distance threshold ( $\gamma$ ).

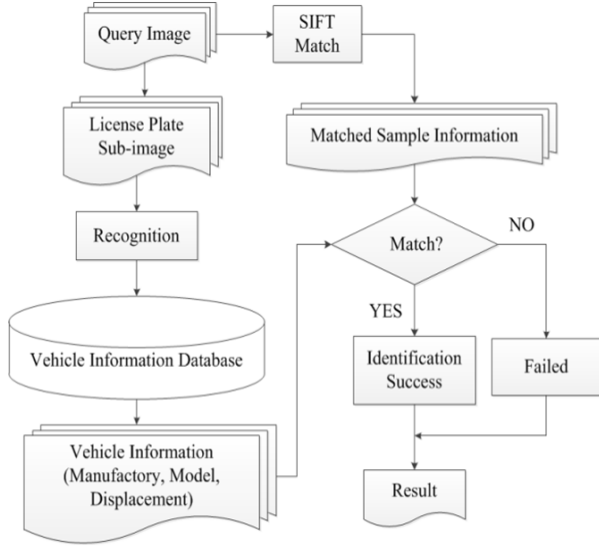


Figure 3. The verification process flowchart.

define the parameters for a similarity transformation between the query and database features. The Hough Transform identifies clusters of features with a consistent transform representation (2D location, scale, orientation and a record of the keypoints' parameters relative to the matched sample) by using each feature to vote for all the sample images that are consistent with the feature. When the clusters of the features are found to vote the same pose of the sample, the matched sample has a higher probability of the correct one than any other feature. The one which has the highest votes in the database is considered as the most possible matched result of the query image.

After previous tasks, clusters that are identified by Hough Transform are then entrance a new procedure to keep stability of geometric transformation. The affine transformation of a model point  $[x, y]^T$  to a query point  $[u, v]^T$  should satisfy the matrix relationship shown in (3):

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} p_1 & p_2 \\ p_3 & p_4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (3)$$

Where  $[t_x, t_y]^T$  is the translation value of the similarity transformation, and  $p_i (i = 1, 2, 3, 4)$  are the affine rotation, scale, and stretch parameters. Because we wish to get the transformation vector, so (3) could be rewritten to the matrix shown in (4):

$$\begin{pmatrix} u \\ v \\ \vdots \end{pmatrix} = \begin{pmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ t_x \\ t_y \end{pmatrix} \quad (4)$$

If (4) is considered as a system of  $A * x = B$ , then the parameter  $x$  can be determined by the following equation:



Figure 4. Successful license plate detection and characters segmentation result.

$$x = [A^T A]^{-1} A^T B \quad (5)$$

Then, the RANSAC method [23] is applied for all Hough Transform clusters found, and the affine transformation with the maximum number of inliers (or minimum number of outliers) is estimated. Lowe [7] reports that it is possible to have reliable recognition with as few as three feature correct matches.

### C. License Plate Recognition

The first task for the proposed VTI system is capturing the vehicle's information, and then comes the verification process; the matching process is shown in Fig. 3. In the rear-view vehicle image, the license plate region is steady-going and fixed in size [24]. Based on this property, the number plate can be fast and robust detected, then the characters on the license plate can be segmented into individual ones, as showed in Fig. 4. Next, combined features [16, 18, 19, 25] would be captured for accurate localization and character recognition. There are many methods of vehicle license plate localization and recognition that have been presented in existing literature [24-27] as yet, and need not be repeated here.

### D. Text Area Detection

The license plate is located, and using their position and size, it can also be obtained from the text areas on the car tail. The accuracy of the text area segmentation will directly affect the matching result. We segment the sub-images using a method which combines a prior knowledge with brim characteristic.

#### 1) Robust Text Area Localization

Using the license plate location and size, a robust text area could be located. Based on the priori knowledge, we know that the text area is mostly located at the car trunk and beside the taillights. As is shown in Fig. 5, positions of the ROI could be obtained.

Based on the experience, the text area could be located for the most part follow the equation:

$$\begin{cases} l_l.x = rect.x - te * \Delta w \\ l_r.x = rect.x \\ l_t.y = rect.y - ts * \Delta h \\ l_b.y = rect.y - te * \Delta h \end{cases} \quad (6)$$

Where  $(l_l.x, l_l.y)$  is coordinate of the upper left corner of the left text area sub-image,  $(l_r.x, l_b.y)$  is coordinate of the lower right corner of the left text area sub-image, and  $te$ ,  $ts$  are normalization coefficient in order to achieve a reasonable position for the text area (default  $te = 1$ ,  $ts = 2$ ). The right



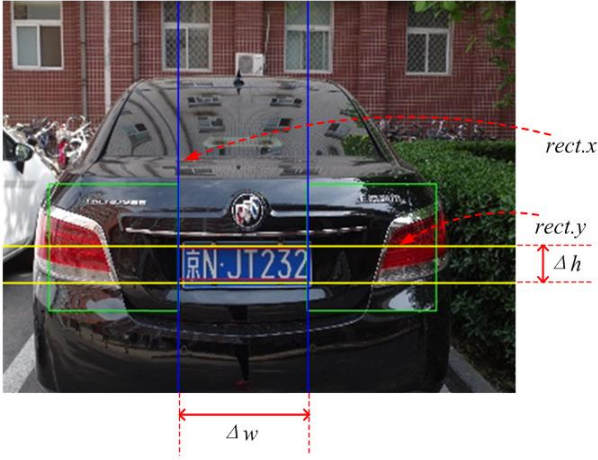


Figure 5. Notations for finding the upper and bottom boundaries of ROI for robust text areas. The parameters  $rect.x$  and  $rect.y$  are the left or top of the license plate,  $\Delta w$  and  $\Delta h$  are the width, height.

text area sub-image corresponds to a similar law, and it is no need to spell out.

## 2) Accurate Text Area Localization

After detecting the ROI, robust text areas have been segmented, the next step is detecting text in the ROI images. Most of the available methods for text detecting use gray-scale or binary images, because methods for color analysis are time-consuming or involve elaborate processing. After image preprocessing, we using edge detection and projection to detect the text area (an edge in gray-level images is changed suddenly in value from white to black). Projection is a favored algorithm for text detection; it is based on alterations of edges in limited areas. Our method uses horizontal projection to find the candidate regions (see Fig.6 (a)). Then candidates screening is implemented by threshold processing by the following equation:

$$p(y) = \begin{cases} 0 & \text{if } p(y) < T \\ p(y) & \text{if } P(y) \geq T \end{cases} \quad (7)$$

Where  $p(y)$  is the horizontal projection image, and  $T$  is threshold value which is obtained by empirical knowledge.

The upper and lower boundaries of the text region can be determined after the threshold processing (see Fig.6 (b)). Similarly, the left and right boundaries of the text region can be confirmed (see Fig.6(c)).

## IV. EXPERIMENTS AND RESULTS

An automatic system for vehicle type identification based on car tail text information was implemented in this paper to evaluate the performance of our system. The process of license plate recognition has been completed by previous work, no need to display the experimental results here.

Twenty-five vehicle types are collected in the paper for performance evaluation, and they are the most common models of all, such as Toyota Camry, Toyota Vios, Honda CRV, Nissan Tenna, Ford Mondeo, Benz C200, etc. To train the SIFT classifier, a database containing 200 text information classes of the car tail were collected, which contains color

images with image sizes varying from  $215 \times 85$  to  $1005 \times 650$ . Fig. 7 shows some of the samples. The text information of the car tail usually contains three parts: manufactory, model and displacement. In mostly literatures for vehicle recognition or model verification, they use vehicle logo to recognize the vehicle type, it maybe causes information loss.

For the testing set, we test the method with sets of images (size:  $1200 \times 900$  pixels) taken by PENTAX K-50 camera, while 1200 sub-images from 684 rear-view vehicle images are captured after text area localization processing. The testing images are all captured without making any controlled condition; the system is close to the real environment. It encompasses 25 categories of vehicles under sunny, cloudy, backlighting or shadow conditions.

In our system, we have considered as much information as possible; consequently, the vehicle type matching is designed to guarantee uniqueness. Firstly, the query image is divided into two or three parts: including information about the vehicle manufacturer, model and displacement; after that, the three sorts of information matching with samples which stored in the database respectively; ultimately the result could be obtained. In the existing methods, the usual plans are matching vehicle logo or vehicle face; while various vehicle types may have similar face or the same logo, it is hard to differentiate between the query one from others. Our method has improved this current situation.

In Section II, we have demonstrated the vehicle text area detection and segmentation, and the result has shown in Fig.6. Firstly, we capture the region of interest by a prior knowledge as is shown in Fig.5, after segmenting the ROI, accurate localization processing operates for obtaining the sub-image that used up to matching scheme. We use a serious of preprocessing procedures (image smoothing, edge extraction and binaryzation) to enhance image information. Next, the statistical law of the characters of car tail text area horizontal projection is used to select in the ROI of the text area and determine the top and bottom borders. As Fig.6 shows, the text area presents a higher peak value. What's more, the vertical

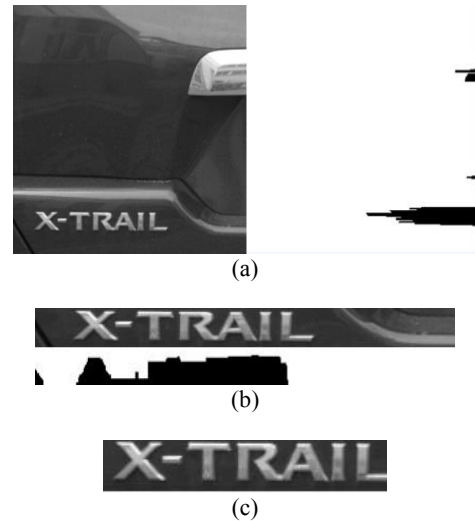


Figure 6. (a)ROI sub-image after robust localization and its horizontal projection. (b)Vertical projection of the text area sub-image. (c).Accurate localization result.

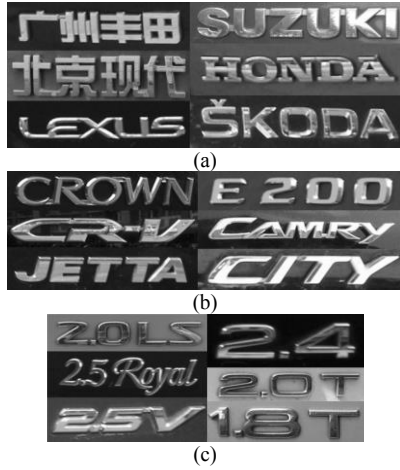


Figure 7. (a) Samples of manufacturer information. (b) Samples of model information. (c) Samples of displacement information.

projection shows continuous projection imaging. So we can determine the right and left borders. Computation, nature of performance of text-segmentation process is 88% on average. The detection process requires about 420 ms.

Fig. 8 shows some successful results. The results suggest that our system is robust against large variations in color, size, geometric distortion and even some local defect. Lines connect the matched features between the query image and a sample image. A successful match occurs when the two keypoints correspond to the same location and a failed match when two keypoints come from different locations. These parallel lines ensure the good matching effect, while several amorphous ones indicate they are wrong matching pairs. For different experimental conditions, we vary the threshold value ( $\gamma$ ).

We also give an indicative example of the NN matches to show the distinction between matched sample and the others. There are 158 keypoints ( $KP_s$ ) have found in the query image, and matched feature number (NNs) has also been added up (see Fig. 9).

In order to ensure the accuracy of the matching results, it

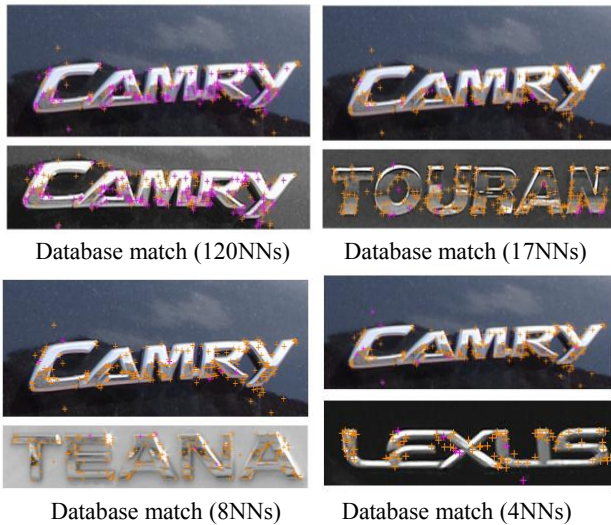


Figure 9. NN match for a sample query image.

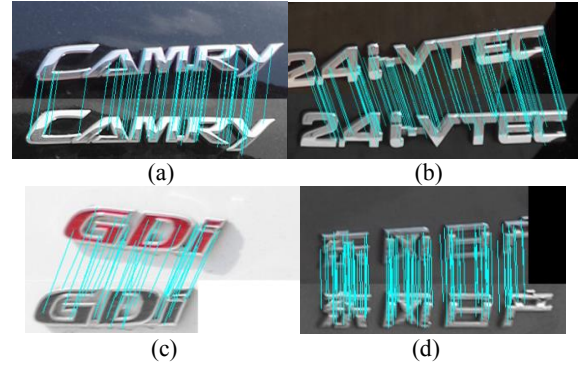


Fig.8 Results of the text recognition. The upper images are text sub-images which are segmented from the query images (from training set) after text localization processing, and lower ones are samples from the database. Query images are in a RGB color format, while sample images are all in grayscale.

needs to not only minimum within-clusters, but also maximum between-clusters. As is shown in Fig.9, the correct match has the maximum amount of keypoints, while the false match has a small number. It could be also regarded as the similarity between the query image and sample images. The features and the matched points have been marked in Fig.9, though there are several matches in the unmatched samples, most of them are false ones.

Because of various numbers of keypoints can be captured from different query images, the accuracy of text recognition may be different either. In our experimentation, we test images with different sizes under various resolutions. As is shown in Fig. 10, one image of little features has a low recognition rate. So, SIFT-based algorithms require the original image to have a certain level of resolution; otherwise, the properties of the feature would not be stable when the image is scaled down. In our testing set and training set, the average number of images' keypoints is close to 150. The accuracy analysis of our text information recognition method is shown in Table I.

The recognition rate of the information is determined by its contents. Compared with manufacturer and model information, displacement information is simpler and mostly composed of

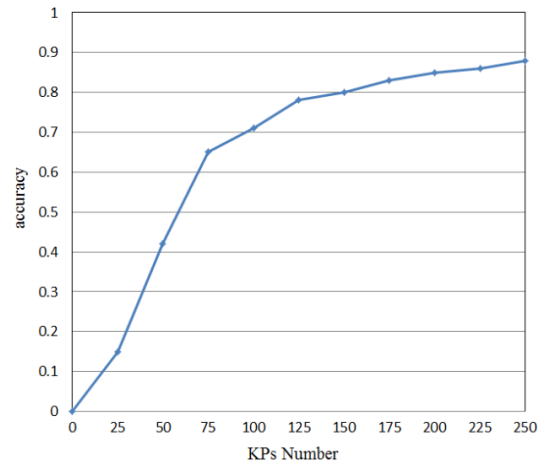


Figure 10. The polygonal line shows the accuracy of query images with various numbers of keypoints matched to a database.

TABLE I. EXPERIMENTAL RESULT FOR THE THREE SORTS OF INFORMATION

Query information	True	False	Total	Precision
Manufacturer	293	107	400	73.25%
Model	507	81	588	86.22%
Displacement	196	16	212	92.45%

figures and English letters, so it becomes the most easily part to recognize. For manufacturer information, the precision has the worst result by reason of many of the query sub-images including the content of Chinese characters. On the basis of the existing technology, it is hard to recognize for Chinese characters. In addition to weather, backlighting and shadow can also affect the performance of text area detection and recognition. And the worst false alarm is due to the influence of the shadow in the sunny days. Computation, the overall recognition success rate is 83%.

The VTI system has a high running speed, with a combined detection and recognition time of about 4.8 s (420 ms + 4.38 s) on average, which is suitable for applications. The recognition speed of the three sorts of information is shown for comparison in Table II. Once we get the matched vehicle information, a search on the vehicle information file will process, which costs approximately 180 ms.

## V. CONCLUSION

In this paper, we have proposed a vehicle type identification system based on car tail text information, which is a new application for SIFT algorithm. By testing hundreds of the rear-view vehicle images, the results show that the proposed method has good efficiency. And the accuracy is still within "acceptable" limit.

Experimental results have proved the viability of our proposed system; however, there are still defects, such as lack of training samples, low precision rates in the Chinese characters recognition. Future work will focus on the improvement of SIFT to adapt to low-resolution images and to accurately recognize the Chinese characters.

## REFERENCES

- [1] A. P. Psyllos, C. N. E. Anagnostopoulos, and E. Kayafas, "Vehicle Logo Recognition Using a SIFT-Based Enhanced Matching Scheme," *IEEE Transactions on Intelligent Transportation System*, Vol.11, No.2, pp. 322-328, 2010.
- [2] A. P. Psyllos, C. N. E. Anagnostopoulos, and E. Kayafas, "M-SIFT: A New Method for Vehicle Logo Recognition," *IEEE International Conference on Vehicle Electronics and Safety*, 2012, pp. 261-266.
- [3] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IEEE Transactions on Information Theory*, Vol. 8, No. 2, pp.179-187, 1962.
- [4] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *The 4th Alvey Vision Conference*, Vol. 53, No. 3, 1988, pp. 147-151.
- [5] H. P. Moravec, "Rover Visual Obstacle Avoidance", *International Joint Conference on Artificial Intelligence*, Vol. 3, No. 4, Vancouver, Canada, 1981, pp. 785-790.
- [6] C. Harris, "Geometry from Visual Motion," *Active Vision*, MIT Press, Cambridge, pp. 263-284, 1992.
- [7] D. G. Lowe, "Object Recognition from Local Scale-invariant Features," *International Conference on Computer Vision*, Corfu, Greece, 1999, pp. 1150-1157.

TABLE II. VEHICLE TEXT INFORMATION RECOGNITION RATE

Query information	Time(second)
Manufacturer	5.89
Model	3.64
Displacement	3.62

- [8] D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.
- [9] M. Brown and D. G. Lowe, "Invariant Features from Interest Point Groups," *British Machine Vision Conference*, 2002, pp. 253-262.
- [10] J. S. Beis and D. G. Lowe, "Shape Indexing Using Approximate Nearest-neighbour Search in High-dimensional Spaces," *Conference on Computer Vision and Pattern Recognition*, Puerto Rico, 1997, pp. 1000-1006.
- [11] D. H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition*, Vol. 13, No. 81, pp. 111-122, 1981.
- [12] L. Dlagnekov, *Video-based Car Surveillance: License Plate Make and Model Recognition*, Master Thesis, University of California at San Diego, 2005.
- [13] I. Zafar, E. A. Edirisinghe, and B. S. Acar, "Localized Contourlet Features in Vehicle Make and Model Recognition," *Proc. SPIE 7251, Image Processing: Machine Vision Applications II*, Vol. 7251, 2009.
- [14] J. W. Hsieh, L. C. Chen, and D. Y. Chen "Symmetrical SURF and Its Applications to Vehicle Detection and Vehicle Make and Model Recognition," *IEEE Transactions on Intelligent Transportation System*, Vol. 15, No. 1, pp. 6-20, 2014.
- [15] U. Iqbal, S. W. Zamir, and M. H. Shahid, "Image Based Vehicle Type Identification," *International Conference on Information and Emerging Technologies*, 2010, pp. 1-5.
- [16] D. Trier, A. Jain, and T. Taxt, "Feature Extraction Methods for Character Recognition-A Survey," *Pattern Recognition*, Vol. 29, No. 95, pp. 641-662, 1996.
- [17] M. Bosker, "Omnidocument Technologies," *Proceedings of the IEEE*, Vol. 80, No. 7, pp. 1066-1078, 1992.
- [18] M. H. Glaubergerman, "Character Recognition for Business Machines," *Electronics*, Vol.29, No.2, pp. 132-136, 1956.
- [19] D. Shen and H. H. S. Ip, "Discriminative Wavelet Shape Descriptors for Recognition of 2D Pattern," *Pattern Recognition*, Vol. 32, No. 98, pp. 151-165, 1999.
- [20] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An Algorithm for Finding Bestmatches in Logarithmic Expected Time," *ACM Transactions on Mathematical Software*, Vol.3, No.3, pp. 209-226, 1977.
- [21] P. V. C. Hough, *Method and Means for Recognizing Complex Patterns*, U.S. Patent 3069654, 1962.
- [22] E. Grimson, "Object Recognition by Computer: The Role of Geometric Constraints," *The MIT Press*, Cambridge, MA, 1990.
- [23] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, Vol. 24, No. 6, pp. 381-395, 1981.
- [24] M. Diligenti, M. Gori, M. Maggini, and E. Martineli, "Adaptive Graphical Pattern Recognition for the Classification of Company Logos," *Pattern Recognition*, Vol. 34, No. 10, pp. 2049-2061, 2001.
- [25] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic License Plate Recognition (ALPR) A State-of-the-Art Review," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 23, No. 2, pp. 311-325, 2013.
- [26] H. Sheng, C. Li, Q. Wen, and Z. Xiong, "Real-Time Anti-Interference Location of Vehicle License Plates Using High-Definition Video," *IEEE Intelligent Transportation Systems Magazine*, Vol. 1, No. 4, pp. 17-23, 2009.
- [27] G. S. Hsu, J. C. Chen, and Y. Z. Chung, "Application-Oriented License Plate Recognition," *IEEE Transactions on Vehicular Technology*, Vol. 62, No. 2, pp. 552-561, 2013.

# Design of an Educational Adventure Game to teach computer security in the working environment

Ciro D'Apice, Claudia Grieco, Rossella Piscopo  
DIEM, Università degli Studi di Salerno  
Fisciano (SA), Italy  
cdapice, cgrieco, rpiscopo@unisa.it

Luca Liscio  
CRMPA  
Fisciano (SA), Italy  
liscioluca@gmail.com

**Abstract**— One of the emerging requirements for learning in the enterprise is finding new ways to keep the learner engaged. The use of advanced learning technologies that exploit a gameful design should increase engagement and encourage students to make connections between the simulated environment and the real world. This article presents game design guidelines for the development of an Educational Adventure Game and how they have been applied during the development of SIRET Security Game, a game that teaches workers the importance of following computer security policies.

**Keywords:** *educational adventure game; game design; assessment.*

## I. INTRODUCTION

The SIRET project<sup>1</sup> is currently investigating the potential of advanced learning technologies in the working environment.

The most important challenge in designing learning content for the enterprise is finding ways to facilitate *deep learning* instead of *shallow learning*. Shallow learning is mnemonic and repetitive acquisition of knowledges such as word definitions, object properties, sequences of events. Serious Games (i.e. videogame designed for a primary purpose other than pure entertainment) can instead be designed to promote deep learning, that consists in understanding cause/effect mechanisms, being able to explain what learned, developing critical reasoning, managing limited resources, applying old solutions to new problems. These skills are all highly desired in the modern era.

Serious Games that have strong elements of simulation and storytelling provide numerous benefits in an enterprise context: the worker is brought to make connections between the simulated environment and the real world, and analysis and reflection are crucial to achieve the learning goals [2]; simulations provide authentic experiences that involve the student both spatially, providing the illusion of being in the place of the events, and emotionally, through its participation in the narrative.

People often hear concepts like Online Identity Theft, web site sabotages and illegal computer operations (like money transfer from unaware people's bank accounts) and think that the greatest dangers to company security are hackers or organized crime.

A recent research from Cisco Italy reveals that, frequently, higher risks come from inside [1]. We are not talking about boycotting or bad intentions of the employees, but about lack of proper training, unapplied or unverified control processes that cause careless behavior to put company computer security at risk, especially in Italy.

In particular, according to Cisco: 55% of Italian respondents admits to use a unique password for sites and applications and does not change it periodically; company networks are often used for personal, potentially risky activities such as personal banking and online shopping (58%); 70% thinks that their behavior cannot put the safety of the company at risk. In companies that provide for computer security policies, these are often thought as bothersome by employees who, at times, even try to get around them. For these reasons, training workers about possible damages from loss or theft of information and about ways of avoiding them is becoming crucial, but it is also crucial to keep learners engaged and to get the message really across.

This paper will describe the design and development of the Educational Adventure Game **SIRET Security Game**, which tries to assess these concerns. Section II will introduce the concepts of Educational Adventure Game and Storytelling. Section II A will provide related work on the subject, while Section III will describe the existing design guidelines for Educational Adventure Games. In Section IV, the SIRET Security Game will be described. Finally, in Section V we draw the conclusions and describe future works.

## II. EDUCATIONAL ADVENTURE GAMES AND STORYTELLING

An Adventure Game is a video game in which the player assumes the role of protagonist in a story driven by exploration and puzzle-solving [3]. The setting is composed of 2D sceneries linked to each other, that the protagonist has to explore to progress in the plot. In order to advance the story, players have to solve a series of puzzles, that typically require to interact with scene objects and with the people they meet.

---

<sup>1</sup> SIRET – Integrated System of Recruiting and Training - PON01\_03024 – The **National Operational Programme** for "Research and Competitiveness" 2007-2013 (NOP for R&C)  
DOI reference number: 10.18293/DMS2015-002

This genre is also called “Point and Click” since mouse clicks are the principal means of interaction within the game. These games are characterized by a strong focus on storytelling, comedy and puzzles to solve and by the lack of action sequences or time limit to solve quests. This combination is good for eLearning [4]: educational content can be conveyed in the interweaving narrative and included in puzzles and in the game scenarios, while the player can assimilate the concepts slowly thanks to the absence of time limits. It is also worth noticing that the development cost of these games is limited [11] and this suits well the academic environment, where budgets are not as big as in videogame companies. Adventure Games created for educational purposes are called **Educational Adventure Games**.

The strong focus on plot makes Educational Adventure Games one of the Serious Game genres that exploit the most the Storytelling paradigm, more than roleplaying games (RPGs) and action games [16]. Storytelling means the narration of a story and the process by which it is conveyed to the public[5]. Bruner pointed out the elements characterizing the narrative, defining it as a “*unique sequence of events, mental states, happenings involving human beings as characters or actors: these are its constituents. But these constituents do not, as it were, have a life or meaning of their own. Their meaning is given by their place in the overall configuration of the sequence as a whole - its plot or fabula*” [6]. Narrative is a favorite instrument for the development of cognitive skills and the dissemination of knowledge in the field of business education [7].

A recent classification of Serious Game Genres [8] noticed that games with strong focus on role playing, narrative and simulation are the most useful for the development of Soft Skills like critical thinking and verbal communication. That is the reason why we chose Educational Adventure Games as the genre for developing a Serious Game for eLearning in the enterprise.

#### A. Related Work

Many researchers approach the design of Serious Games using methodologies taken directly from software project management. For example, in the TIE project [29] the Serious Game *Pappi World* was developed following the SCRUM agile methodology [30] in which a software is developed using small incremental steps (called *Sprints*). An incremental approach was also followed during the development of SIRET Security Game, but because of the peculiar characteristics of Educational Adventure Games, it was also needed to identify and follow design guidelines proper of the genre, which will be described in Section III.

Educational Adventure Games have been used in education since the 90s [9] and are still very much in use [4]. Some examples in the science fields [10] are *Electro Adventure* and *Twisted Physics*, which teach respectively Classical mechanics and Electromagnetism. In these games the player explores a building and finds a problem-solving exercise in each room. In *Twisted Physics* the character gets the key to the next room only after solving the problem: this motivates players and

ensures that learning contents are addressed with a precise order.

The most famous tools for the creation of Adventure Games are Adventure Game Studio (AGS)<sup>2</sup> and Visionaire Studio<sup>3</sup>. They both offer a Drag&Drop interface to create games visually, limiting the code that needs to be inserted. However, they are not products designed to produce eLearning content; therefore, their games cannot easily be integrated with Learning Management Systems (LMS) or inserted in a SCORM package. There are three research tools which focused specifically on Educational Adventure Games: StoryTec [4], RealChamber [21] and eAdventure [11]. StoryTec allows creating Narrative Game-based Learning Objects (NGLOBs), in which each scene of the game can be annotated with the skills it provides and with the prerequisites to view it, while players' actions may cause an update of their cognitive profile. StoryTec games, however, can be executed only in the StoryTec player, so there is a problem of interoperability. RealChamber is an example of the potential of Educational Adventure Games for learning in the enterprise. It is a commercial adventure game building tool in which photographs of real locations can be used as the 2D sceneries of the game. In this way it is possible to obtain a certain level of realism while keeping the game computationally efficient. The game engine chosen for our project is **eAdventure**. eAdventure is an open source library developed by the Universidad Complutense de Madrid. It offers most of the main functions of an Adventure Game library but also includes useful functions for learning content creation. eAdventure games, in fact, can be exported using Learning Object standards (SCORM 1.2, LAMS, IMS-CP and so on) and can dialogue with LMS to obtain the learning state of the player and to transmit the learner score.

### III. GAME DESIGN GUIDELINES

Adventure Games are a genre with strong conventions, whose user interaction mechanisms have been standardized over time [4]. These games also need a limited set of graphical resources. Designing an Adventure Game therefore essentially involves creating:

- a) the story to narrate;
- b) the world the protagonist explores;
- c) the puzzles the player has to solve.

For Educational Adventure Games we believe there are other two necessary elements:

- d) which learning contents to convey and the most effective way to convey them
- e) strategies to assess the player/student performance.

Jane McGonigal in [13] introduced the concept of “Gameful design” for eLearning content: the goal is to create a didactical material (in our case, a Serious Game) that has the spirit, and not just the mechanics, of a good game. Gameful

<sup>2</sup> <http://www.adventuregamestudio.co.uk>

<sup>3</sup> <http://www.visionaire-studio.net>



Design is based on the “PERMA” approach (**P**ositive emotions; **E**ngagement (or flow); **R**elationships; **M**eaning; **A**ccomplishments). Instead of giving extrinsic motivations for playing (points, levels, achievement badges), Gameful Design provides intrinsic motivations: the very act of playing is a reward for the students. It is a vision much more focused on people and on their positive experience rather than on videogame techniques and mechanics.

While designing story, world and puzzles, we chose to follow guidelines that increase characteristics described in the PERMA approach, because keeping learner motivation high is a key factor in education in the working environment.

**Puzzles** need to have clear rules and objectives [15]. Also, there should be a good balance of difficulty levels in order to make the game neither too simple nor too hard, to keep the player Engagement’s levels high. It is better to start with simple problems of immediate solution that serve to explain the game mechanics and give the player immediate positive feedback (**A**ccomplishment). Next, the game should gradually raise the difficulty level. Some possible kinds of puzzles are: set of actions to complete in a specific sequence; combining different elements (people and objects); using common object in an unconventional way (e.g. a coconut as a bucket). To increase **P**ositive **E**motions and sense of **A**ccomplishment, the player can be rewarded for completing a task with animated sequences, new areas or new powers to solve more complex puzzles [15].

The **World** should be “interactive” [24]: the player must be able to interact with characters and objects as much as possible, even if it is not needed to continue the story. First, the main setting should be chosen (Realistic or Fantasy? Modern day or Historical period?) For educational purposes, it is important to help the player in relating the game experience to problems of the real world, so the setting should simulate places where the competences taught in the game could be realistically applied. Designing the world means also creating a map of all the locations the main character has to visit, so it is strongly linked to story and puzzle design.

As for the **Story**, it is important to involve the player by creating drama and empathy. This is achieved by providing engaging narrative contents (**M**eaning). As explained by McGonigal [13] if players recognize narrative patterns familiar to them they will be more inclined to pay attention and will assimilate more concepts. One of the most used narrative patterns is The Hero’s journey, common in ancient myths as well as modern day adventures. The concept of the Hero’s Journey was described by mythologist Joseph Campbell in his book *The Hero with a Thousand Faces* [20]: the protagonist leaves the familiar world behind, overcomes a series of trials, receives a reward for his accomplishments and returns home to live a peaceful life. Story design includes characterizing the protagonist and the people he or she will interact with (Non-Playable Characters or NPCs). The fields of cryptography and computer security have long benefitted from the use of *character archetypes*, i.e. characters playing a specific role in the narrative (hero, villain, etc.) and identified using conventional names (Alice, Bob, Oscar and so on) [22].

An important factor in Story Design for Adventure Games is deciding whether allowing the player to **lose**. Losing means:

- a) Receiving unequivocal communication of mission failure (death of the main character, "Game Over" message, etc.)
- b) Allowing the game to enter an *Unwinnable State*, a state in which it is no longer possible to continue (e.g. the player must open a door, but she forgot to take the key from her car and her car has been stolen).

The last kind of player defeat is currently considered a design mistake [28] and should be avoided at all costs. For Educational Adventure Games, it is important to encourage exploration of all the game features, allowing the player to also make mistakes. In order to avoid frustrations, we believe mistakes should not lead to the interruption of the game and the player must be offered opportunities to try again.

As for **Learning contents**, video games are better suited to teach processes, methods of participation and dynamic actions, while simple lists of notions are not appropriated [14]. The learning content should be an integral part of the game, which must simulate real-world applications of the competences we want to teach. Detailed theoretical explanations, however, can be included within the game experience as sequences of texts and images that serve to deepen the concept learned through play. The structure of Adventure Games helps the game designer in placing educational contents[10]. In Adventure Games the player moves within a predetermined set of physical contexts (well defined places like rooms, corridors, etc.), so it is easy to associate learning objectives to each of these contexts, while the road to travel becomes a literal representation of the *learning path*. Another technique described by Amory [17] is dividing the game in different *Acts*, in order to subdivide the learning contents into groups of increasing difficulty.

For **Assessment**, Serious Games have a significant advantage over simple tests: extensive opportunities for tracking user actions, which lead to a more complete student judgment. One of the most used assessment strategies is *In Process Assessment* [18]: player performance is measured during the game itself according to the interactions with game elements. Player’s assessment can be used to adapt in real time the complexity of the game to the player’s skills [19]. This is useful to present a level of difficulty neither too easy nor too hard, so as to keep the player in a state of Flow, the feeling of getting lost in the media element which makes the game intrinsically motivating [13].

Oostendorp et al [19] distinguish five different components of Serious Games that can be affected by adaptivity: the game environment, whose layout can be made easier for players in trouble; game mechanics (e.g. shooting can be made more or less precise according to difficulty levels); NPC attributes, that can change according to the player’s competences; game narrative (order of events); game scenes (whole sections of the game can be included or erased according to the learning objectives). In Educational Adventure Games, that have



standardized game mechanics and fixed layouts, adaptation can be focused on changing narrative, scenes and NPC attributes.

Adaptivity is also useful from an educational point of view, since it allows designing recovery strategies if the system notices that a student fails in learning a concept. As an example, in the Storytelling Complex Learning Objects created in the ALICE project [12] if the player fails assessment tests during the game, the story adapts itself by changing the point of view (Role), the scenario or the formative content, allowing them to see things from another perspective. In the medical Serious Game JDoc, instead, if players fail a mission the game places a new character (a nurse) in the scene, which provides suggestions on how to proceed correctly.

#### A. Evaluation of Game Effectiveness

In order to evaluate the game effectiveness, we referred to the work by Yusoff et al. [25] which identified the desired characteristics of Serious Games. These features are summarized in Table 1 and constitute a “checklist” that the designer has to follow in order to be sure of covering all the necessary requirements. An important factor in measuring game effectiveness is *transfer of learnt skills*, i.e. measuring effective transfer of knowledge and skills through play. This is usually done [26] by making a group of students experience the game and measuring their cognitive state (using a test focused on the subject taught by the game) before and after the playing activity. Eventual variations in the cognitive state are attributed to the game’s didactic abilities.

TABLE I. SERIOUS GAMES FRAMEWORK ELEMENTS

Serious Games Framework elements	Definition
Clear Instruction	In-game Instructions explain how to use the game
Intended learning outcomes	Learning outcomes are clearly exposed
Instructional content	The instructional content is well structured and subdivided.
Game mechanics	It is easy to interact with the game and system answers are consistent and easy to understand.
Linearity	Students are not confused and are able to follow the plot.
Attention Span	An adequate attention span is required
Interaction	The challenge fits with the target audiences giving the learners the feeling of comprehension and satisfaction.
Learner control	The difficulty level is adequate to the game target, avoiding frustration while being a satisfactory challenge.
Game achievement	The in-game assessment influence the game flow, adjusting it to match the learner’s abilities
Reward	The player feels motivated because of in-game rewards whenever they finish an important task
Intermittent feedback	There is a system of feedbacks and helps that supports the user.
Situated and authentic learning	Students can make correlations with real life applications of the competences they are taught.
Transfer of Learnt Skills	Effective transfer of knowledge and skills through play

## IV. SIRET SECURITY GAME

In the Serious Game SIRET Security Game the player takes the role of an employee of a famous Corporation and has to defend corporate data from spies and saboteurs.

The main character has to complete a series of missions in order to be considered worthy of the position of Computer Security Officer. This game is meant to convey the principles of information security: defense against viruses (Malware, Spyware etc.), fraud protection, Cryptography principles. The target audiences are public administration and company workers that want to understand how to implement security policies.

The game is not aimed at Information Technology experts, but to employees of other areas, in order to help them take responsibility and understand that the safety of the company depends on them.

Game mechanics follow the elements of classical Adventure Games: scene elements with which to interact are called *hotspots*; hotspots can be clicked to choose the right action to apply on them (examine, use, talk to, etc.); the main character has an inventory containing objects that can be combined with each other; interactions with NPCs are done following a “Dialog Tree”, a graph of all the possible dialogue options.

The following paragraphs will describe how Game Design Guidelines have been applied to the development of SIRET Security Game.

#### A. Story, World and Puzzles

The game plot contains many elements of classic narrative and of the Call to Adventure pattern [20]. Harry, the main character, is the **Hero** archetype, who receives a mission from his boss, meets the **mentor** (a security officer offering help and advices), solves a series of **quests** and defeats the **villains** (in this case the saboteur Oscar and the spy Eve). Harry’s boss acts as the **Threshold Guardian** [16], a character that tests the learner’s knowledge, by checking which of the quests have been completed and rewarding the player when his work is completed. Some characters in the game have been taken from the archetypes of *Applied Cryptography* [22]: Harry’s friends Alice and Bob, the enemies Eve and Oscar.

The world in which Harry moves is a small office (Figure 1. ) composed of 2D pictures created from 3D objects, in order to maintain a certain level of realism while keeping the game computationally efficient. Also for computational efficiency, characters are animated using *sprite sheets*, sets of single graphic images (called *frames*) that are rendered in quick succession to give the illusion of movement.

The player can interact with most of the people and objects in the scene, sometimes also for amusing purposes. In-game puzzles require the player to learn computer security principles in order to offer advice to Alice and Bob, and collect and combine scene objects to act on the information learned.

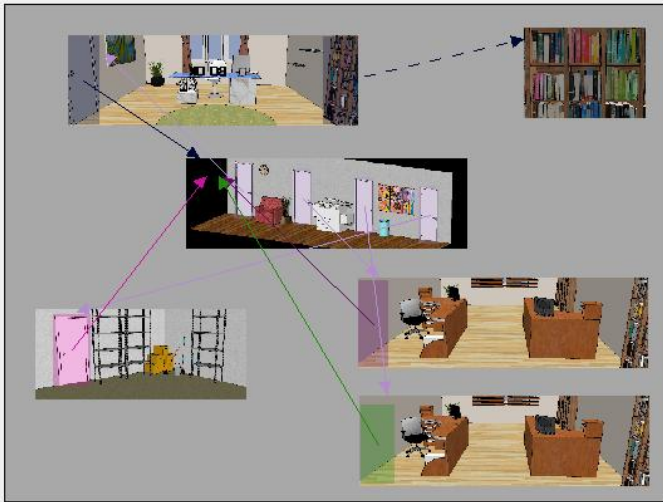


Figure 1. Scene Map of SIRET Security Game.

### B. Learning Content Presentation Strategies

In SIRET Security Game learning content can be conveyed using:

- “Book” objects in-game that, if read by the protagonist, convey detailed information.
- Pictures and texts in *Cutscenes*, non-interactive sequences of the game.
- Interactions with NPCs.
- Sets of procedures that the player is asked to follow in order to achieve a specific goal.

We tried to deliver as many concepts as possible through simulations of real situations and interactions with NPCs (Figure 2), in order to maintain a good level of immersiveness.



Figure 2. The protagonist interacts with NPCs.

To allow players to learn the cryptography principles it is better to present the problem as a series of actions to be taken to prevent a spy from intercepting an important message, instead of immediately providing the theoretical notions.

As summarized in Table 1, it is essential that the intended Learning outcomes are clearly exposed in-game: students at any part of the learning experience should keep track of what the current learning goals are and what goals they have obtained so far. For this reason the player at the start of the game receives a *journal*, Harry's personal notebook divided into two sections:

- Quests:** it contains all the security problems the protagonist has to solve. This section keeps track of completed and uncompleted missions.
- Knowledge:** it summarizes all the concepts the player has acquired. It is a Reference Library that automatically updates each time a learning goal is completed.

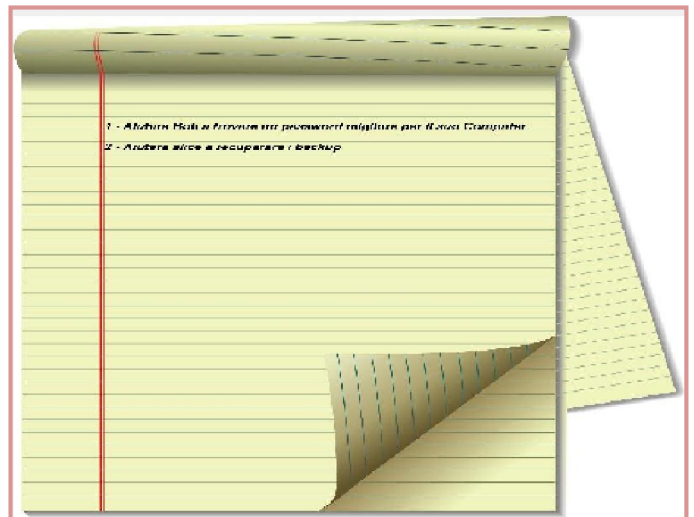


Figure 3. Harry's Journal – Quests Section.

We decided to divide contents following Amory's approach [17]: the game is divided into two Acts of increasing difficulty. Inside each act, different challenges are posed to the player in each room of the office building, but he or she is free to roam across all the rooms and solve the missions in the preferred order. That is to help maintain a certain level of engagement: if players are stuck in resolving a specific problem, they can resolve other quests first and avoid being bored or frustrated.

### C. Assessment and adaptation Strategies

The assessment strategy we chose to follow is In Process Assessment [18]. From a programmatic point of view, the Adventure Game genre makes designing this strategy easier. Performance tracking is possible by inserting *triggers* (event generators) in the game [23]. The event can be triggered when a certain system variable reaches a specific value (data trigger) or when the player interacts with a scene object (trigger box).

In SIRET Security Game:

- The system exploits user variables to keep track of scores for each of the learning concepts.
- Interactions with people or scene objects trigger events that update user variables or keep track of user actions.

This kind of assessment allows the teacher to understand the reasons behind the player's actions [18]. Assessment also takes advantage of the Assessment Profile feature of eAdventure [11].

Figure 4. Assessment Profile.

Each profile (Figure 3) is made up of a set of rules (conditions on user variables). When the conditions are met, rules are applied to assess student behavior.

In SIRET Security Game if the system detects that the player is in obvious difficulty (they take too long to solve the game, they perform too many incorrect attempts) the player receives a call from the **mentor**. The mentor is Dave, a senior security officer, which offers help and direct advices to solve the game puzzles. The game adapts, therefore, to the player's actions.

Asking the mentor for tips greatly simplifies the game. For this reason a student who uses the tips should be penalized in the final vote. On the other hand, the purpose of the game is teaching new information, so asking for help cannot always be considered a negative action. For this reason the mentor will give three different kinds of help:

- Veiled: clearer explanation of the problem that the protagonist must solve, to help the player to focus on the goal. No suggestions are given on how to fix the problem.
- Concrete: the player is given a first hint on what to do to solve the problem.

- Obvious: the mentor explains step by step the moves that the player needs to do to solve the problem.

Players will not be penalized for a veiled help. A concrete help will cause a small penalization, while an obvious help will cause a strong penalization.

#### D. Evaluation of Game Effectiveness

The Serious Game Framework elements of Table 1 have been used as a basis for developing a questionnaire for testing game effectiveness. The questionnaire is composed of a series of statement and the reviewer has to answer how much it agrees (from a scale of 1 to 5) with each statement. A first, internal review of the game has been done using the questionnaire: the puzzles offer an adequate challenge level, the player is rewarded for completing tasks and it is easy to imagine a real life application of the information the player acquires. The review feedback prompted the addition of more in-game instructions to explain game mechanics.

## V. CONCLUSIONS AND FUTURE WORKS

**SIRET Security Game** has been implemented using the eAdventure platform. The game is divided into two acts, for a total playing time of nearly half an hour. It can be exported as a SCORM package, allowing it to be integrated into any LMS and to update the student's learning state according to the game results. After the first internal review, a more detailed evaluation of the game will be done in the future by subject experts: computer security officers of the public administration will experience the game and use the game effectiveness questionnaire to check whether all the Serious Game Framework requirements have been met. Future works also include measuring the effective transfer of knowledge and skills through play. In order to measure the transfer, students of Computer Science could be involved: they will experience the game and a pre and post-game questionnaire will measure any variation in their cognitive state.

## ACKNOWLEDGMENT

The research reported in this paper has been performed during the project SIRET – Integrated System of Recruiting and Training PON01\_03024 – The **National Operational Programme** for "**Research and Competitiveness**" 2007-2013 (NOP for R&C). Special thanks go to all the partners involved in the SIRET project.

## REFERENCES

- [1] Cisco Italia, "Una ricerca Cisco svela quali sono i comportamenti dei dipendenti che minano la sicurezza IT delle aziende Italiane", <http://www.cisco.com/web/IT/press/cs14/20141027.html>, Accessed: 2015-02-3
- [2] P. Jacobs and I. Bone. "The need for interactive narrative in educational management simulations." Interact, integrate, impact: proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE) edited by G Crisp, D Thiele, I Scholten, S Barker and J Baron. Vol. 2. pp. 618-623, 2003.

- [3] E. Adams, *Fundamentals of game design*. Pearson Education, 2013.
- [4] F. Mehm, S. Göbel, and R. Steinmetz. "Authoring and re-authoring processes for educational adventure games." 6th European conference on games based learning. Vol. 323. 2012.
- [5] G. Genette. *Narrative discourse revisited*. Cornell University Press, 1988.
- [6] J. Brunner. "Acts of Meaning: Four Lectures on Mind and Culture." 1990.
- [7] R. C. Schank. *Tell me a story: A new look at real and artificial memory*. Charles Scribner's Sons, 1990.
- [8] B. Botte, C. Botte, and M. Sponsiello. "Serious Games between simulation and game: A proposal of taxonomy." *Journal of e-Learning and Knowledge Society* 5.2, 2009.
- [9] B. Cavallari, J. Heldberg, and B. Harner. "Adventure games in education: A review." *Australasian Journal of Educational Technology* 8.2, 1992.
- [10] H. M. Halff. "Adventure games for science education: Generative methods in exploratory environments." *Proc. AIED05 WORKSHOP5: Educational Games as Intelligent Learning Environments*, pp. 12-20, 2005.
- [11] J. Torrente, A. Del Blanco, E. J. Marchiori, P. Moreno-Ger and B. Fernández-Maniñón. "<e-Adventure>: Introducing educational games in the learning process." In *Education Engineering (EDUCON)*, 2010 IEEE, pp. 1121-1126.
- [12] M. Gaeta, V. Loia, G. R. Mangione, F. Orciuoli, P. Ritrovato, S. Salerno. "A methodology and an authoring tool for creating Complex Learning Objects to support interactive storytelling", *Computers in Human Behavior*, 31, 620-637, 2014.
- [13] I. McGonigal. "We Don't Need No Stinkin'Badges: How To Reinvent Reality Without Gamification." *GDC Vault*. [www.gdcvault.com](http://www.gdcvault.com), 2011.
- [14] N. Fortugno, E. Zimmerman, "LEARNING TO PLAY TO LEARN - Lessons in Educational Game Design", [Online]. Available: <http://www.ericzimmerman.com/texts/learningtoplay.html> Accessed: 2015-02-30
- [15] B. Tiller, A. Larry, "21 Adventure Game Design Tips"
- [16] M. D. Dickey, "Game design narrative for learning: Appropriating adventure game design narrative devices and techniques for the design of interactive learning environments." *Educational Technology Research and Development* 54.3, pp. 245-263, 2006.
- [17] A. Amory, "Another Country: Virtual Learning Spaces," in *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Honolulu, Hawaii, USA, 2003.
- [18] S. Chen e D. Michael., *Proof of learning: Assessment in serious games.*, 2005.
- [19] H. van Oostendorp, E.D. van der Spek, and J. M. Linssen. "Adapting the complexity level of a serious game to the proficiency of players.", 2014.
- [20] J. Campbell, *The Hero with a Thousand Faces* Novato, California: New World Library, 2008 ISBN 978-1-57731-593-3.
- [21] Gamification.it. *Real Chamber al Salone del Libro: la piattaforma gamificata per le tue avventure grafiche* [Online]. Available: <http://www.gamification.it/gamification/real-chamber-al-salone-del-libro-la-piattaforma-per-le-tue-avventure-grafiche/>
- [22] B. Schneier, *Applied cryptography: protocols, algorithms, and source code in C*. John Wiley & sons, 2007.
- [23] A. Sliney and D. Murphy, "Using serious games for assessment." in *Serious Games and Edutainment Applications.*, London, Springer, 2011, pp. 225-243.
- [24] RinkWorks, *Developing your own Adventure Game*, [Online] Available: <http://www.rinkworks.com/smash/tutorial/own.shtml> Accessed: 2015-02-30
- [25] A. Yusoff, R. Crowder and L. Gilbert., "Validation of serious games attributes using the technology acceptance model.," in *Second International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*. IEEE, 2010.
- [26] F. Bellotti, B. Kapralos, K. Lee, P. Moreno-Ger and R. Berta, "Assessment in and of serious games: an overview." *Advances in Human-Computer Interaction*, 2013, 1.
- [27] A. Graesser, P. Chipman, F. Leeming and S. Biedenbach, S. "Deep learning and emotion in serious games". *Serious games: Mechanisms and effects*, 2009, pp. 83-102.
- [28] The Interactive Fiction Wiki, "Cruelty Scale," [Online]. Available: [http://www.ifwiki.org/index.php/Cruelty\\_scale](http://www.ifwiki.org/index.php/Cruelty_scale) . [Accessed on 2015-02-30]
- [29] Bifulco, I., Francese, R., Lettieri, M., Liscio, L., Passero, I., & Tortora, G.. *The TIE Project: Agile Development of a Virtual World Serious Game on Waste Disposal*, 2011. In *DMS* (pp. 204-209).
- [30] Schwaber, K; & Beedle; M. *Agile Software Development with SCRUM*. Prentice Hall, 2002.

# Reward Points Calculation based on Sequential Pattern Analysis in an Educational Mobile App

Boshi Li

School of Computing and Information Systems  
Athabasca University  
Edmonton, Canada  
Snickers8523@gmail.com

Maiga Chang

School of Computing and Information Systems  
Athabasca University  
Edmonton, Canada  
maiga.chang@gmail.com

Rita Kuo

Independent Researcher  
Buffalo, USA  
rita.mcs1@gmail.com

Kristin Garn

Mathtoons Media  
Kelowna, Canada  
mathtoons@gmail.com

**Abstract**—In recent years, learning on smartphones has become a significant trend in education. The educational mobile app, Practi, provides a platform that can let students practice their knowledge of math and science. Practi gives students reward points when they finish a course or solve a question to encourage them to keep using the app. Students can use these reward points to redeem in-app items. However, a pre-defined amount of reward points does not always fit every student's situation. For instances, when most of students feel the question is difficult, the students who solve the question easily may be given more rewards. On the other hand, when a student achieves mastery in the required skills of a particular course or set of questions, he or she should be awarded more points. In order to make Practi capable of giving proper reward points according to students' question-solving behaviours, this research designs an Apriori based algorithm that can extract students' behaviours patterns and give students appropriate reward points according to the result of pattern comparisons.

**Keywords:** Association Rules; Sequential Patterns; Difficulty Analysis; Educational Game; Reward points; Apriori; Algorithm; Mobile app

## I. INTRODUCTION

Canada leads many other countries regarding of the percentage of its classrooms with access to high speed Internet [1]. Although Canadian provinces such as Alberta have been studying the benefits of allowing students to use their own mobile devices in class [2] and a large number of Canadian students use mobile devices in schools [3], few report being able to use their own personal devices.

One area of concern in Canada is the widening skills gap as students exit high school and enter post-secondary unprepared for academic rigour [4]. The latest results from the Organization for Economic Co-operation and Development (OECD) for International Student Assessment shows Canadian scores in mathematics dropping

DOI reference number: 10.18293/DMS2015-047

significantly [5]. Many Canadian schools are now looking for innovative ways to raise students' math skills in a short period of time and most believe that technology will be of help.

**Practi** is an educational mobile application that lets students engage in meaningful, gamified skills practices on their own iOS or Android devices by completing quizzes, interacting with classmates and tracking their own performance as Fig. 1 shows.

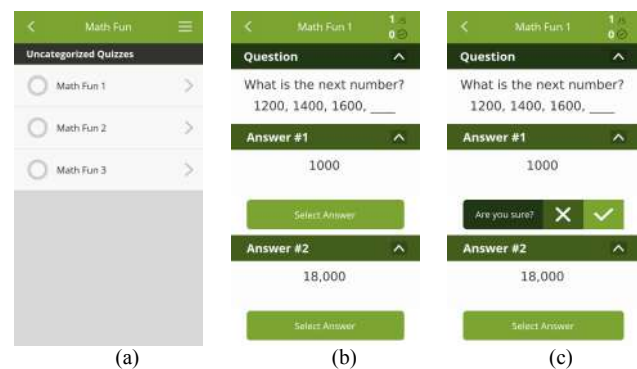


Figure 1. User Interfaces of Practi

Fig.1(a) shows a list of quizzes created for the student mobile practice app, Practi. Fig.1(b) shows a question and its answers. Practi will confirm with the student when he or she chooses a correct answer for the question as Fig.1(c) illustrates. The goal of Practi is to help students foster deeper subject engagement so that more students will practice what they learned to be successful in academics subjects. Having rewards (e.g., symbolic stars, gifts, and reward points) for students' practice results is a proven method to increase student retention [6] and the use of the rewards will serve to inspire and lengthen skills practice. This paper is organized as follows: In Section II, we review



related research of educational data mining approaches and adopt one of the approaches. Section III analyzes students' behaviours of solving a question, extracts students' problem solving patterns, and understanding of the perceived difficulty towards the question that most of students have by finding the frequent patterns. Section IV introduces the proposed algorithm that analyzes a question's perceived difficulty and gives fairness reward points for students according to their behaviour patterns. Section V briefly explains our evaluation plan for the proposed algorithm's accuracy, usability and effectiveness. At the end, Section VI summarizes the research and discusses the possible future work that we can do further.

## II. EDUCATIONAL GAME DATA MINING

### A. Educational Data Mining

Currently, Practi collects large amounts of students' data while students answer questions and interact with the app. How to find useful information from the data is a noteworthy topic. In the area of educational data mining research, researchers use different kinds of techniques, e.g. decision tree, neural networks, and association rules extraction, to find implicit and interesting information [7].

Brijesh and Pal have used decision trees to evaluate students' study performances and to identify students who need extra attention by their teachers at the end of semester [8]. In a recent study by Moucary and colleagues developed a hybrid system based on neural networks and data clustering to predict students' Grade Point Averages according to their foreign language performances [9]. Their system allowed teachers to identify students' capabilities and performances at an early stage to give students advice on registering for a courses and maintaining higher retention rates.

### B. Association Rule Extraction Approaches

Association rule extraction is a method that can be used for discovering relationships among patterns in large databases. It is widely used in medicine, business, and education. Batal and colleagues designed an algorithm that can identify frequent time-series symptom patterns from electronic medical databases to help doctors diagnose the possible diseases their patients may have [10]. Merceron and Yacef also designed an association rule extraction algorithm for a web-based educational system [11]. They extracted students' sequential patterns of learning so that the system would be capable of understanding students' learning progress and give proper feedback to students.

A supermarket has a database to store the transactions made by its customers. A transaction may contain more than one item sold in the supermarket. For instances, a customer may purchase potatoes only and another customer may purchase a box of juice and potatoes at the same time. When a customer purchases only potatoes in a transaction, the transaction's length is one. On the other hand, a transaction's length is two when the transaction contains two items. Association rule extraction algorithms first find

frequent items from the database. We set a threshold to find which items are frequently being seen in the database.

A frequent item may have more than one item. Similarly, if a frequent item is composed of a single item, then the frequent item's length is one. On the other hand, a frequent item's length will be more than one if it is composed of multiple items. For instance, if more than 20% of transactions in the supermarket's database contains both of a box of juice and potatoes (meaning than 20% of customers purchase these two products together at a single visit to the supermarket), then the combination of the two items is also treated as a frequent item with a length of two. In general, association rule extraction algorithms will set a maximum length that frequent items should have in order to determine a searching termination point.

### C. Finding Patterns of Time-Series Behaviour

Since students' question answering behaviours in Practi are also time-series, association rules can be extracted and used to discover the most frequent behaviours that students have while solving a particular questions. In this paper, we analyze students' behaviours when they repeatedly solve a particular question in Practi. The proposed algorithm first finds patterns of students' time-series behaviours while solving a particular question. Second, the algorithm analyzes these patterns and finds the frequent patterns that allow the system to understand the perceived difficulty students have towards each question they attempt to solve. Third, the algorithm calculates the proper number of points to give to the student according to his or her question solving patterns, and the performance of other student who solved the question before.

## III. STUDENTS' QUESTION ANSWERING BEHAVIOUR ANALYSIS

### A. Identifying Patterns

In order to get an idea of how students solve a question in Practi, the research team categorized students' question solving behaviors into 8 patterns. Table I shows the eight basic patterns.

TABLE I. EIGHT PATTERNS OF STUDENTS' QUESTION SOLVING BEHAVIOURS

Patterns	Difficulty	Equation	Weight	Symbol
Correct and attempts (L) with time (L)	Easy	$=4+2+1$	7	A
Correct and attempts (L) with time (H)	Easy	$=4+2-1$	5	B
Correct and attempts (H) with time (L)	Easy	$=4-2+1$	3	C
Correct and attempts (H) with time (H)	Normal	$=4-2-1$	1	D
Skipped and attempts (L) with time (L)	Normal	$= -4+2+1$	-1	E



Skipped and attempts (L) with time (H)	Hard	=-4+2-1	-3	F
Skipped and attempts (H) with time (L)	Hard	=-4-2+1	-5	G
Skipped and attempts (H) with time (H)	Hard	=-4-2-1	-7	H

The first column shows what a student did to solve the question using the Correct and attempts (L) with time (L), pattern as an example. This pattern represents a student who has finished solving a question with the correct answer in fewer attempts and in shorter time frame than average. (The second column is the corresponding visualized data retrieved from the database.) The “Difficulty” column represents the researchers’ perceived level of difficulty of the students corresponding behaviours.

Assigning weight for each pattern can make algorithm capable of recognizing patterns and calculating students’ pattern sequences. The research team assigned a numeric value for each case in a pattern. Practi allows students to revise their answers until they have chosen the correct answer or skipped the question. Answering a question correctly and skipping a question are two basic cases of solving a question. We used plus 4 (+4) for the Correct case and minus 4 (-4) for the Skipped case. Before a student correctly answers a question or decides to skip a question, he or she can attempt to answer the question many times. We used Less (L) and High (H) to represent whether or not a student’s trial number is lower or higher than the average. For the Less case, a plus 2 (+2) is given and a minus 2 (-2) is given for the High (H) case. Similarly, we used Low (L) and High (H) to represent whether or not a student spent much more time than the average and gave the question solving pattern plus 1 (+1) and minus 1 (-1).

The “Equation” column shows the weight calculation for each pattern. Take pattern #1 in the first row as an example. This pattern represents the students solving a question correctly (Correct) by trying less than the average (i.e., attempts L) and spending less time than the average (i.e., time L). We sum up each value of the behaviours that occurred in pattern #1 and the calculated result is shown in the “Weight” column to let the algorithm to recognized students’ behaviours and their perceived. The last column is the symbol which is used to represent the corresponding pattern in this paper.

#### B. Finding the Frequent Patterns

After converting all students’ question solving behaviours into patterns, the next step was to find patterns that had been seen more frequently. For example, if a pattern is repeatedly seen in the database, (representing, say, more than 8% of patterns in the database) then the pattern is a frequent pattern. Besides the frequency of a pattern occurring in the database, the research team also set the frequent pattern’s maximum length to three.

#### C. Understanding Students’ Perceived Question Difficulty

Once frequent patterns were found, the research team used these patterns to get an idea of how most of students’ perceptions of the questions in terms of their difficulty levels. Each frequent pattern’s weight was calculated based on the percentage that the pattern occupied in the database. With the calculation results, we can calculate each question’s perceived difficulty.

### IV. REWARD POINTS CALCULATION

#### A. Frequent Patterns Finding

Practi records all actions the students make when they solve a question. Table II is an example database which contains 12 students’ pattern sequences while solving a particular question, #2212. Each student may solve same question multiple times, hence, each student’s pattern sequence has different lengths.

TABLE II. AN EXAMPLE DATABASE CONTAINS 12 STUDENTS’ PATTERN SEQUENCES OF SOLVING QUESTION #2212

Question ID	User ID	Pattern sequence
2212	Andy	H, F, D, D, B, A, E
2212	Ben	D, C, E, D, C, D
2212	Carl	H, D, E, C, A, E
2212	David	A, A, B, E, C, E, E
2212	Anthony	A, C, A, A, B
2212	Derek	E, D, C, D, A, A, B
2212	Evan	F, D, A, A, B
2212	Bill	G, G, D, C, D, A, A, B
2212	Adam	F, G, C, A, A, B
2212	Edwin	F, F, D, D, B, B
2212	Denny	C, D, C, D, A, A, B
2212	Edgar	E, E, B, B, A, A, B

Algorithm 1 shows the procedure for finding frequent patterns. To limit the time spent on searching, the research team set a limit of 3 for a **pattern’s maximum length**. At the first scan (i.e., when  $i$  equals to one and only Lines #6 to #11 will be executed), the algorithm finds all frequent patterns whose length is 1 and stores them in the array CandidateList. The threshold of frequency  $\sigma$  is set to 20% means that the patterns that are not so often seen in the database will be filtered. Taking Table II as an example, pattern “A” occurs 11 times and there are 12 transactions in the database. Hence, the percentage of pattern “A” is 91.66% and is more than 20%. Therefore, pattern “A” is taken as a frequent pattern at first scan.

During the second scan (when  $i$  equals to 2), the algorithm extends a frequent item which only has one pattern (i.e., the item’s length is one) from the array CandidateList with any possible pattern that the students may have. The algorithm then checks if the extended items happened frequently (i.e., the percentage the items existed

in the database is higher than the threshold  $\sigma$ ). If three basic patterns “A”, “B”, “C” are frequent items, then “A” is extended with the eight basic patterns and forms “A, A”, “A, B”, “A, C” to “A, H”. The percentages with different combinations are 58.33% for “A, A”, 41.66% for “A, B”, 8.33% for “A, C” and so on. Because more than 20% of the transactions contain “A, A” and “A, B” patterns, “A, A” and “A, B” are treated as frequent items. The algorithm uses the same procedure for the other two frequent items, “B” and “C”, until the found frequent item’s length is larger than the predefined maximum length.

---

#### ALGORITHM 1: Finding frequent patterns

---

**Input:** All patterns  
**Output:** Frequent patterns

```

1  Max Frequent Pattern’s length = 3
2  threshold of frequency  $\sigma = 20\%$ 
3  BasicPatterns = { pattern A to H }
4  CandidateList =  $\{\Phi\}$ 
5  P = new patterns
6  for j = 1 to BasicPatterns.length do
7    percentage = (number of transactions which
    contain BasicPatterns[j]) / (number of
    transactions stored in the database)
8    if (percentage  $\geq \sigma$ ) then
9      CandidateList = CandidateList  $\cup$ 
      BasicPatterns[j]
10   end if
11 end for

12 for i = 2 to Max Frequent Pattern’s length do
13   for j = 1 to CandidateList.length do
14     for k = 1 to BasicPatterns.length do
15       P = CandidateList[j] + BasicPatterns[k]
16       percentage = (number of transactions that
       contains P) / (number of transactions
       stored in the database)
17       if (percentage  $\geq \sigma$ ) then
18         CandidateList = CandidateList  $\cup$  P;
19       end if
20     end for
21   end for
22 end for
23 return CandidateList;
```

---

#### B. Question’s Difficulty Analysis

After all frequent patterns in the database are found, the algorithm uses these frequent patterns to analyze how most of students perceive a question’s difficulty. In order to achieve this, the algorithm first needs to know each frequent pattern’s meaning.

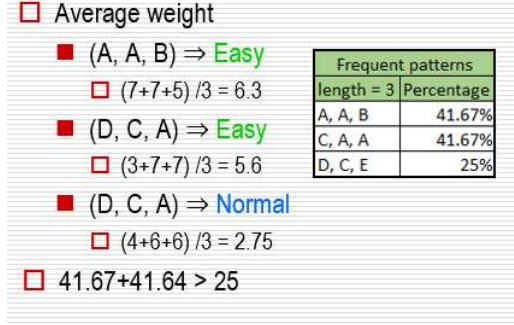


Figure 2. Calculating frequent patterns’ weights and summing up the percentages of the patterns with the same meaning.

Take frequent pattern “A, A, B” in Fig. 2 as an example: the first symbol in the pattern is “A” whose weight is 7. The algorithm will calculate the average weight of a pattern accordingly – the average weight of pattern “A, A, B” is 6.3. According to Table I:

$$7 \leq \text{Easy} \leq 3. \quad (1)$$

$$-3 < \text{Easy} < 3. \quad (2)$$

$$-3 \geq \text{Easy} \geq -7. \quad (3)$$

The average weight of pattern “A, A, B” satisfies (1). The result means that 41.67% of students have pattern “A, A, B” while solving the question and their perceived difficulty of the question is easy. Similarly, pattern “C, A, A” also represents Easy and the overall percentage of students who think the question is easy is 83.34%, i.e. 41.67% plus 41.67%. On the other hand, the pattern “D, C, A” represents the students who feel the question difficulty is Normal and there are only 25% of students who have this pattern while solving the question. As 83.34% is larger than 25%, which means most of the students think the question is **Easy**.

#### C. Points Awarded

As previously mentioned earlier, students are allowed to practice the same questions multiple times and the objective of Practi is to engage students in practicing the ones they struggle with. In order to motivate students to practice, the algorithm gives extra reward points for students according to students’ performances of past and current trials. Fig. 3 shows the criteria for calculating extra points for students according to a question’s perceived difficulty as well as students’ question solving patterns in the past and current trials.

Extra points				
Question is <b>easy</b>				
Past \ Current	Easy	normal	Hard	
Easy	0%	0%	0%	
normal	5%	0%	0%	
Hard	15%	10%	0%	

Extra points				
Question is <b>normal</b>				
Past \ Current	Easy	normal	Hard	
Easy	0%	0%	0%	
normal	10%	0%	0%	
Hard	20%	15%	0%	

Extra points				
Question is <b>hard</b>				
Past \ Current	Easy	normal	Hard	
Easy	0%	0%	0%	
normal	15%	0%	0%	
Hard	25%	20%	0%	

Figure 3. Criteria of calculating extra points

Past patterns represent how a student perceived the difficulty of a question during earlier attempts. The algorithm uses a similar method. It calculates the average weight for past patterns in order to obtain the perceived difficulty. Using Table II as an example, Ben just finished his fifth trial of solving the question, #2212.

The pattern of the four trials in the past is “D, C, E, D”. In order to know how difficult he feels the question was in the past, the algorithm calculates the average weight of the pattern “D, C, E, D”, that is  $(1+3+(-1)+1)/4 = 1$ . This number indicates Ben felt the question was Normal according to (2). On the other hand, Ben’s current pattern suggests that he now feel the question is Easy. Due to the fact that most students feel this question is Easy as Fig. 2 shows, the algorithm gives Ben 26.25 reward points which include 25 basic reward points and 5% extra reward points according to Fig. 3.

## V. EVALUATION PLAN

### A. Accuracy Testing

This paper proposed an automated algorithm to calculate reward points for students. However, the algorithm’s accuracy should be evaluated with the help of teachers. The research team plans to find five to six teachers to comment on whether or not the given reward points are appropriate for the students according to the replays of their question solving processes. If teachers’ comments are positive and consistent, then the algorithm is effective. If not, then the research team will revised the rules of calculating extra points (as Fig.3 shows) for students to fit most of teachers’ feedback.

### B. Evaluating the Effectiveness of the Algorithm

Giving proper reward points for students to get students motivated in practicing more is the research’s goal. Thus, an experiment with two groups is designed. The students in the first group will be given reward points by the algorithm and the students in the second group will always receive same reward points for solving a question first time. In the end of the experiment, a questionnaire contains learning motivation scale and usability analysis items is given to the students. The research team will use quantitative data analysis approaches to figure out whether or not students’

learning motivation gets enhanced and how the students think of the usability of the proposed algorithm.

## VI. CONCLUSION

The proposed algorithm uses an association rules extraction approach to find the frequent patterns of students’ question solving behaviours. With these frequent patterns we can know how most of students perceive a question’s level of difficulty. Once the algorithm identifies a question’s difficulty, it can calculate reward points for students based on past and present patterns of solving the question in the past and right now. This paper proposed a new way to give appropriate reward points for students according to their performance improvement and efforts.

## REFERENCES

- [1] R. Hirshhorn, Assessing the Economic Impact of Copyright Reform in the Area of Technology-Enhanced Learning, [online] 2011, <https://www.ic.gc.ca/eic/site/ippd-dppi.nsf/eng/ip01106.html#note42> (Accessed: 03 June 2015).
- [2] Alberta Education, Bring Your Own Device: A Guide for Schools, Technology in Schools, [online] 2012, <http://www.k12blueprint.ca/sites/default/files/byod%20guide%20revised%202012-09-05.pdf> (Accessed: 03 Jun 2015).
- [3] A. Joyce, Canadian Students are Massively Mobile, [online] 2014, [https://www.londondrugs.com/on/demandware.static/Sites-LondonDrugs-Site/Sites-LondonDrugs-Library/default/v1427703642968/pdf/news/08-14-14\\_BTS\\_Cell.pdf](https://www.londondrugs.com/on/demandware.static/Sites-LondonDrugs-Site/Sites-LondonDrugs-Library/default/v1427703642968/pdf/news/08-14-14_BTS_Cell.pdf) (Accessed: 03 June 2015).
- [4] C. Alphonso, Canada’s fall in math-education ranking sets off alarm bells, The Globe and Mail, [online] 2013, <http://www.theglobeandmail.com/news/national/education/canadas-fall-in-math-education-ranking-sets-off-red-flags/article15730663/> (Accessed: 03 June 2015).
- [5] P. Brochu, M. A. Deussing, K. Houme and M. Chuy, Canada’s students slipping in math and science, OECD finds, CBC news Canada, [online] 2013, <http://www.cbc.ca/news/canada/canada-s-students-slipping-in-math-and-science-oecd-finds-1.2448748> (Accessed: 03 June 2015).
- [6] B. Magerko, C. Heeter J. Fitzgerald, and B. Medler, "Intelligent Adaptation of Digital Game-Based Learning," In the Proceedings of the 2008 Conference on Future Play: Research, Play, Share, Toronto, Ontario, Canada, pp. 200-203, November 2008.
- [7] C. Romero, and S. Ventura, "Educational data mining: A survey from 1995 to 2005," Expert Systems with Applications, Vol. 33, No. 1, pp. 135-146, July 2007.
- [8] B. K. Baradwaj, and S. Pal, "Mining educational data to analyze students' performance," International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, pp 63-69, January 2012.
- [9] C. E. Moucary, K. Marie, and Z. Walid, "Improving student’s performance using data clustering and neural networks in foreign-language based higher education," The Research Bulletin of Jordan ACM Chapter – ISWSA, Vol. II, pp. 27-34, [online], April 2011. <http://ijj.acm.org/volumes/volume2/no3/ijjvol2no3p1.pdf> (Accessed: 03 June 2015)
- [10] I. Batal, D. Fradkin, J. Harrison, F. Moerchen, and M. Hauskrecht, "Mining Recent Temporal Patterns for Event Detection in Multivariate Time Series Data," In the Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data mining, Beijing, China, pp. 280-288, 2012.
- [11] A. Merceron, and K. Yacef, "Educational Data Mining: a Case Study," In the Proceedings of 12th International Conference on Artificial Intelligence in Education, Amsterdam, Dutch, pp. 467-474, November 2005.

# Multimedia data integration and processing for E-government

Flora Amato\*, Francesco Colace<sup>†</sup>, Luca Greco<sup>†</sup>, Vincenzo Moscato\* and Antonio Picariello\*

\*Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione. University of Naples "Federico II", ITALY

Email: {flora.amato,vmoscato,picus}@unina.it

<sup>†</sup>Dipartimento di Ingegneria dell'Informazione, Ingegneria Elettrica e Matematica Applicata. University of Salerno, ITALY

Email: {fcolace,lgreco}@unisa.it

**Abstract**—Knowledge management has become a challenge for almost all E-government based applications where one of the main problem is the efficient management of great amounts of data. In order to efficiently access the information embedded in very large document repositories, techniques for semantic document management are required. They ensure improvement for a large and intense process of dematerialization and aim at eliminating or at least reducing, the amount of paper documents.

In this work, we present a novel model of digital documents for the improvement of the dematerialization effectiveness. This model represents the starting point for an information system that is able to manage the document streams in an efficient way. It takes into account E-government applications needs like the compliance with the laws and regulations in force and the adaptability to evolving technologies. At the best of our knowledge, the proposed model is one of the first attempts to give a single and unified characterization for the management of multimedia documents, pertaining to a bureaucratic domain as the E-government one, on which semantic procedures are used for the transformation of non structured documents (pertaining to specialized domain) into structured data, suitable for automatic processing.

Furthermore, an architecture for the management of documents life cycle is proposed, which provides advanced functionalities for semantic processing, such as giving formal structure to document informative content, information extraction, semantic retrieval, indexing, storage, presentation, together with long-term preservation.

**Keywords**—Ontology Learning, Ontology Population, Natural Languages Processing

## I. INTRODUCTION

E-Government processes are dedicated to the improvement of the efficiency, expensiveness and accessibility of public administration services: dematerialization activities, introduced for properly managing bureaucratic documents, are among the main tasks of the E-government works.

The core aspect related to a novel and efficient dematerialization process is the idea standing beyond the common document concept, that can be defined as the representation of acts, facts and figures directly made or by means of electronic processing, and stored onto an intelligible support<sup>1</sup>.

In other words, a document consists of objects such as text, images, drawings, structured data, operational codes,

programs and movies, that, according to their relative position on the support, determine the shape and, consequently the structure of the documents.

During the various E-government processing phases, which differs depending on applications domains, a document is processed and eventually stored on various kinds of media, as papers and photographic films.

In order to manage documents properly, Document Management Systems (DMS) are used. They were introduced in the early 1970 for converting paper documents into electronic images stored in computers.

Nowadays DMS are becoming the basis of most business information systems, giving user control over company knowledge, providing efficient retrieval and desktop integration, reducing error rates in documents manipulation and thus improving business performance.

With the use of standards for knowledge representation, DMS are evolving, from search engine, toward systems able to integrate semantic search procedures into companies business processes. These systems, however, are limited to provide additional semantic functionalities to existent document management features. At the best of our knowledge, there are a variety of semantic-based approaches to modeling multimedia content focusing on single media (e.g. images or sounds only) but exist only a few experiments[3] for processing more complex multimedia documents such as those dealt with in this work.

The aims of this semantic-based processing are to structure input documents and to allow automatic retrieval of targeted information, based on formal representation of the domain associated to the documents, defined in a semi-automatic way, starting from the processable documents themselves.

In this work, we propose a new model of multimedia document, suitable for E-government activities, that takes into account the requirements of the E-government applications, which, depending on authorities, final users or time, produce different representations of the same multimedia contents. For describing the proposed model and system architecture, we focused on the E-Health domain. This particular domain implies a proper massive document processing. Knowledge management activities, In particular, must be performed

<sup>1</sup>This definition accords, for example, to the Italian civil law [1]

in reliable, effective and error-free way. Hence, E-Health organizations needs to be supported with approaches aimed at assessing clinical guidelines, and supporting their correct and ecient execution. The reported examples, in particular, are taken from the sub-domain of the Electronic Clinical Records. According to the International Organization for Standardization (ISO) denition, an electronic clinical record means a repository of patient data in digital form, stored and exchanged securely, and accessible by multiple authorized users. It contains retrospective, concurrent, and prospective information, and its primary purpose is to set objectives and planning patient care, document the delivery of care and assess the outcomes of care [2].

For this reason we model presentation and informative content in a separate way, allowing the solution, among other things, of open problems related to technology evolution, different document formats and access rights. The proposed model is the starting point for an information system which integrates and processes, in the most efficient way, different multimedia data types (like as images, text, graphic objects, audio, video, composite multimedia, etc.). In particular, it allows: *i)* documents structuring *ii)* automatic information extraction from digital documents; *iii)* semantic retrieval; *vi)* semantic interpretation of the relevant information presented in the document, *v)* storing and *vi)* long term preservation.

The proposed system combines Object-Relational Database (ORDBMS) technologies, Natural Language Processing (NLP) techniques, proper domain and structural ontologies, and inference rules in order to automatically extract significant concepts instantiated to each document and to provide semantic querying facilities for users.

In the process for representation and use of domain, specific knowledge ontologies play an important role, helping to cope documents with metadata annotations for supporting the process of information structuring and retrieval.

The quality of the information retrieved is thus improved by exploiting the possibility to enrich and then refine the set of the retrieved documents by using reasoning techniques on the ontologically-defined relations[4].

The work is organized as follow: in the next section, an overview on Knowledge Modeling Methodologies is presented. In the third section the method for semantic processing implemented in the proposed system are introduced. In section 4 we report the preliminary experimental results and in section 5 present the related works together with a discussion of the original contribution of our proposal. Finally in section 6 we give our concluding remarks and we outline our future work.

## II. RELATED WORKS

First, we briefly report the state of the art on the Systems developed for the Document Management and then we focused on the system managing multimedia documents.

Starting from the 1980s, a number of vendors began to develop systems to manage paper-based documents. They include the management not only of printed and published

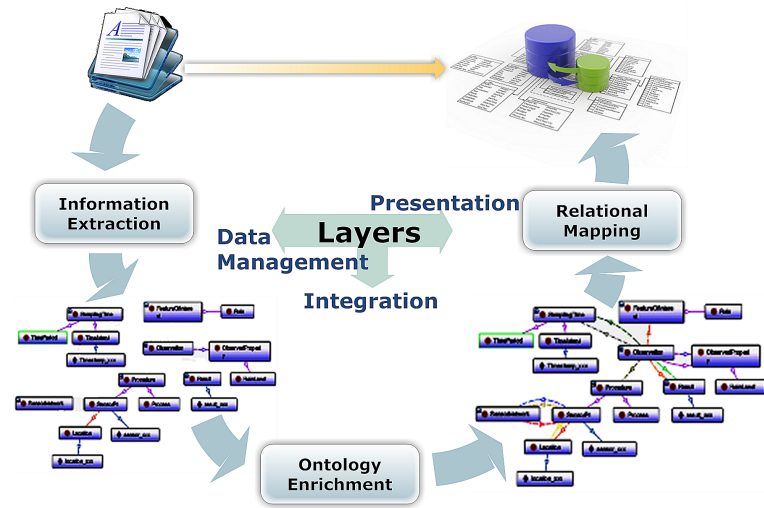


Fig. 1. General Schema of Documents Processing

documents, but also of photos, prints, etc. Recent Document Management Systems (DMS) are dedicated to the management of digital documents. This kind of systems commonly provides facilities for document processing as storage, versioning, metadata, security, as well as indexing and retrieval capabilities. In recent years numerous DMS projects suitable for specialist domains have been realized. These systems propose features for content characterization, offering for example, templates for documents semi-automatic generation. Nowadays DMS are moving toward semantic functionality, including advanced features for contents management like semantic search.

In Italy, numerous projects are presented for several specialist domains as the ASTREA Project realized by the Judicial Systems Research Institute (IRSIG) for the CNR (National Research Centre) during the years 2002-2006 and TAPA project realized in 2004 for the Anti-trust Authority (AGCM). Another relevant experience to be mentioned is the ESTRELLA project (European project for Standardized Transparent Representations in order to Extend Legal Accessibility), financed by the European Union (2006-2008). For what concerns the state of art in multimedia information management system one of the main research objective is the automatic indexing of multimedia data on the basis of their content in order to make query processing easier, more effective and efficient. In the following, supported by the related state-of-the-art, we describe the major challenges in developing reliable image and text database systems. The goal of image retrieval systems is to find out images from databases while processing a query provided by a user. In the last decade, most of researches are focused on Content Based Image Retrieval (CBIR). The CBIR is characterized by the ability of a system in retrieving relevant information on the base of image visual content and semantics expressed by means of simple search-attributes or keywords. Traditionally, CBIR addresses the problem of finding images relevant to



the users information needs from image databases, based principally on low-level image global descriptors (color, texture and shape features) for which automatic extraction methods are available, see [7] for details. More recently, it has been realized that such global descriptors are not suitable to describe the actual objects within the images and their associated semantics. Two main approaches have been proposed to cope with this deficiency: the first approach segments the image into multiple regions, and different descriptors are built for each region; the second approach exploits salient points identification techniques. Following the first approach, different systems like, SIMPLcity and Blobworld [8] have been developed. The second approach avoids the problem of segmentation altogether by choosing to describe the image and its contents in a different way. By using salient points or regions within an image, in fact, it is possible to derive a compact image description based around the local attributes of such points [9]. Our proposal [4] follows the second approach avoiding the problem of early segmentation and exploits color, texture and shape features in the principled framework of Animate Vision, according to which is the way that features are dynamically organized in the Where-What space that endows them with information about the context in terms of categories. Finally, more recent systems, such as Cortina and ALIPR [12] have as goal the automatic classification of images on the base of low-level features and high-level human annotations.

The textual processing phase requires the use of different techniques from interdisciplinary fields. Both theoretical and applicative fields have to be considered: the first for defining lexical dictionaries for legal domain, the second for organizing, storing and retrieving information of interest[11].

### III. KNOWLEDGE MODELING

In order to manage the composition of different multimedia data, their semantic relations and the structure imposed for bureaucratic documents, the defined document model is divided in three levels, as described in the following:

- **Data Management Layer:** describes the semantic content of each single media element (such as a text fragment or an image), providing functionalities for working on a single media; for example, information extraction and indexing over text are performed in this layer.
- **Integration layer:** describes the relations between the heterogeneous media components of the document, providing functionality for the integration of different data sources. For example the property of a text fragment of referring to an image belongs to this layer.
- **Presentation layer:** regulates the way by which the information has to be shown to users. It provides different representations of the same informative content, according to the formats, the final user's access rights and the available technology.

This approach allows the management of heterogeneous contents, by working on form and content independently,

enabling solutions of open problems related to technology evolution: in order to give a concrete example, it permits to give an immutable legal validity to the content of a document even if the format of representation changes, evolving with technology. On different layers of the document, information is semi-automatically tagged according to the concepts contained in the *domain ontologies*: associations among concepts and their instances are picked out. A general schema of documents processing is depicted in Fig. 1. Different ontologies can be used for the tagging process according to the different domains of interest. **Domain Ontology** is exploited to formalize the concepts of interest in the reference domain and relationships among them.

An example of top-level fragment of ontology in the domain of E-Health is depicted in Fig. 2, showing the relevant concepts and the semantic associations among them, occurring in a medical record. Specialized Domain ontology [13] can be divided into: **Structure Ontology** that describes how information is organized within the document and models the associations between the internal sections of the document and the set of concepts that can be found in it, and **Lexical Ontology** that contains the terms of the general language and can be used to refer wide-ranging concepts presented in the documents, not enclosed in the domain of reference.

#### A. System Architecture

Starting from the model, we have proposed an architecture, implemented in a prototype, for the management of the medical records life cycle. As already stated, medical records contain text that can be supplied with multimedia information as pictures (e.g. in radiographies), video streaming (e.g. in echographies) and audio. It is composed of three main modules: one for the text processing, one for the processing of the other media typologies of data, one for the management of the different formats of presentation, according to the requirements of the E-government applications. The modules that compose the system architecture, depicted in fig. 3 are described in the following: **Text processing** module aims at extracting relevant information from text, associating concepts to the terms of the text and defining relationships between them. The text is processed by a series of procedures each of which produces information usable by other procedures [14]:

- **Structural Analysis:** performs text segmentation and the related classification in order to identify the different sections constituting the structure of the document.
- **Linguistic Analysis:** performs procedures of Morpho-Syntactic analysis on the text (such as text tokenization and normalization, Part-of-Speech Tagging and lemmatization, complex terms analysis) combined with statistic procedures (such as the computation of opportune indices) enabling the extraction of relevant terms from the corpus to process. These terms and the information about them, refined with the help of domain experts, constitutes a lexicon that is exploited for the building



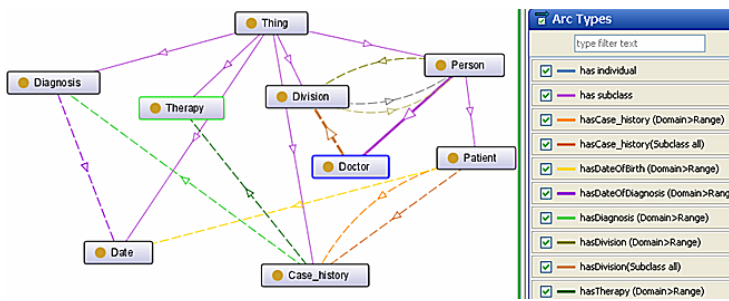


Fig. 2. A Fragment of Domain Ontology for electronic medical record

of the set of concepts used for domains formalization, performed by using ontologies.

- **Semantic Analysis:** by using the information of the early analyses, it detects properties and associations among terms, defining the concepts and relationships, allowing for ontology building and documents annotation.

The **Multimedia Data Processing** module has the aim of classifying each multimedia element, associating concepts from the domain ontology. It is composed of two components implementing innovative methods that have been presented in recent works [4][10]:

- **Analyzer** identifies relevant media parts and produces a low-level description that permits to create a series of indices to help the tagging and the retrieval.
- **Classifier** uses the indexing information to deduce which concepts, from the domain ontology, are being associated to media elements.

The **Presentation** module performs the dual task of combining the information about the heterogeneous contents and managing the ways by which they are presented to different users, according to the policies of the Entity (as the Public Administration), the final user's access rights and the available technology.

This module has also the aim to map the semantic information in a standard data format. For example, in E-Health domain, the module aims to map in Standard HL7 format the data semantically enriched with information about concepts and the implicit and explicit relations among them, coded in RDF triple. The module wraps the RDF data sets, translating the file into an XML based document, according to HL7 specifications, it works applying one or more XSLT transformation, according to the HL7 data format.

The list of the XSLT transformation rules is downloadable from our Document Processing project web site<sup>2</sup>.

The system is based on a multimedia database management system: it supports different multimedia data types (e.g. images, text, graphic objects, audio, video, composite multimedia, etc.) and, in analogy with a traditional DBMS, facilities for the indexing, storage, retrieval, and control of the multimedia data, providing a suitable environment for using and managing multimedia database information.

<sup>2</sup><http://www.unina.it/DOCProject.html>

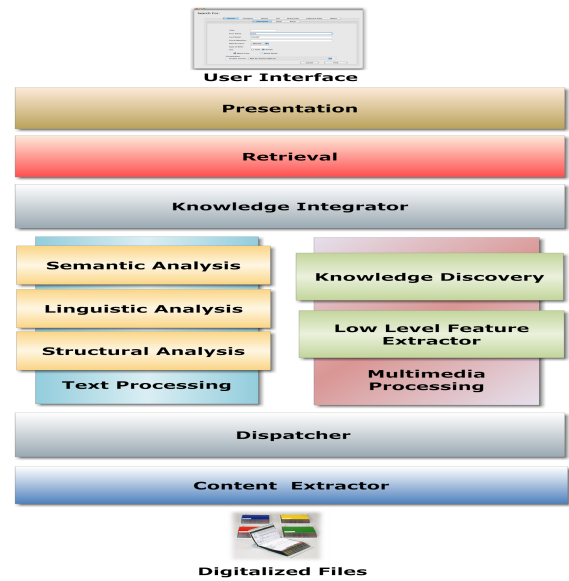


Fig. 3. System Architecture

More in details, a MMDBMS meets certain requirements that are usually divided into the following broad categories: multimedia data modelling, huge capacity storage management, information retrieval capabilities, media integration, composition and presentation, multimedia query support, multimedia interface and interactivity, multimedia indexing, high performances and distributed multimedia database management.

All document management system applications are designed on the top of a MMDBMS in order to support E-government processes in a more efficient way, in particular for those tasks regarding: automatic information extraction from documents, semantic interpretation, storing, long term preservation and retrieval of the extracted information.

The architecture of the proposed MMDBMS system, shown in Fig. 3, can be considered a particular instance of the typical MMDBMS architectural model and it is a suitable support for the management of E-government documents. The main components of the system are the modules delegated to manage the *Information Extraction and Indexing* process and those related to *Retrieval and Presentation* applications. All the knowledge associated to E-government documents is managed by proper *ontology repositories*.

In the current implementation of the system we realized three main separate subsystems that are responsible for the information extraction and the presentation tasks: one for the text processing related to e-doc, another one for processing the other kinds of multimedia information, in particular images, and the last one for presentation aims according to the requirements of public administrations.

The multimedia indexing and information extraction modules can be also specialized for other kinds of multimedia data like audio and video. In this case ad-hoc preprocessing

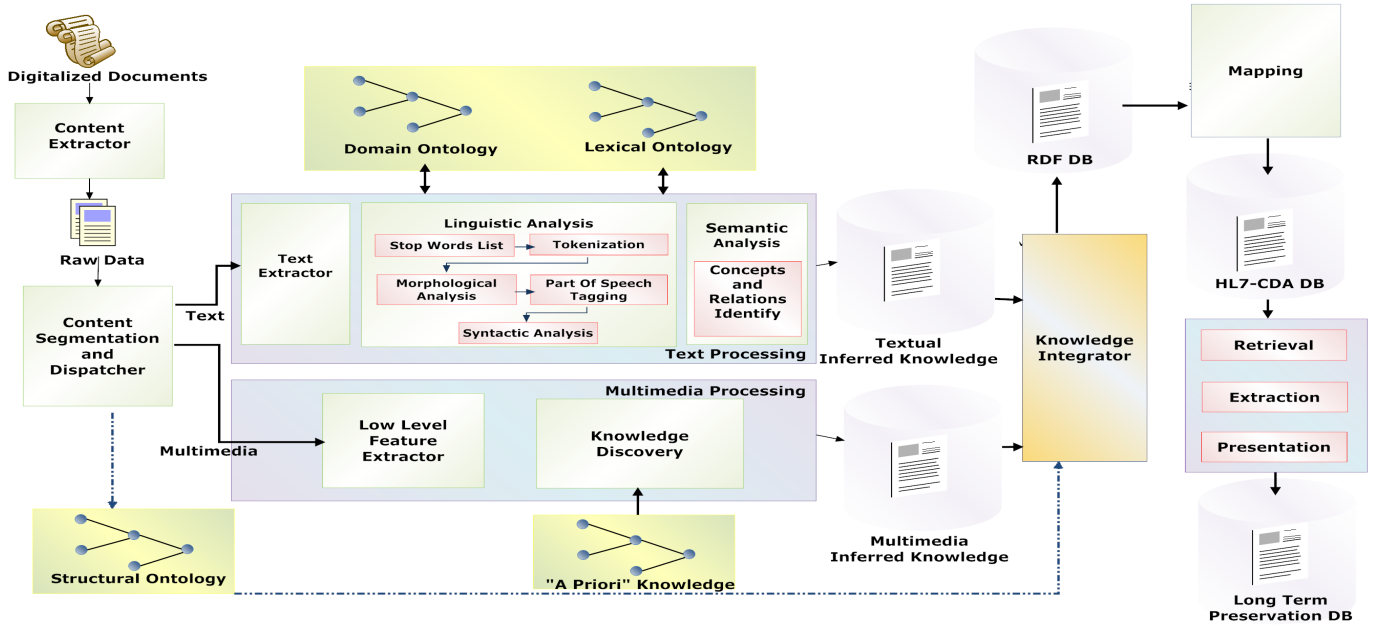


Fig. 4. Semantic Document Processing

components able to perform a *temporal segmentation* of multimedia flow are necessary to efficiently support the indexing process.

#### B. Proposed Multimedia Document Processing Methods

The whole process of document management, performed by the designed architecture, can be divided in Domain formalization and Final users application.

Domain formalization stage has the aim of codifying, with proper data structures (ontologies) the information of interest pertaining to the domain which the documents belong to. The information associated to contents is codified in terms of relevant concepts and relationships between them. Final users application stage implements the functionalities of document processing offered to the users in order to perform automatic operations on documents, such as searches by contents, long-term preservation and information representation according to different formats and different access policies.

1) *Information Extraction and Ontology Population* : Once associations between document segments and ontology fragments have been resolved, we proceed in populating concepts and relationships in the ontology fragment, by adding proper instances detected in document segments. Relevant information are then extracted, document segments are annotated and results are presented in *RDF* triples containing the properties identified in the segments.

Concepts and relations are extracted by exploiting an inference mechanism performed by a Rule-Based System. A generic rule is formed by a combination of token and syntactical patterns, which codifies the expert domain knowledge. In order to derive instances of relevant concepts or relationships, rules exploit:

- Named Entity Recognition (NER) functionality

- Morpho-Syntactic information obtained from NLP procedures performed in the Lexical Analysis,

eventually using subsumption on *TBox-Module* for deriving more specific concepts.

The detected instances can be shown by using tools like KIM[5], that highlights the associations among detected instances and the concept defined in the domain ontology.

The extracted relevant information is presented in *RDF* triples.

2) *Information Retrieval* : Once relevant information related to the domain of interest has been codified for document corpus, it is possible to execute semantic-based searches which are able to retrieve information by contents and not only by key-words.

The system we propose combines ORDBMS technologies, NLP techniques, proper domain structural ontologies management, and inference rules in order to retrieve significant concepts related to each document and to provide extended querying facilities for users. In particular, one of these facilities is the ability to perform advanced searches that overcome the limit imposed by “keyword-based” traditional queries. It also allows for a “content-based” access to documents database.

Traditional information retrieval systems, based on the comparison of sequences of characters, are in fact able to identify relevant concepts only if they are expressed with the same terms within the text: the search is always limited to the specific key-words inserted into the query and it excludes all the text parts where those keywords do not specifically appear. For instance, when searching for the word “house”, the system will ignore the documents where the words “home” or “residence” appear, even if they represent, in many contexts, the same concept. We

exploit, thus, semantic characterization of the document content, in order to improve the quality of the information retrieval. The domain specific knowledge is represented by means of Ontologies, that contain concepts and relationships. Instances of such elements are indicated in the documents by means of semantic annotations, performed by information extractions procedures.

When a user submits a query, the system identifies the concepts associated to the terms used in the query. These concepts are represented by means of ontologies as *synsets*, which are the set of linguistic elements linked by a synonymy relationship, i.e. terms that can be used in the same statement without modifying its whole meaning. Furthermore, same terms can be used with different acceptations (the meaning in which a word or expression is understood). In this case, different synsets are related to different meanings. If these ambiguities are present, the system will provide features to discriminate the synset of interest in the search.

Once users have selected the desired synset (all synsets are chosen if no selection is specified) a *query expansion*[6] mechanism is used in order to perform queries on corpus where all lemmas in the selected synsets become lemmatized keywords for a text-based search.

Query expansion techniques are used for dealing with the problem of word mismatch in information retrieval: retrieval system users and authors, in fact, often use different words to describe the same concepts in documents. The adopted query expansion approach requires that the query is expanded using lemmatized terms with the same meaning of the words used in the query. Thus, words within the same synsets are used for expansion and the match is not performed between single terms but between list of terms, which concern the concept to be retrieved in documents.

The collection of all the documents retrieved from these searches constitutes the results of the semantic-based query. A ranking algorithm is used to score results depending on a similarity measure, based on Tf-Idf index evaluation.

Notice that all query words and all relevant terms present in documents (which are also used for indexing purposes), have been reduced to their lemma, in order to make the search independent from different declinations and conjugations.

#### IV. IMPLEMENTATION OF THE MULTIMEDIA DOCUMENT PROCESSING

We implemented a prototypal version of the system that realizes the described data management procedures.

The proposed *Multimedia Document Management System* has the following main features:

- it exploits a unified data model that takes into account content-based and document-based characteristics;
- it uses ontological support for managing the semantics of data;
- it has a multi-layer architecture with different kinds of user interfaces;

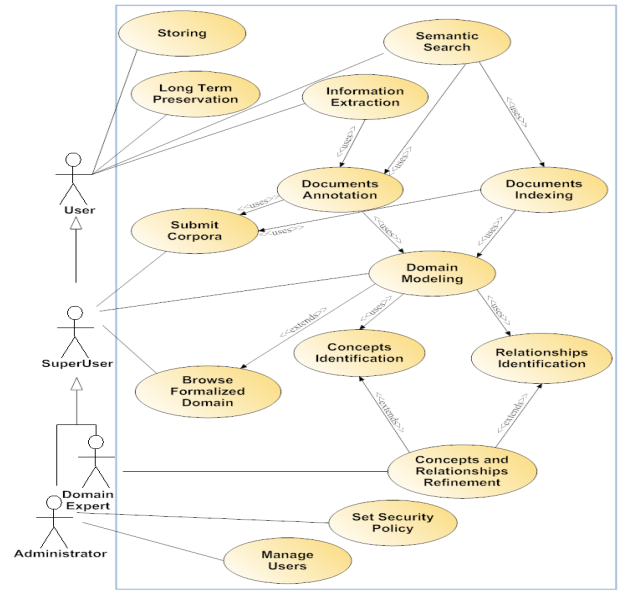


Fig. 5. Use Case Diagram of System Functionalities

- it provides advanced functionalities for document indexing and semantic retrieval. The system features are depicted in the user diagram in Fig. 5, in which the different functionalities (described in the previous paragraphs) are accessible only to the user with the appropriate privileges. Users can query the system, performing searches by content and information extraction, and use storing functionalities. Super-Users can also submit new medical documents or integrate the existent one, on which starting the process of domain modeling. Domain Experts can refine the modeled knowledge and the Administrators can manage users' proprieties and security policies too.

Fig. 6 shows at glance the Component Architecture of our system. Resources in the system are *Digital Documents* (DD) that are managed by a dedicated component, named *Digital Document Repository* (DDR). Its objectives are, from one hand, to allow for interoperability among the different data formats by providing import/export procedures and, from the other one, to manage security in the data access. Moreover, documents can be organized in specific *folders* to easy management and retrieval.

According to the introduced data model, it is possible to associate a digital document with a set of *semantic concepts* – retrievable by semi-automatic information extraction procedures and related to single content units of a document – and set of *keywords* – defined as particular properties of the whole document.

In the early stage, documents acquired by means of apposite OCR techniques are stored in the DDR and undergo the information extraction processing described in the following.

In the indexing stage, digital documents are picked up from DDR by a particular module called *Knowledge Discovery System* (KDS). The KDS analyses digital documents



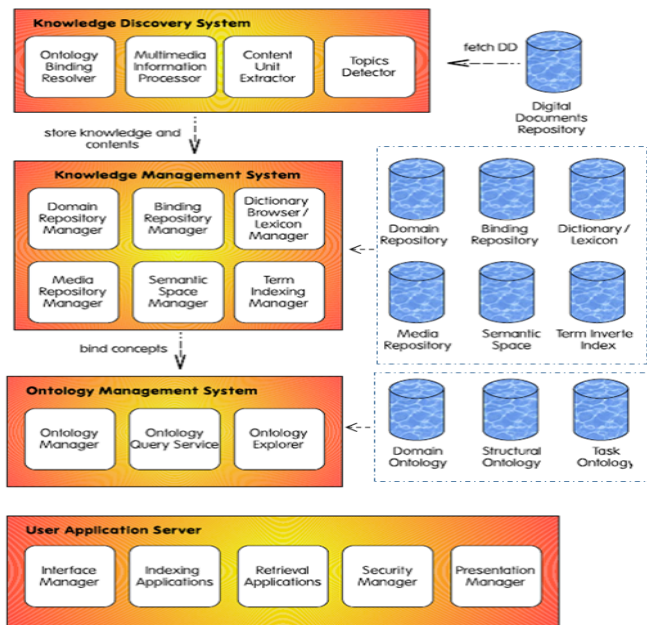


Fig. 6. Component Architecture

with the goal of obtaining useful knowledge from raw data. In particular, a *Content Unit Extractor* has the task of extracting (by a human-assisted process) content units from a document (and of generating an instance that can be stored in the system knowledge base), while, the *Multimedia Information Processor* sub-module infers knowledge in terms of semantic concepts from the different kinds of multimedia data[4] (e.g. text, audio, video, image). Furthermore, a *Topics Detector* sub-module operates on the not-structured view of a document and aims at detecting by natural language processing the most relevant topics for the whole document. Eventually, the *Ontology Binding Resolver* sub-module has the objective of creating for each discovered concept/topic a *binding association* with a node of domain ontology.

The extracted knowledge is then stored in the *Semantic Knowledge Base (SKB)* managed by a *Knowledge Management System (KMS)*. The KMS performs indexing operations on the managed information, providing features for the browsing and the retrieval of the documents. The components of the SKB (and the related KMS managing modules) are described in the following.

- *Dictionary* (for each supported language) - It contains all the terms of a given language with the related possible meanings and some linguistic relationship (e.g. WordNet). Each dictionary is managed by an apposite management module, called *Dictionary Browser*.
- *Lexicon* - It contains all the terms known by the system: dictionary terms and named entities (names of people and organizations). The lexicon is managed by a proper module, called *Lexicon Manager*.
- *Term Inverted Index* - It is the data structure used for indexing terms inside documents. For each term known by the system (and contained in the lexicon) a *posting*

*list*, that contains identifiers of documents and contents referring to terms with the related frequency, is created. The inverted index is managed by a *Term Indexing Manager*.

- *Semantic Space* - It allows for the storage of atomic pieces of knowledge belonging to document content units, which are called *document segments*. It is an abstraction of a shared virtual memory space (with read/write methods) by which applications can exchange multimedia data. This space is called “semantic” because each element is associated to a particular structural ontology that allows for relating segments of the same content unit to content units of different documents. The *Semantic Space Manager* provides functionalities for reading, writing, removing and searching tuples in the space.
- *Domain Repository* - It contains the description of application domain concepts and it is managed by a *Domain repository Manager*.
- *Binding Repository* - It contains the associations between document and domain repository concepts and it is managed by a *Binding Repository Manager*.
- *Media Repository* - It is an Object Relational DBMS able to manage different kinds of multimedia contents. It is managed by a particular module, called *Media Repository Manager* able to support classical multimedia query for the different kinds of multimedia data – e.g. *query by example/feature* for images, *query by content/keywords* for images and text, and so on.

The semantics associated to the data contained in the knowledge base is then managed by the *Ontology Management System (OMS)*, that contains the ontology models used by the system. In particular, we exploit three kinds of ontologies (managed by an *Ontology Manager*): (i) a set of *domain ontologies* that relate the semantic concepts in a given domain, (ii) a set of *task ontologies* that determine the role/meaning of a content unit in a document and (iii) a set of *structural ontologies* that code the relationships between contents and segments. The *Ontology Explorer* allows browsing of the concepts in the ontologies, while the *Ontology Query Service* is a component devoted to execute queries on the ontologies.

From the user point of view, the features provided by the system are the *indexing* of documents and the *semantic retrieval* of information. The application interfaces are realized both as web services and desktop programs (and managed by an *Interface Manager*). Finally, two modules are provided for the *security* and the *presentation* management.

## V. PRELIMINARY EXPERIMENTAL RESULTS

In this section we report some experiments we have carried out for evaluating the impact of the proposed system on enhancing user effort in indexing about 10000 medical records, properly anonymized, coming from an Italian health care organization. To set up our experimentation, we chose a sub-set of the collected data (constituted by 2000 randomly

chosen documents) as training set for training the classifier used for text segmentation. The objective in this experimentation is to evaluate the system correctness (precision) in automatically discovering relevant concepts of a medical document and in particular:

(1) Personal Data, (2) Diagnosis, (3) Diary of significant events, (4) Hospital discharge.

Relevant concepts discovery procedures exploit a domain ontology built from scratch from the medical records dataset, with the help of domain experts. Table 1 shows the related results and in particular the number of documents that has a given value of precision (100%: all the concepts have been correctly discovered, 75%: three concepts have been correctly discovered, 50% two and 25% only one concept has been correctly discovered).

In the majority of cases for which precision is 50% correct relevant concepts are the Personal Data and Hospital discharge, thus in our approach the most difficult concept to discover is that related to diary of significant events, probably due to the fact that such diaries are written in free text, also by different categories of medical users.

Eventually, table shows, on the right side, the average indexing times with respect to the document size<sup>3</sup>.

Precision	Documents	Doc. size	Indexing Time
100%	984	<150K	1, 2 s
75%	2024	150K ~ 300K	1, 8 s
50%	3713	300K ~ 500K	2, 5 s
25%	2498	500K ~ 1000K	2, 9 s
0%	781	>1000K	4, 8 s

TABLE I  
INDEXING PRECISION AND INDEXING TIMES

## VI. CONCLUDING REMARKS

In this work, we have defined a novel system for automatic processing of documents, based on semantic technologies. The realized semantic-based functionalities, as well as search by contents and information extraction, are based on the modeling of the relevant information of the domain of interest, codified by ontologies. Even if it is possible to provide as input data structures containing significant information, for example in form of lexicon for refinement purpose, the proposed system is able to define a formal representation for the domain of interest, in terms of concepts and relationships. The domain representation is built on the basis of the documental corpus, analysed in the early domain formalization phase. The formalization procedure is semi-automatic, because domain expertise can be exploited in order to refine ontologies, automatically built in a previous stage. The system, intended to be the core of an E-government information system, exploits the use of Linguistic and Semantic Analysis in order to transform unstructured (or semi-structured) documents into structured,

automatically processable records, codified by RDF triples. The system is designed for the management of documents belonging to specialized domains; the restricted area of specialization reduces the intrinsic semantic ambiguity of the words, related to the generalist domain, allowing more accurate information management operations. In order to perform semantic based document processing, we have defined a model for multimedia digital document, particularly suitable for processing data from E-government activities. The model is a starting point of a general framework for structuring, presenting and retrieving relevant information for a specialized domain. Experimental results (not reported for brevity) have shown encouraging results. Future direction will be devoted to improve the interoperability among the available procedures.

## REFERENCES

- [1] Deliberation of 13 dicembre 2001, n. 42, published on Gazzetta Ufficiale della Repubblica Italiana n. 296 of 21 dicembre 2001
- [2] Colantonio, S., Esposito, M., Martinelli, M., De Pietro, G., Salvetti, O. (2012). "A knowledge editing service for multisource data management in remote health monitoring". IEEE Transactions on Information Technology in Biomedicine, 16(6), 1096-1104.
- [3] Khoo, Michael, et al. "Towards digital repository interoperability: The document indexing and semantic tagging interface for libraries (distil)." Theory and Practice of Digital Libraries. Springer Berlin Heidelberg, 2012. 439-444.
- [4] Colace, F., De Santo, M., Greco, L., Moscato, V., Picariello, A. (2015). "A collaborative user-centered framework for recommending items in Online Social Networks". Journal of Computers in Human Behavior. 2015. Doi : <http://dx.doi.org/10.1016/j.chb.2014.12.011>.
- [5] B. Popov, A. Kiryakov, A. Kirilov, D. Manov, D. Ognyanoff, M. Goranov. "KIM – Semantic Annotation Platform". Book Chapter of The SemanticWeb - ISWC (2003). pp. 834 – 849. ISBN 978-3-540-20362-9-. Springer Berlin / Heidelberg.
- [6] Z. Jiuling , D. Beixing ,L. Xing , Concept Based Query Expansion Using WordNet, pp. 52-55, 2009 International e-Conference on Advanced Science and Technology, 2009.
- [7] R. Datta, and D. W. J. Joshi, "Image retrieval: ideas, influence, and trends of the new age", ACM Computing Survey, vol. 40, n. 2, pp. 5–64, 2008.
- [8] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blob world: image segmentation using expectation-maximization and its application to image querying", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, Issue 8, pp. 1026–1038, 2002.
- [9] J. S. Hare, and P. H. Lewis, "On image retrieval using salient regions with vector-spaces and latent semantics", Image and Video Retrieval (CIVR 2005), Singapore, Springer Ed., 2005.
- [10] Chianese A. and Piccialli F. "Designing a smart museum: when Cultural Heritage joins IoT." Next Generation Mobile Apps, Services and Technologies (NGMAST), 2014 Eighth International Conference on. IEEE, 2014.
- [11] Chianese, Angelo, and Francesco Piccialli. "SmaCH: A Framework for Smart Cultural Heritage Spaces." Signal-Image Technology and Internet-Based Systems (SITIS), 2014 Tenth International Conference on. IEEE, 2014.
- [12] B. S. Manjunath and et al. Cortina, "Searching a 10 million images database", Technical report, Sep 2007.
- [13] F. Amato, A. Mazzeo, V. Moscato, A. Picariello. "A System for Semantic Retrieval and Long Term Preservation of Multimedia Documents in E-Government Domain". To Appear in International Journal of Web and Grid Services, Vol. 5, No. 4, Inderscience Publishers, pp. 323.338(16), 2009.
- [14] F. Amato, A. Mazzeo, A. Penta, A. Picariello, "A semantic document management system for legal applications", International Journal of Web and Grid Services, Vol. 4, No. 3, Inderscience Publishers, pp. 251–266(16), 2008.

<sup>3</sup>All experiments presented in this Section were conducted on a Linux Cluster of 3 machines, each one mounting a 2GHz Intel Core i7 processor with a 8 GB, 1600 MHz DDR3

# Learning a Semantic Space by Deep Network for Cross-media Retrieval

Zhao Li<sup>† \* \*</sup>, Wei Lu<sup>†</sup>, Egude Bao<sup>†</sup>, Weiwei Xing<sup>†</sup>

<sup>†</sup> School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China

<sup>\*</sup> Shandong Computer Science Center(National Supercomputing Center in Jinan), Jinan Shandong 250014, China

<sup>\*</sup> Shandong Provincial Key Laboratory of Computer Network, Jinan Shandong 250014, China

liz@sdas.org {luwei, baoe, wwxing}@bjtu.edu.cn

**Abstract**—With the growth of multimedia data, the problem of cross-media (or cross-modal) retrieval has attracted considerable interest in the cross-media retrieval community. One of the solutions is to learn a common representation for multimedia data. In this paper, we propose a simple but effective deep learning method to address the cross-media retrieval problem between images and text documents for samples either with single or multiple labels. Specifically, two independent deep networks are learned to project the input feature vectors of images and text into an common (isomorphic) semantic space with high level abstraction (semantics). With the same dimensional feature representation in the learned common semantic space, the similarity between images and text documents can be directly measured. The correlation between two modalities is built according to their shared ground truth probability vector. To better bridge the gap between the images and the corresponding semantic concepts, an open-source CNN implementation called Deep Convolutional Activation Feature (DeCAF) is employed to extract input visual features for the proposed deep network. Extensive experiments on two publicly available multi-label datasets, NUS-WIDE and PASCAL VOC 2007, show that the proposed method achieves better results in cross-media retrieval compared with other state of the art methods.

**Keywords**—cross-media retrieval; cross-modal retrieval; deep learning.

## I. INTRODUCTION

Nowadays, with the development of Internet, an enormous amount of multimedia data, e.g., image, text documents and videos, have been generated. These data with various modalities usually co-occur to describe the same objects or events. For example, images are usually accompanied with a textual description to represent the same meaning. Learning the relationships among different modalities is becoming an interesting research topic which can benefit many important applications, such as multimedia retrieval and content creation. In this work, we address the cross-media retrieval problem between images and text documents, i.e., using an image to retrieve text and using text to retrieve images, as illustrated in Fig. 1. Although here we only focus on two modalities, i.e., image and text, our method can be easily adapted to other modalities.

During the past few years, many cross-media retrieval methods have been proposed [1], [2], [3]. As two typical methods, Canonical Correlation Analysis (CCA) [4] and

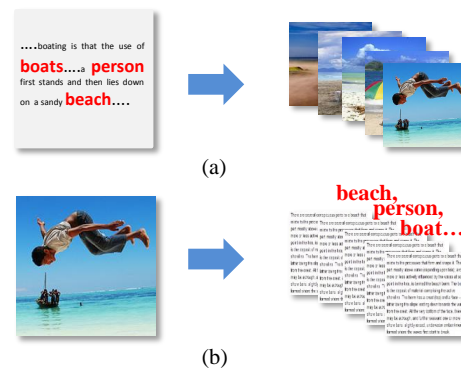


Figure 1: The cross-media retrieval task considered in this paper. (a) Using text as a query to retrieve relevant images. (b) Using an image as a query to retrieve relevant text.

Partial Least Squares [5], [6] are usually adopted to learn a couple of projections to maximize the correlations between two variables. Some methods have been proposed based on CCA. One of them is a Semantic Correlation Matching method [1], which leverages multi-class logistic regression to produce an isomorphic semantic space for cross-media retrieval. In [2], a generic framework called Generalized Multiview Analysis was presented to address multimedia problems. More recently, [3] proposed a multi-view CCA model by introducing a semantic view to achieve a better separation for multimedia data of different classes in the learned isomorphic space.

Although these methods have made contributions to the solution of cross-media retrieval tasks, their performance is still far from satisfactory. This is because the performance of cross-media retrieval between images and text is highly dependent on visual feature representation, but traditional feature extraction techniques have been undergoing a bottleneck period for image understanding in the past few years. Recently, significant progress has been made in image classification due to the development of convolutional neural networks (CNN) [7], [8], [9]. Especially, [9] has demonstrated promising results for image classification in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [10]. More recently, [11] has released an



open-source implementation of an eight-layer network called DeCAF, which is trained on ImageNet with 1,000 classes, and has demonstrated that CNN features are suitable as general features for various tasks.

In this paper, we propose a deep learning method to address the cross-media retrieval problem with multi-label. As shown in Fig. 2, two independent fully-connected deep networks are trained to project input feature vectors of images and text into an isomorphic semantic space with higher level abstraction. Specifically, we employ DeCAF to extract the visual feature of each image as the input feature vector of Image\_Net. For Text\_Net, the input feature vector can be extracted by many traditional feature extraction techniques, such as the bag-of-words. Since the proposed method is supervised, the correlation between two modalities can be built according to their shared ground truth probability vector. Both two networks have two hidden layers and one output layer, and the squared loss is employed as the loss function. Experiments on NUS-WIDE and PASCAL VOC 2007 demonstrate a significant performance improvement over other methods.

The remainder of this paper is organized as follows. We briefly review the related work of cross-media retrieval in Section II. In Section III, we present the proposed method in details. After that, experimental results and analysis are reported in Section IV. Finally, Section V presents the conclusions.

## II. RELATED WORK

Based on Canonical Correlation Analysis (CCA) [4], some methods [1], [2], [3], [12] are proposed to learn a common space for multimedia data of different modality, in which the distance between two media objects with similar semantics could be minimized while those with different semantics could be maximized. Specifically, Gong *et al.* [3] presented a multi-view CCA method via introducing a third view, i.e., semantic view, to better separate the multimedia data with different semantics in the learned latent common space. The semantic view representation can be obtained through supervised information as well as clustering analysis. Similarly, a cluster CCA method, which also focuses on learning discriminant common space to maximize the correlation of different kinds of multimedia data, was proposed by [12]. In this work, the separation for multimedia data with of different semantics was achieved via an unsupervised way.

Besides, with the ever-growing large-scale multimedia data on the Internet, much attention has been devoted to nearest neighbor search. To address this time-consuming problem, some hashing-based methods [13], [14], [15] have attracted a lot of interest. In [13], a cross view hashing (CVH) method was proposed to generate hash codes by minimizing the distance of hash codes for similar data while maximizing the distance for dissimilar data. Wu *et al.* [14] presented a sparse hashing method to obtain sparse code sets

for the data of different modalities through joint multimedia dictionary learning.

In addition, with the development of deep learning, some deep models [16], [17], [18] have been proposed to address multimedia problem. Specifically, Andrew *et al.* [17] adapt the CCA into the deep model to learn complex nonlinear transformations of different multimedia data. Based on Restricted Boltzmann Machine, Ngiam *et al.* [16] proposed to learn a shared representation between different modalities of multimedia data.

## III. THE PROPOSED METHOD

In this section, we will detail the proposed deep learning method for cross-media retrieval. We will first describe the architectures of the two deep networks as well as the training parameters, and then introduce the Euclidean loss function used in the training process.

### A. Network Architecture

As shown in Fig. 2, we build two independent networks, i.e., Image\_Net and Text\_Net, to map images and text from their input feature spaces into a common semantic space respectively. Each network consists of two hidden layers and one output layer. For Image\_Net, we employ DeCAF for image feature extraction. Specifically, each image is firstly resized to  $256 \times 256$  and fed into DeCAF, which is pre-trained on the ImageNet dataset with 1,000 classes. Different from the previous work [11], which used CNN features (DeCAF<sub>5</sub>, DeCAF<sub>6</sub> or DeCAF<sub>7</sub>; refer to [11] for more details) to represent a given image, we utilize the 1000 dimensional predictive scores of each image as the input visual feature of the proposed deep network. The reason for this choice is that the predictive scores provide a probability distribution over 1,000 classes from the ImageNet dataset and the relationship between this kind of visual features and ground truth can be easily built. For Text\_Net, since textual features usually have greater discriminative power than traditional visual features (e.g., SIFT and HOG), the relationship between textual features and ground truth can be more easily built. Therefore, many feature extraction techniques, such as bag-of-words, can be employed to extract the input textual features for Text\_Net.

Denote  $\mathbf{h}^{(0)} \in R^{d_0}$  as the input feature vector of Image\_Net (or Text\_Net).  $d_t$  is regarded as the output dimension of the  $t$ th layer (the input can be considered as the 0th layer for convenience). The outputs of the subsequent three layers (two hidden layers and one output layer) can be defined as

$$\mathbf{h}^{(t)} = \sigma \left( W_t \mathbf{h}^{(t-1)} + \mathbf{b}_t \right), \quad t = 1, 2, 3, \quad (1)$$

where  $\mathbf{h}^{(t)}$  is the output vector,  $W_t \in R^{d_t \times d_{t-1}}$  is the matrix of weights and  $\mathbf{b}_t \in R^{d_t}$  is the vector of biases.  $\sigma(\cdot)$  is the activation function. In our work, we use the rectified linear units (ReLU) as the nonlinear activation function.

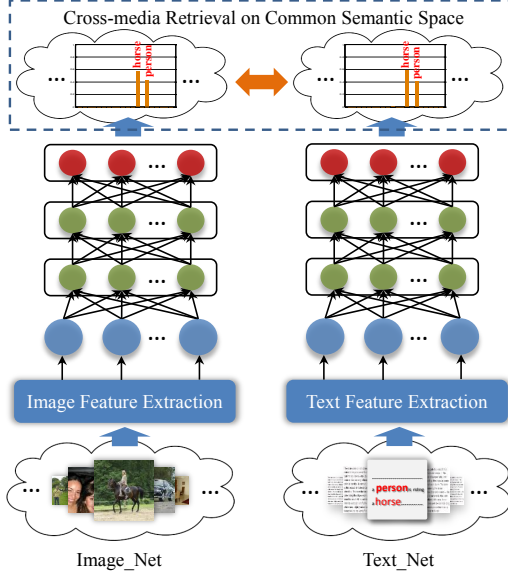


Figure 2: Two independent deep networks, Image\_Net and Text\_Net, are trained to project images and text from their original feature spaces into a common semantic space for the cross-media retrieval. Both networks have three layers, including two hidden layers and one output layer. Blue, green and red nodes represent input features, hidden units and output units, respectively.

The two networks are trained by stochastic gradient descent with momentum of 0.9. Besides, each hidden layer is followed by a dropout operation with a dropout ratio of 0.5 to combat overfitting. The global learning rate of these two networks is set as 0.01 at the beginning and dynamically changed according to the Euclidean loss (see below).

### B. Euclidean Loss

To achieve the target that pairs of image and text can retain similar feature representation in the common semantic space, we utilize Euclidean loss as the cost function to optimize both Image\_Net and Text\_Net. The output of the last layer is fed into a  $c$ -way ( $c = d_3$ ) softmax, which generates predictive scores over the  $c$  class labels. Suppose there are  $n$  image-text pairs in the training set. The predicted probability for the  $j$ th class of the  $i$ th input vector  $\mathbf{h}_i^{(0)}$  (image or text) can be defined as

$$P(y = j | \mathbf{h}_i^{(0)}) = \frac{\exp(f_j(\mathbf{h}_i^{(0)}))}{\sum_{k=1}^c \exp(f_k(\mathbf{h}_i^{(0)}))}, \quad (2)$$

where  $f(\cdot)$  can be considered as the mapping from the input layer to the output layer and  $f_j(\mathbf{h}_i^{(0)})$  is the activation value of  $\mathbf{h}_i^{(0)}$  on class  $j$ .  $y$  indicates the class label. Since the proposed method is targeted at multi-label problems, we can form a label vector  $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{ic}]$  for each image-text pair.  $y_{ij} = 1$  ( $j = 1 \dots c$ ) if the given sample is annotated

with class  $j$ , and otherwise  $y_{ij} = 0$ . We define the ground truth probability vector of  $\mathbf{h}_i^{(0)}$  as  $\hat{\mathbf{p}}_i = \mathbf{y}_i / \|\mathbf{y}_i\|_1$  and the predictive probability vector as  $\mathbf{p}_i = [p_{i1}, p_{i2}, \dots, p_{ic}]$ , where  $p_{ij} = P(y = j | \mathbf{h}_i^{(0)})$  ( $j = 1 \dots c$ ). Then, the cost function to be minimized can be defined as

$$J = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^c (p_{ik} - \hat{p}_{ik})^2. \quad (3)$$

## IV. EXPERIMENTS

### A. Dataset and Settings

We evaluate the performance of the proposed method compared with other methods on two publicly available datasets.

**NUS-WIDE [19]:** The dataset contains 269,648 images. Each image is accompanied with 81 ground truth labels (classes) and 1,000 text annotations. We select those pairs belonging to one of the 20 most frequent classes and ignore those pairs containing images without any ground truth label or text annotation. Then, a subset of 52,829 pairs for training and 35,216 pairs for testing can be obtained for evaluation.

**PASCAL VOC 2007 [20]:** There are 9,963 images of 20 classes in this dataset. A set of 399 words provided by [21] is employed as the textual description for each image. We conduct experiments on *trainval* and *test* splits, which contain 5,011 and 4,952 pairs respectively.

We experimentally set  $d_1 = 512$ ,  $d_2 = 256$ ,  $d_3 = 20$  for both Image\_Net and Text\_Net. Euclidean distance is used to measure the similarity between features in the semantic space. We compare the proposed method with three popular methods, including two unsupervised methods, Canonical Correlation Analysis (CCA) [4] and Partial Least Squares (PLS) [5] which only utilize the pair-wise information, and the linear version of one supervised method, Multi-view CCA (Multi-CCA)<sup>1</sup> [3]. We vary the embedding dimensionality for these methods, i.e., 20, 128 and 512, and report the best performance. All the methods project images and text from their input feature spaces into a 20 dimensional common semantic space.

To get effective feature representation of visual representations for both NUS-WIDE and PASCAL VOC 2007, DeCAF<sup>2</sup> is employed in our method to extract the 1,000 dimensional predictive scores as the visual feature for each image. We use the 1,000 dimensional bag-of-words features provided by [19] as the textual features for NUS-WIDE and use the 798 dimensional tag ranking features (relative and absolute) provided by [21] as the textual features for PASCAL VOC 2007. Besides, to validate the effectiveness of the visual features used in this work, we do a comparison with other visual features based on CCA, PLS and multi-view CCA. For NUS-WIDE, we use the 500 dimensional Bag-of-SIFT-Words (SIFT-BoW) [22] features provided by [19]

<sup>1</sup><http://www.unc.edu/~yunchao/crossmodal.htm>

<sup>2</sup><https://github.com/UCB-ICSI-Vision-Group/decaf-release>

Table I: Cross-media retrieval performance on the NUS-WIDE dataset (mAP scores).

Method	I2T	T2I	Average
CCA (SIFT-BoW)	0.226	0.205	0.216
CCA (DeCAF)	0.277	0.262	0.270
PLS (SIFT-BoW)	0.316	0.181	0.249
PLS (DeCAF)	0.203	0.314	0.259
Multi-CCA (SIFT-BoW)	0.353	0.280	0.317
Multi-CCA (DeCAF)	0.439	0.315	0.377
Proposed	<b>0.486</b>	<b>0.409</b>	<b>0.448</b>

Table II: Cross-media retrieval performance on the PASCAL VOC 2007 dataset (mAP scores).

Method	I2T	T2I	Average
CCA (GIST+HSV+SIFT-BoW)	0.368	0.345	0.357
CCA (DeCAF)	0.638	0.618	0.628
PLS (GIST+HSV+SIFT-BoW)	0.380	0.348	0.364
PLS (DeCAF)	0.351	0.574	0.463
Multi-CCA (GIST+HSV+SIFT-BoW)	0.475	0.436	0.456
Multi-CCA (DeCAF)	0.709	0.577	0.643
Proposed	<b>0.781</b>	<b>0.689</b>	<b>0.735</b>

as the visual representations. For PASCAL VOC 2007, the 776 dimensional visual features, each of which contains a 512 dimensional GIST [23] feature, a 64 dimensional color feature (i.e., HSV) and a 200 dimensional SIFT-BoW feature, provided by [21] are employed as the visual representations.

### B. Evaluation Metrics

In this paper, we consider two retrieval tasks, i.e., using image to retrieve text documents (I2T) and using text document to retrieve images (T2I). Retrieval performance is evaluated by mean average precision (mAP), which is one of the standard information retrieval metrics. In particular, given a set of queries, the average precision (AP) of each query is defined as:

$$AP = \frac{\sum_{k=1}^R P(k)rel(k)}{\sum_{k=1}^R rel(k)},$$

where  $R$  denotes the number of retrieved results.  $rel(k) = 1$  if the item at rank  $k$  is relevant,  $rel(k) = 0$  otherwise.  $P(k)$  is the precision of retrieved results ranked at  $k$ . We can get the mAP score by averaging AP for all queries. Since NUS-WIDE and PASCAL VOC 2007 are two multi-label datasets, it is regarded as a relevant result if the retrieved result shares at least one class label with the query.

### C. Results

Table I and Table II report our experimental results on NUS-WIDE and PASCAL VOC 2007, respectively. We can observe that the proposed method makes a significant improvement over any compared method on NUS-WIDE and PASCAL VOC 2007 (44.8% and 73.5%). This is because the proposed two deep networks can effectively build the relationship between the input feature vectors and

the shared ground truth probability vectors with Euclidean loss function. In addition, the improvement of our method may also depend on the effective visual features. It can be observed that DeCAF works better than other image feature extraction techniques. This observation is reasonable, since DeCAF is pre-trained with about 100 million labeled images (i.e., ImageNet), which can make the learned visual features have sufficient representational power. Fig. 3 shows some examples accompanied with their visual features and textual features in the learned common semantic space on PASCAL VOC 2007. From Fig. 3, it can be seen that each *image-text* pair usually has a similar probability distribution in the common semantic space, which can further validate the effectiveness of our method.

## V. CONCLUSIONS

In this paper, we proposed a deep learning method for cross-media retrieval. We trained two independent deep networks to map input feature vectors of images and text documents into an isomorphic semantic space, respectively. Especially, we took 1,000 dimensional predictive scores produced by an open-source CNN implementation called DeCAF, which is pre-trained on the ImageNet dataset with 1,000 classes, as the input visual features of Image\_Net. Extensive experimental results on NUS-WIDE and PASCAL VOC 2007 show that the proposed method can achieve a better performance in cross-media retrieval task compared with other methods.

## REFERENCES

- [1] N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. Lanckriet, R. Levy, and N. Vasconcelos, "A new approach to cross-modal multimedia retrieval," in *MM*, 2010, pp. 251–260.
- [2] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *CVPR*, 2012, pp. 2160–2167.
- [3] Y. Gong, Q. Ke, M. Isard, and S. Lazebnik, "A multi-view embedding space for modeling internet images, tags, and their semantics," *IJCV*, vol. 106, no. 2, pp. 210–233, 2014.
- [4] D. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.
- [5] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *Subspace, Latent Structure and Feature Selection*, 2006, pp. 34–51.
- [6] A. Sharma and D. W. Jacobs, "Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch," in *CVPR*, 2011, pp. 593–600.
- [7] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *ICCV*, 2009, pp. 2146–2153.

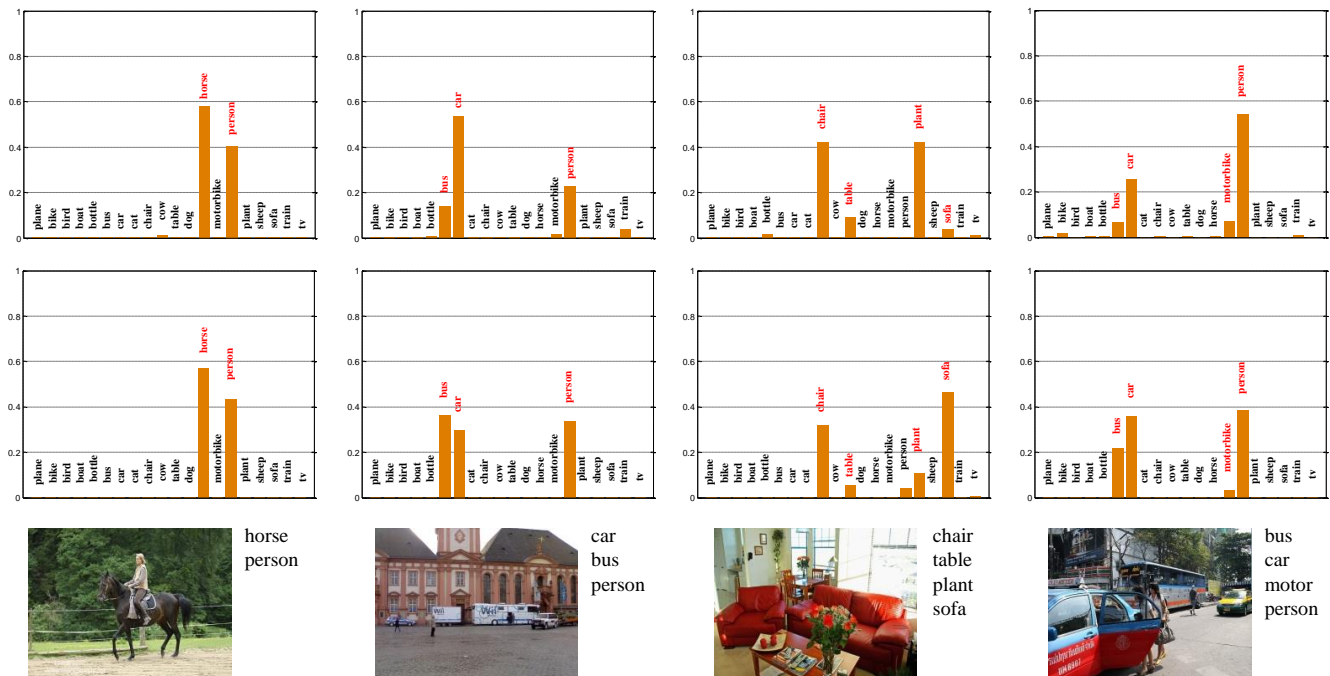


Figure 3: Some examples on PASCAL VOC 2007. The first and second rows plot the 20 dimensional semantic feature vectors of image and text corresponding to examples given in the third row. Ground truth labels are highlighted with red color.

- [8] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” in *ISCAS*, 2010, pp. 253–256.
- [9] A. Krizhevsky, I. Sutskever, and G. Hinton, “Imagenet classification with deep convolutional neural networks,” in *NIPS*, 2012, pp. 1106–1114.
- [10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *CVPR*, 2009, pp. 248–255.
- [11] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, “Decaf: A deep convolutional activation feature for generic visual recognition,” *arXiv preprint arXiv:1310.1531*, 2013.
- [12] N. Rasiwasia, D. Mahajan, V. Mahadevan, and G. Aggarwal, “Cluster canonical correlation analysis,” in *AISTATS*, 2014, pp. 823–831.
- [13] S. Kumar and R. Udupa, “Learning hash functions for cross-view similarity search,” in *IJCAI*, vol. 22, no. 1, 2011, p. 1360.
- [14] F. Wu, Z. Yu, Y. Yang, S. Tang, Y. Zhang, and Y. Zhuang, “Sparse multi-modal hashing,” *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 427–439, 2014.
- [15] M. M. Bronstein, A. M. Bronstein, F. Michel, and N. Paragios, “Data fusion through cross-modality metric learning using similarity-sensitive hashing,” in *CVPR*, 2010, pp. 3594–3601.
- [16] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, “Multimodal deep learning,” in *ICML*, 2011, pp. 689–696.
- [17] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, “Deep canonical correlation analysis,” in *ICML*, 2013, pp. 1247–1255.
- [18] A. Frome, G. S. Corrado, J. Shlens, S. Bengio, J. Dean, T. Mikolov *et al.*, “Devise: A deep visual-semantic embedding model,” in *NIPS*, 2013, pp. 2121–2129.
- [19] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, “Nus-wide: a real-world web image database from national university of singapore,” in *CIVR*, 2009, p. 48.
- [20] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (voc) challenge,” *IJCV*, vol. 88, no. 2, pp. 303–338, 2010.
- [21] S. J. Hwang and K. Grauman, “Accounting for the relative importance of objects in image retrieval,” in *BMVC*, 2010, pp. 1–12.
- [22] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [23] A. Oliva and A. Torralba, “Building the gist of a scene: The role of global image features in recognition,” *Progress in brain research*, vol. 155, pp. 23–36, 2006.

# Intelligent Agent and Virtual Game to support education in e-health

Enrica Pesare, Teresa Roselli, Veronica Rossano

Department of Computer Science

University of Bari

Via Orabona, 4

{enrica.pesare, teresa.roselli, veronica.rossano}@uniba.it

**Abstract** — The education in e-health play an important role in particular in context of chronic diseases. The empowerment process aims at enhancing the patient knowledge and skills in order to allow them to be aware of their health status. The paper presents a serious game that is addressed to young children with type I diabetes and uses the coaching strategy. The coach has been designed and developed as an Intelligent Agent using a knowledge base about the diabetes intervenes to suggest the best action to do or to correct wrong behaviours. A first pilot test was conducted to measure the learning effectiveness and usability of the game. The results were promising.

*Virtual game; Intelligent game; Serious game; Coaching strategy.*

## I. INTRODUCTION

The success and effectiveness technology use to support education and training are universally recognized. In the recent years, the use of ICT is spreading as tools to support also therapeutic education. Even the European Community is sponsoring and encouraging this trend, as can be noticed in the ICT calls of Horizon 2020 work programme. Many ICT solutions, in fact, are addressed to the patient empowerment [7, 8, 13, 15, 22, 24]. The concept of empowerment can have different meanings: strengthening, self enhancement, power increase, increased personal responsibility and knowledge; the patient should constantly monitor her/his status and learn to recognize potentially dangerous situations for her/his health.

In this case, the process of empowerment is fundamental: it begins from the first hospitalization when the diagnosis is confirmed and lasts for life. The use of new technologies, in these cases, can also act on the psychological state of the patient and make training more effective. This is especially true when patients are very young, as in the case of type I diabetes mellitus, that occurs in children and it is caused by a congenital lack of insulin production. Because of patients' young age, it is essential that parents learn quickly how to keep the blood sugar balance to ensure them a good health, and for the patients themselves to learn the glycemia self-management, to allow them to enjoy the same activities of their peers in safety. Moreover, the empowerment process should support the acquisition of knowledge and skills that allow the young people to play an active role in the management of their disease.

With these premises, the authors, in collaboration with a team of practitioners of the "Metabolic Diseases and Medical Genetics Unit" of "Giovanni XXIII" Pediatric Hospital in Bari and with the APGD (Association of Apulian Young Children with Diabetes) have developed different technological solutions to support learning and training processes [7, 8, 22, 24]. The paper presents a serious game addressed to young people. The coaching strategy has been used in order to supply some suggestion to the player during the game. The coach has been designed and developed as a simple reflex agent: based on condition-action rules in the knowledge base, it suggests the best action to do or to correct wrong behaviours. Teaching strategies

In the specific context of patient empowerment the motivation and the user engagement are two key factors to assure a high level of learning effectiveness [19]. In addition, if the final users are young people, as in case of type I diabetes, it is important to use teaching strategies able to maximize their outcomes without overwhelming them.

## A. Coaching

Coaching represents one of the teaching strategies most popular in learning/teaching settings. It was defined by Goldstein & Carr [11], who used the athletic paradigm to illustrate how the computer could act as a coach in educational settings. They used this concept, borrowed from sports and athletics, in order to transfer serious intellectual skills. The coaching strategy is less invasive for the learners than the Socratic method, where in order to stimulate critical thinking the teacher uses a form of inquiry and debate based on asking and answering questions. In coaching, the observation of the teacher (coach) is non-invasive; for example, in soccer the coach stays outside the football field, where he observes and suggests actions that can contribute to improve the player's performance. Since the coaching is a non-invasive strategy, it represents one of the best strategies for the patient's empowerment. In this case, in fact, the intervention of a trainer is essential to help the trainee in the self-empowerment process because patients should become aware of their abilities and use them to improve their health status.

One-to-one tutoring methods such as coaching are very effective. Empirical evidences showed that students, who are

supported by a skilled tutor in a one-to-one setting, performed better than the students who learned in a classroom setting with the same material [2]. Since a one-to-one tutor is expensive, the use of intelligent agents could combine personal characteristics of the subject with the need of a training process organization to ensure high learning performances.

### B. Game based learning and Serious Game

The origins of game-based learning theory can be found in the Vygotskij's works, then, thanks to the spreading of digital games, it has gained high attention. In particular, Prensky [25] states that the digital game-based learning is a way to create new opportunities and effective learning tools for students who grew up with computers and video-game since childhood. As a matter of fact, the game-based learning promotes at the same time fun and engagement and aims at making the achievement of educational goals easier, more student-centered, and therefore more effective.

But those opportunities are not only reserved to the digital natives; in fact, as stated by Papastergiou [23], games are powerful learning environments to promote both knowledge and skills acquisition, thanks to their capability to support several pedagogical aspects [20, 21, 23].

For these reasons, in the latest years the Serious Games (SGs) are becoming one of the widely tools used in learning

contexts. De Freitas 2006 [10] defines SGs as "...applications using the characteristics of video and computer games to create engaging and immersive learning experiences for delivering specified learning goals, outcomes and experiences." It is clear that, this kind of immersive and experiential tools for learning will be useful not only in traditional learning contexts, where the knowledge has to be acquired, but also in training contexts where the previous knowledge has to be applied. Furthermore, SGs make step forward in comparison with serious playing [27]: by combining the realism of simulation with the clearness of the game instructions and goals. They provide a controlled environment to practice and experience several options: the game environment prevents users from anxiety and fear of consequences. The effectiveness of SGs has been proven in several domains, such as science [3, 4, 14, 18] and humanities [5, 16, 26, 28], but also in the e-health field where the need of joining knowledge and practical skills is more pushing, as reported in [1, 6, 7, 8, 9, 13, 15, 22, 24].

These are the premises of our research work where the coaching strategy has been integrated in a serious game to create a virtual game aimed at supporting young patients in the self-empowerment process in the context of type 1 diabetes. In particular, the patients are asked to acquire the knowledge and the skills to help them to self-manage their clinical conditions and especially to prevent serious hypoglycaemic events and reducing psychological burdens.

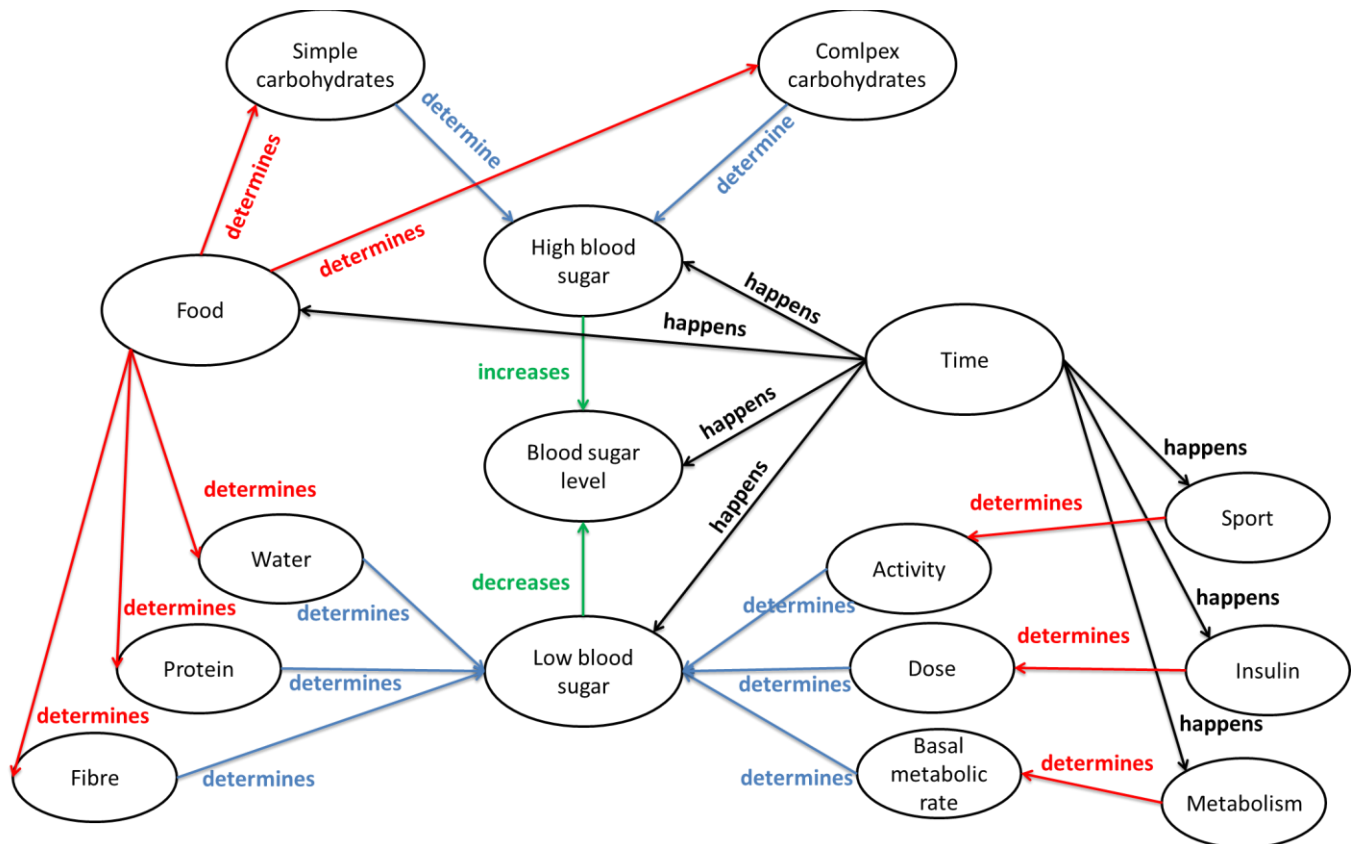


Figure 1. Knowledge model representation



## II. VIRTUAL COACH

Virtual Coach is a virtual game that combines pedagogical strategies, such as coaching and game-based learning [25], with the technological issues of serious game [27] and intelligent agent in order to educate the young users to adopt an adequate lifestyle and to control the glycemic balance.

The system is a game simulation that allows the user to take care of a virtual character, supported by the presence of a doctor, as virtual coach. The coach intervenes to give suggestions and to correct wrong attitudes. The use of the game is a specific communication strategy to make the training process more pleasant for the young and to increase user engagement and motivation, essential elements to make the learning effective [23].

### A. The game mission

The mission of the game is to help a virtual patient, named Mario, to make the right choices during everyday life. The ultimate aim is to train the patient to handle the normal activities, such as breakfast, homework, play, and so on.

### B. The game organization

In the game the user faces with two types of days: a school day and a non-working day, such as Sunday or public holiday. During the day different activities are scheduled, according to the type of the day, and different scenarios have been designed. Each scenario represents the most important moments of the patient's life. For diabetic children, for example, it is important to understand what behaviour to adopt during meals and physical activities in order to control the blood sugar levels.

### C. Learning objectives

In order to reach the specific learning goal, different basic issues have been defined with the practitioners: what kind and how many physical activities could be done; what foods and how many of them could be eaten; and when and how many insulin should be injected.

### D. Knowledge topics

As said before, in the virtual game there is a coach, represented by a doctor avatar, which intervenes to suggest the right behaviour or to correct wrong user's actions. To allow the coaching activities, a model of the knowledge about the type I diabetes has been defined (Fig. 1). In particular, the knowledge base has to be aware of the blood glucose metabolism, in other words it is necessary that the virtual game knows why and when the blood glucose increases and why and when it decreases. For example, the blood glucose increases faster when you eat fruits rather than when you eat a slice of bread. In this context, the representation of the *Time* in the knowledge base has a key role.

## III. THE GAME ARCHITECTURE

The game has been designed as a client-server web application (Fig. 2). The different components are:

- the *knowledge base* (server side) developed using the Prolog language;
- the *inference engine* (server side) composed of two different modules, one to calculate the blood sugar level and one for the fluent calculus;
- the *user interface* (client side) to allow the user to interact with the system.

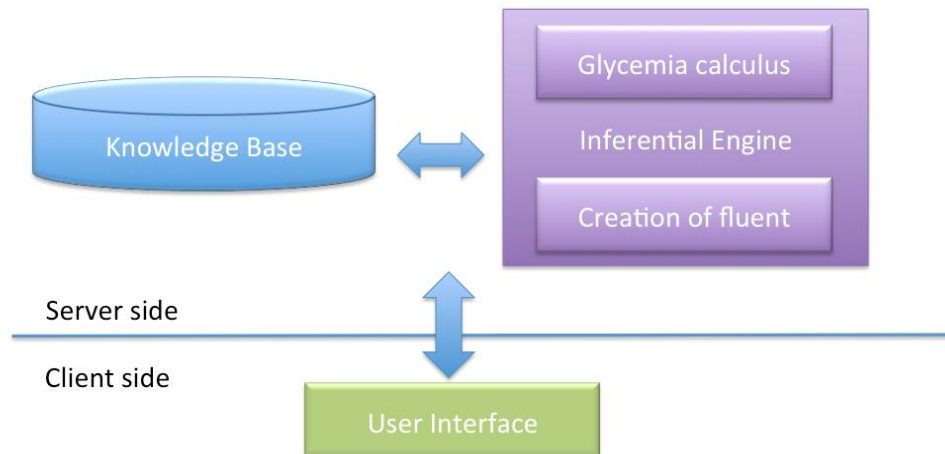


Figure 2. Game architecture

### A. The knowledge base

As said before in the blood sugar metabolism it is essential understand how the level of glucose can vary during the day. Different kind of foods and macronutrients (simple or complex carbohydrates, proteins, fiber, water) as well as different kind of activities (basal metabolism, sport activities or insulin injection) have different impact on the blood sugar levels according to the time (Fig.1).

For this reason the fluent calculus has been used to formalize the increasing and decreasing of glycaemia:

- *fapporto\_glicemico* ( $T_i$ ,  $T_f$ ,  $Varmin$ ): represents the increasing of glycaemia in a specific slot of time;
- *fdiminuzione\_glicemica* ( $T_i$ ,  $T_f$ ,  $Varmin$ ): represents the decreasing of the glycaemia in a specific slot of time.

$T_i$  is the start time of the fluent, that is the time when an event occurs;  $T_f$  is the end time of the fluent, calculated according to the knowledge base rules;  $Varmin$  is the variation of the glycaemia per minute, that depends on the kind of event.

For example, the event that improves the glycaemia is the food. Let us suppose that the player gives to Mario some cookies at 4.00 PM, the fluent *fapporto\_glicemico* is started. Both  $T_f$  and  $Varmin$  depend on the quantity of carbohydrates and the type, in the specific case the glycaemia will increase of 0,5 mg/dl per gr and the fluent will last 120 minutes, since the cookies contain complex carbohydrates. This allows the blood sugar level in Mario to be calculated. In the same way, events, such as insulin, sport, metabolism, will be responsible for the reduction of glycaemia. In those cases the *fdiminuzione\_glicemica* fluent will be activated.

All the events are the results of the possible player interaction within the virtual game.

### B. The inference engine

The inference engine has two main components:

- Creation of the fluents module
- Glycaemia calculus module

The first module is addressed to the definition of the different type of events that can occur in the virtual game. It defines which type of fluent should be activated. Moreover, the module has to define the parameters of the fluent based on of the different type of events. When the fluent is created, an assertion is written in the knowledge base in order to calculate the glycaemia.

The second module allows the glycaemia to be calculated in a specific slot of time and allows the transition to the following scenarios with a specific blood sugar value. In order to calculate it, the fluent *glicemia* ( $Glicemia$ ,  $Tglic$ ) was defined. The parameters represent the value of glycaemia ( $Glicemia$ ) at a specific time ( $Tglic$ ) in the game. When the player starts the game a random value of glycaemia is defined by the system.

In order to well understand how the knowledge base is used, let us suppose that when the day starts at 7.00 AM the random value of Mario glycaemia is 100,  $Tglic$  is 0, the fluent *glicemia*(100,0) starts. At 7.00 the scenario in the game is usually the breakfast, the player gives to Mario 50gr of complex carbohydrates (cookies). Since the action of eating cookies means that the blood sugar level increases, the fluent *fapporto\_glicemico* is started. Using a mathematical function that emulates the blood sugar metabolism, both  $T_f$  and  $Varmin$  are calculated, their values, as said before, depend on the quantity and type of food eaten. In this specific case,  $T_f$  will assume the value of 120 minutes and  $Varmin$  is 1.3875. When the player changes the scenario, let us suppose that are the 8.00 AM, since 60 minutes are passed in the game, the new glycaemia value is calculated by the fluent *glicemia* taking into account that the *fapporto\_glicemico* fluent is active until 9.00 AM.

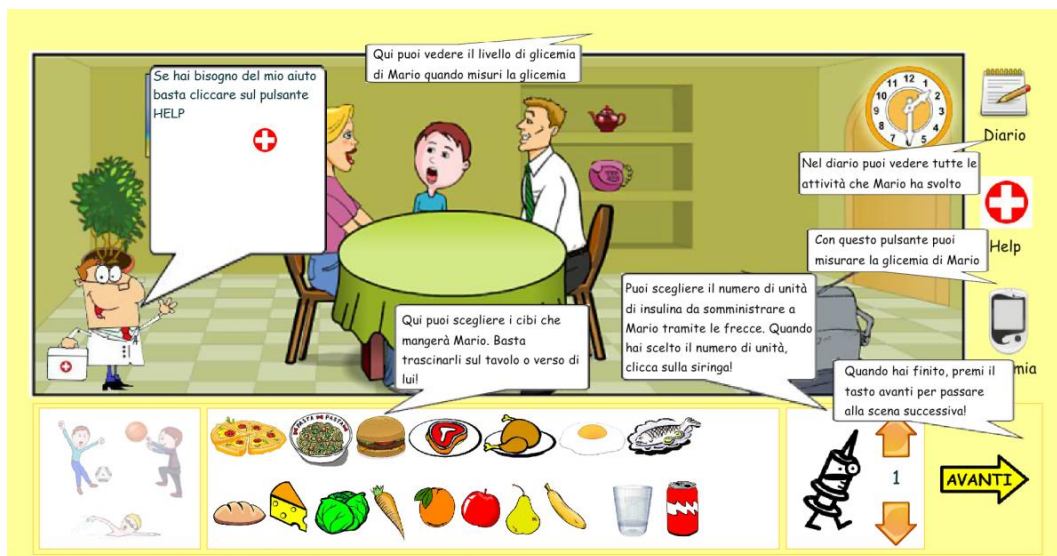


Figure 3. The introduction to the interaction



Figure 4. The virtual coach gives some advice before Mario plays with his friends

### C. The user interface

The virtual game starts with the doctor avatar (the virtual coach) that introduces the game and the characters (Fig. 3).

The doctor avatar is always visible in the left side of the main scenario, where Mario is interacting with other characters (Fig. 4). The main elements in the right side of the screen are:

- The clock used to allow the player to monitor the time flow. It is useful to understand how the blood sugar level change during the day.
- The diary where the system writes all the actions performed by the player, such as eating food, playing sport, and measuring blood glucose (Fig 5). The diary is useful to allow the player to think over the relationship between her/his actions and the glycaemia metabolism.

- The help button to ask the coach help. He can provide some useful tips during the game.
- The glucometer to visualize a sliding bar with the glycaemia values.

The main elements at the bottom of the screen are the available actions. Different options are proposed according to the current scenario (Fig. 3 and Fig. 4). From the left side there are:

- different types of sport to be played;
- the foods that can be offered to Mario;
- the insulin doses to be supplied;
- the button to go ahead in the game.



Figure 5. The glycaemia diary

#### IV. PILOT STUDY

In order to measure the learning effectiveness of the virtual game a hypothesis test was conducted. The defined hypothesis where:

- $H_0$  (null hypothesis) there is no statistically significant difference between the user knowledge about blood sugar levels management before the use of the virtual game and after it;
- $H_a$  (alternative hypothesis) there is a statistically significant difference between pre and post user knowledge about blood sugar levels management.

##### A. Participants

The pilot study involved twenty children, aged from 8 to 12, patients of the “Metabolic Diseases and Medical Genetics Unit” of “Giovanni XXIII” Pediatric Hospital in Bari under the supervision of Dr. Elvira Piccinno and Dr. Elda Frezza, medical experts.

##### B. Pre-experimental design

The pilot study was conducted using the “one group pre-test post-test study” [12].

The benefit of this design is the use of a pretest to determine baseline scores and a post test in order to measure the outcome of the treatment. In this case, the treatment is the use of the virtual game and the expected outcomes is the improving of the knowledge about type I diabetes and blood sugar levels management. Both pretest and posttest were accurately defined with the practitioners.

##### C. Procedure

The pilot study consisted in two sessions, each one lasted an hour.

During the first session all subjects underwent an individual pre-test to assess their prior knowledge about diabetes, and they were asked to use the virtual game for at least half an hour in order to become familiar with the game.

The second day (second session), they were asked to use the virtual game, and then to answer the post-test in order to evaluate the knowledge gain.

##### D. Results

In order to evaluate the pilot study results the Wilcoxon signed-rank test was used [29]. The Wilcoxon is a non-parametric test used as an alternative to the Student t-test when the population cannot be assumed to have a normal distribution [17]. After the computation of statistical significance, if the p-value is higher that the predefined level (fixed at 0.05) the  $H_0$  (null hypothesis) could be accepted, otherwise  $H_a$  (alternative hypothesis) has to be accepted.

In Table 1 all the results obtained by the subject are reported. The first observation is that in almost all cases, except two users, there was an improvement of knowledge between pretest and posttest.

Moreover, p-value obtained is lower then 0.05, then the alternative hypothesis has to be accepted: there is a significant difference between the pretest and the posttest.

Table 1. Pretest and Posttest results

User	Pretest	Posttest	Difference
1	7	9,50	2,5
2	6	8,50	2,5
3	7,5	8,00	0,5
4	3,5	5,00	1,5
5	6	8,50	2,5
6	6	8,00	2,0
7	5	7,50	2,5
8	0	2,50	2,5
9	6	8,00	2,0
10	8	9,50	1,5
11	9	9,00	0,0
12	9	9,50	0,5
13	6	7,50	1,5
14	6,5	8,25	1,8
15	3,75	6,00	2,3
16	9,5	9,50	0,0
17	5	7,50	2,5
18	7	9,00	2,0
19	6	7,50	1,5
20	7,5	9,75	2,3
<b>Mean</b>	<b>6,14</b>	<b>7,57</b>	<b>1,4</b>

In addition, a deeper analysis of the collected data allowed us to find that the virtual game had a great impact on the ability to recognize the different kind of carbohydrate. In fact, the mean knowledge gain was higher on this topic than the knowledge gain on hypoglicemia and hyperglycaemia (Fig. 6). The main reason of this difference is that the virtual game is mainly devoted to acquire competences about type I diabetes, such as to distinguish complex carbohydrates from simple carbohydrates and how much of them could be eaten to balance the blood sugar level.

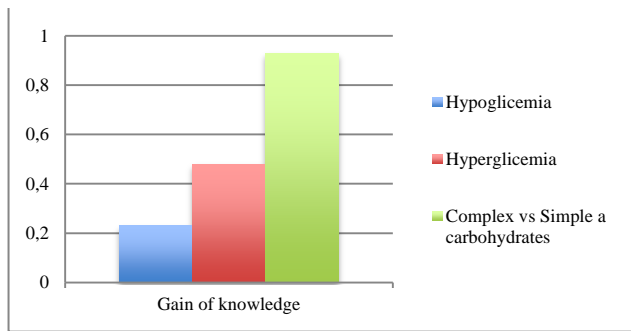


Figure 6. Mean knowledge gain for each topic

#### E. Usability results

Usability test or inspection are the more common methods to understand whether users feel comfortable with the software. In educational context, and in particular, when the multimedia is addressed to patient empowerment it is important to know how much the user likes using the system. For this reason after the post-test the users were required to give some feedback on different aspects of the game.

The learnability was judged positive from the users.

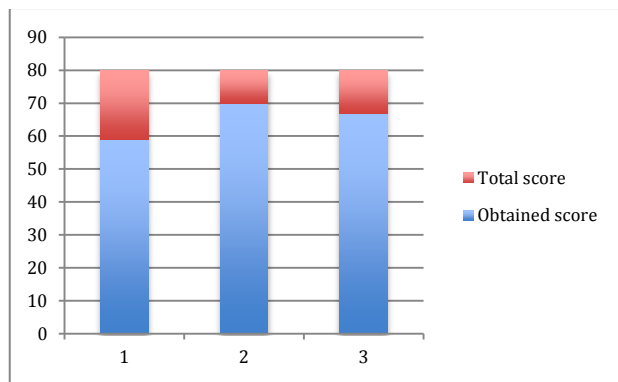


Figure 7. Learnability results

Also the efficiency that measures how fast the user can accomplish a task, was appreciate by the users.

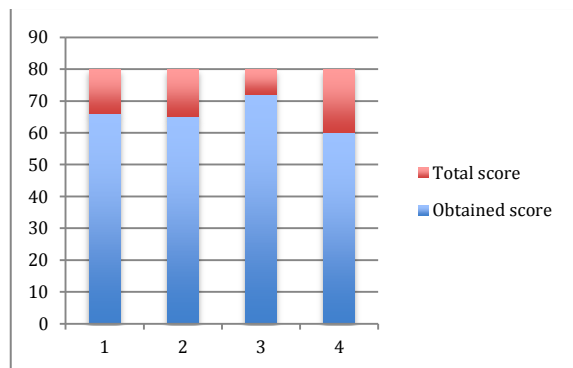


Figure 8. Efficiency results

Finally, the users like using the system, obtained score 66 out 80, and they claim that would suggest the use of the game to their friends. In particular, some of the subjects stated that

the virtual game could be more effective for those children who are new to the type I diabetes.

#### V. CONCLUSIONS AND FUTURE WORKS

To guarantee a good quality of life with chronic diseases it is basic to have a proper lifestyle. Thus, it is important that an empowerment process starts from the first hospitalization or when patient become aware of the disease. This is particularly true in type I diabetes where the patients are young children. In order to do everything, like their friends, they need to learn how the glycaemia metabolism works and how they can avoid hypoglycaemia and hyperglycaemia events. For this purpose our research, conducted with a team of practitioners of "Metabolic Diseases and Medical Genetics Unit" of "Giovanni XXIII" Pediatric Hospital in Bari, led us to design and build different solutions aimed to support the educational processes of young patients. The use of new technologies and the game approach is the key to keep the user motivated and engaged in the learning process. The proposed virtual game uses the coaching strategy in order to allow basic concepts and skill about diabetes to be acquired. The coach, who is represented in the game as a doctor avatar, intervenes when wrong actions are taken by the player. To do this a knowledge base and an inference engine were defined in order to simulate the glycaemia metabolism. A first pilot study has demonstrated both the learning effectiveness and the appreciation of young patients and their parents.

In the next future will be necessary to improve the number of the available choices for the player and to conduct an experiment with a larger sample.

#### ACKNOWLEDGMENT

We would like to thank Antonio Simmaco and Giuseppe Grosso who developed the virtual game during their Master Thesis in Informatics, the APGD (Association of Apulian Young Children with Diabetes) and all the young children that participate to the study.

#### REFERENCES

- [1] Berni, F., Corriero, N., Pesare, E., Rossano, V., Roselli, T.: A Knowledge Management Service for e-health. In: ICERI2013 Proceedings, pp. 488-493. ISBN: 978-84-616-3847-5. (2013)
- [2] Bloom, B. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, 4-16.
- [3] Bos, N., & Shami, N. S. (2006). Adapting a face-to-face role-playing simulation for online play. *Educational Technology Research and Development*, 54(5), 493-521.
- [4] Cagiltay, N. E. (2007). Teaching software engineering by means of computer-game development: Challenges and opportunities. *British Journal of Educational Technology*, 38(3), 405-415.
- [5] 29Chou, C., & Tsai, M. J. (2007). Gender differences in Taiwan high school students' computer game playing. *Computers in Human Behavior*, 23(1), 812-824.
- [6] Corriero, N., Di Bitonto, P., Roselli, T., Rossano, V., Pesare, E.: Simulations of clinical cases for learning in e-health. In: *International Journal of Information and Education Technology*, International



- Conference on Information and Education Technology (ICIET) Melbourne, 2-3 January, 2014.
- [7] Di Bitonto, P., Roselli, T., Rossano, V., Frezza, E., Piccinno, E. (2012). An educational game to learn type 1 diabetes management. In: The 18th International Conference on Distributed Multimedia Systems. Miami Beach, USA, 9-11 August, 2012, p. 139-143, SKOKIE, ILLINOIS: KSI Press, ISBN: 1-891706-32-2
  - [8] Di Bitonto, P., Rossano, V., Roselli, T., Piccinno, E., Ortolani, F., Frezza, E., & Tummolo, A. (2014). Gamification to train young diabetic to manage the insulin metabolism. INTED2014 Proceedings, 3586-3592.
  - [9] Di Bitonto, P., Di Tria, F., Roselli, T., Rossano, V., Berni, F.: Distance Education and Social Learning in e-Health. *International Journal of Information and Education Technology*, vol. 4, no. 1, pp. 71-75 (2014)
  - [10] De Freitas, S. (2006). Learning in immersive worlds. London: Joint Information Systems Committee.
  - [11] Goldstein, I. and Carr, B. (1977) The computer as coach: As athletic paradigm for intellectual education. In Proceedings of the 1977 annual conference (ACM '77). ACM, New York, NY, USA, 227-233.
  - [12] Heffner, C. L., Research Methods, AllPsych Editor available at <http://allpsych.com/researchmethods/preexperimentaldesign/#.VVvFfJNX9E4>
  - [13] Kato, P. M., S. W. Cole, et al. (2008). A Video Game Improves Behavioral outcomes in Adolescents and Young Adults With Cancer: A Randomized Trial. *Pediatrics* 122(2): 305- 317.
  - [14] Kiili, K. (2007). Foundation for problem-based gaming. *British Journal of Educational Technology*, 38(3), 394-404.
  - [15] Lieberman, Debra A., Interactive video games for health promotion: Effects on knowledge, self-efficacy, social support, and health. In: Street, Richard L., Jr., Gold, William R., Manning, Timothy R. (Eds), *Health promotion and interactive technology: Theoretical applications and future directions*, (pp. 103-120). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, (1997).
  - [16] López, J. M. C., & Cáceres, M. J. M. (2010). Virtual games in social science education. *Computers & Education*, 55(3), 1336-1345.
  - [17] Lowry, Richard. Concepts & Applications of Inferential Statistics <http://vassarstats.net/textbook/>.
  - [18] Mayo, M. J. (2009). Video games: a route to large-scale STEM education?. *Science*, 323 (5910), 79-82.
  - [19] McCarley, P. (2009). Patient empowerment and motivational interviewing: engaging patients to self-manage their own care. *Nephrology nursing journal*, 36 (4), 409-413.
  - [20] McFarlane, A., Sparrowhawk, A., & Heald, Y. (2002). Report on the educational use of games. Teachers evaluating educational multimedia.
  - [21] Oblinger, D. G. (2004). The next generation of educational engagement. *Journal of interactive media in education*, 2004(1).
  - [22] Ortolani, F. M. Vendemiale, A. Tummolo, P. Di Bitonto, V. Rossano, Roselli T. E. Piccinno (2014). Learning by doing approach: use of multimedia applications in type 1 diabetic children. In: ESPE DUBLIN - 53rd Annual Meeting. european society for pediatric endocrinology, DUBLIN, 18-20 SEPTEMBER
  - [23] Papastergiou M., Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. In: *Computers & Education*, 52 (1), 2009, pp. 1-12.
  - [24] Piccinno, E., Vendemiale, M., Tummolo, A., Ortolani, F., Frezza, E., Torelli, C., Di Bitonto, P., Rossano, V., Roselli, T. New technologies for promoting hypoglycaemia self- management in type 1 diabetic children. In: 9th Joint Meeting of Paediatric Endocrinology. Milan, 19-22 September, 2013.
  - [25] Prensky, M. Digital Game-Based Learning. McGraw-Hill, New York, 2001.
  - [26] Ravenscroft, A. (2007). Promoting thinking and conceptual change with digital dialogue games. *Journal of Computer Assisted Learning*, 23(6), 453-465.
  - [27] Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research & Development*, 44(2), 43-58.
  - [28] Tüzün, H., Yılmaz-Soylu, M., Karakuş, T., İnal, Y., & Kızılkaya, G. (2009). The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers & Education*, 52(1), 68-77.
  - [29] Wilcoxon, Frank (Dec 1945). "Individual comparisons by ranking methods" *Biometrics Bulletin*, Vol. 1, No. 6. (Dec., 1945), pp. 80-83.



# Surveillance System with SIS Controller for Incident Handling using a Situation-based Recommendations Handbook

Erland Jungert<sup>1</sup> and S.-K. Chang<sup>2</sup>

<sup>1</sup>CustodIT  
S-582 26 Linköping, Sweden  
erland @jungert.net

<sup>2</sup>University of Pittsburgh  
Pittsburgh, PA, USA  
chang@pitt.cs.edu

**Abstract**---Protection of critical infrastructures involves handling of incidents that may range from serious to quite harmless events. Such systems require means for surveillance that involves a type of sensor system that may identify entities that behave in an unusual way. However, this is not sufficient as means for determination of entities that seemingly are behaving in a normal way but whose activities somehow relate to the first category, must be determined. Means for the support of the operators must also be available by the surveillance system. In this work, an approach to a surveillance system with a Slow Intelligence systems controller for incident handling using a situation-based recommendations Handbook, is proposed and discussed.

**Keywords:** security systems, surveillance systems, Slow Intelligence, recommendations handbook

## I. INTRODUCTION

Critical infrastructures are to an increasing degree becoming the target for intruders with the intentions to either destroy such facilities or to take them over. For this reason, systems designed for the surveillance of such critical infrastructures have become necessary. An integrated part of such surveillance systems is the sensor system that may include multiple sensors of varying types, such as video cameras, IR-cameras and radars. The sensor systems are continuously collecting large amount of data that are analyzed by the surveillance system and made available to the operators. Collected data may be represented in various information structures. The generation of such extremely large data volumes will eventually lead to the determination of overwhelmingly large information quantities that must be handled and interpreted by the surveillance system to support its operators. Adequate handling of the incoming information by the operators is more or less impossible unless the information is organized and presented in a suitable way. The support for this should not only be carried out by the operational picture presentation system, the operator will also need recommendations on how to act under various circumstances as the situation that need to be handled may be quite complex and of unexpected nature. Important aspects here are, for

example, relationships existing between entities within and around the facility and in which context the entities are acting and whether they could be determined as direct or indirect intruders. Operator support from the surveillance system is of great importance when dealing with serious incidents such as attacks from threatening intruders or antagonists. The approach taken here is to solve the problem based on an approach to Slow Intelligence [1] and the use of a situation-based recommendation handbook for crisis management [2] and the protection of critical infrastructures.

The main objective of the work discussed in this work concerns an approach to incident handling based upon Slow Intelligence systems (SIS) controller. Secondary to this, some details of a recommendations handbook to support the operator of the surveillance system will also be discussed together with the required information structures of the surveillance system.

This paper is organized as follows. The architecture of the surveillance system is described in section II together with its process steps. In section III the Slow Intelligence system controller is introduced together with various computation cycles of the Slow Intelligence system controller. The situation-based recommendation handbook is discussed in section IV. Section V describes a short scenario and section VI gives an overview of the identified information structures, section VII presents related works and conclusions of the work are discussed in section VIII.

## II. SYSTEM ARCHITECTURE

The architecture of the surveillance system can be seen in Fig. 1 and it is made up by three basic modules i.e. the sensor system, the Slow Intelligence system controller (SISC) and the visual operations control (VOC). The sensor system and its sensors will not be dealt with further in this work but it is expected to be able to detect and identify entities of all relevant types and on command track

entities entering, residing inside or leaving the facility. This will require a system with a large number of sensor types where the capacity for collection, analyses and storage of these data will be necessary. Such sensor systems will be feasible in the near future as the technology for the development of such systems already exists. The SISC module supported by the sensor system will have the capability to identify entities engaged in hostile activities during the entire period of an incident. Another capability of great importance to the surveillance system is to allow for early detection of hostile entities. To be able to carry out all its requirements SISC also needs to have direct access to all information collected, generated or pre-stored in the databases of the surveillance system. The VOC contains, besides the operational picture, the command and control unit, the recommended actions viewer module and the situation-based recommendation handbook. Attached to the SISC module are three different databases, i.e. the Surveillance information database, the Terrain database and the Normal states database, which stores context information and other descriptions related to the normal state of the facility.

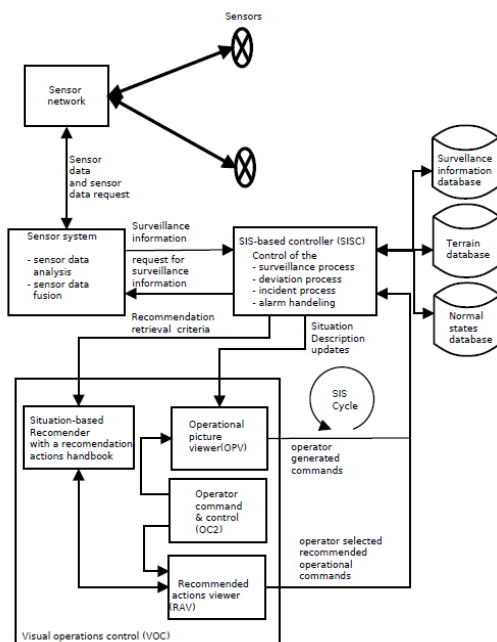


Figure 1. The surveillance system and its main modules, including the sensor system, SISC and the Visual operations control.

The system can be in one of three event states, i.e. *Normal*, *Deviation* and *Incident*. Besides, there is also an alarm state with three different states, i.e. *Inactive*, *Unverified* and *Verified*. In the *Normal* state the surveillance system collects information that is used to determine whether the behavior of all appearing entities are normal. Once some type of deviation from the normal occurs, the Event status will be switched to

*Deviation* and in some obvious cases directly to *Incident*. In the *Deviation* state one or more entities may be subject to further observations and an alarm is released, which will correspond to an unverified alarm that immediately must be verified either automatically by some sensors or manually by a guard. The alarm will be determined either as *false* or *true*. When false the system is switched back to *Normal* and when true is set to *Incident*. In the *Incident* state, the operator, the rest of the staff and other involved must be concerned with bringing the incident to its end. Once the incident is brought to its end the system is switched back to the *Normal* and the alarm state is set to *Inactive*. The various steps in these processes are in short described below.

#### *Normal state*

At all times: for all events collect and store Surveillance information;

Analyze all events; If deviation, set Event status to *Deviation* and Alarm status to *Unverified*; release unverified alarm and proceed to *Deviation* state; if alarm verified set Event status to *Incident*, Alarm status to *Verified*, release verified alarm and proceed to *Incident* state ;

If Event status equal to *Normal*: Proceed.

#### *Deviation state*

At all times: Collect and store Surveillance information for all events; analyze all events; Handle unverified alarm: if alarm *false* set Event status to *Normal*, Alarm status to *Inactive* and switch to *Normal* state; if alarm *true* set Event status to *Incident*, Alarm status to *Verified*, release verified alarm and switch to *Incident* state;

#### *Incident state*

At all times: Collect and store Surveillance information for all entities; analyze all events, determine relations of all other entities; Handle incident: if incident over set Event status to *Normal*, Alarm status to *Inactive* and switch to *Normal* state; If incident still going on, proceed.

### III. THE SIS CONTROLLER

SISC can be seen as an *information hub* of the Surveillance system as almost all information in the system flows through this module. Besides by coordinating the information flow of the system SISC also determines entity to entity and entity to context relations. A further aspect of concern is to determine which of the entities are related to any entity with the event state equal to *Incident*. To determine this is the main task of the Slow Intelligence process. In short, this means that any entity with an identified relationship to an entity determined as an intruder and part of the incident

will be classified and handled as an intruder independently of whether this eventually is the case.

The Slow Intelligence process can be seen as a procedure where first all entities, including also their entity relations, that could be considered antagonist candidates are determined, i.e. whether they could be determined as antagonists that could be participating in an incident. In a second step some of the identified entities are eliminated as they are considered unlikely participants of the incident. This procedure can be described as follows in terms of a pseudo high level programming language.

*Repeat for all entities in Facility when Alarm has been released or when Event-status eql Incident*  
*(if Alarm is released for Entity E<sub>j</sub> then*  
*(if Alarm unverified then (verify Alarm for*  
*Entity E<sub>j</sub> /\* E.g. send out guard \*/*  
*if Alarm verified for Entity E<sub>j</sub> then*  
*if Event-status eql Normal then Exit)*  
*if Event-status eql Incident then*  
*set Incident-entity-set to E<sub>j</sub>*  
*Repeat until no extension of entities in*  
*Incident-entity-set then Exit*  
*/\* enumeration step\*/*  
*for all entities in Incident-entity-set*  
*Determine all relations between all*  
*entities present in Facility*  
*/\* elimination step \*/*  
*Eliminate all entities with harmless relation(s)*  
*to entities in Incident-entity-set*  
*Extend Incident-entity-set with entities with*  
*relevant relations to entities in*  
*Incident-entity-set*  
*Extend Incident entity relations with*  
*the relations between relations in*  
*New-entities and the relation in*  
*Incident-entity-set*  
*set Event-status to Incident to all new*  
*entities in Incident-entity-set))*

#### A. Computation Cycles of SIS Controller

The computational cycles of the SIS Controller described in terms of Slow Intelligence operators and operational cycles as can be seen below. The comments in the description are inserted to make the description more readable.

**Variables:** Incident-entity-set (IES), New-entities (NE), Incident-entity-relations (IER), Entities-in-Facility (EIF), Observed-entity (OE)

**Initial state:** IES = {}, NE = {}, IER = {}, EIF = {OE<sub>1</sub>..., OE<sub>k</sub>..., OE<sub>n</sub>}, OE = {}

*/\*Description in terms of the abstract machine for Slow Intelligence\*/*

Cycle0: [guard 0,2] P<sub>0</sub> +adapt= P<sub>1</sub> =prop+ P<sub>2</sub>

*/\*Information is received from the sensor system and*

*propagated to the operational picture system which is updated; the information concerns entities entering and exiting the facility. Cycle0 proceeds receiving and propagating information as long as Event status equal to "Normal" and when it becomes "Deviation" the cycle terminates, if the Event status is "Incident" control is switched to Cycle 2\*/*

Cycle1: [guard1,0] P<sub>3</sub> +adapt= P<sub>4</sub>

*/\*This cycle is entered for Event status equal to "Deviation". SISC request the operator to verify this status. If the response is "Normal" or "Deviation" control is switched to Cycle0 otherwise if the status is "Incident" the cycle terminates/*

Cycle2: [guard 2,2] P<sub>1</sub> -enum< P<sub>5</sub> >elim- P<sub>6</sub>  
*/\*This cycle proceeds recursively when an incident has occurred to identify all entities that are part of the incident, i.e. entities related to the entity that caused the incident. It terminates when all entities associated with the incident are determined\*/*

Cycle3: P<sub>7</sub> =prop+ P<sub>8</sub>

*/\*An incident is going on and entities participating in the incident must be tracked; this request is propagated to the sensor system\*/*

Cycle4: P<sub>9</sub> +adapt= P<sub>10</sub> =prop+ P<sub>11</sub>

*/\*Incident related information (basically tracks and entity information) from the sensor system is further propagated to the operational picture system/*

Cycle5: [guard5,9] P<sub>12</sub> +adapt= P<sub>13</sub>

*/\*This cycle request information, from the sensor system concerning entities entering the facility; if no new entity is available control is transferred to Cycle9 otherwise the cycle terminates\*/*

Cycle6: [guard 6,9] P<sub>14</sub> -enum< P<sub>15</sub> >elim- P<sub>16</sub>

*/\*In this cycle P<sub>14</sub> is enumerated with respect to the latest acquired entities in Cycle 5 and all relations between the entities in the Incident-entity-set are determined and then the relations are eliminated with respect to whether they are non-incident related. If the relation set is empty control is transferred to Cycle9 otherwise the Incident-entity-set and the Incident-entity-relations are extended with the determined entity and its relations respectively and then the cycle terminates.\*/*

Cycle7: P<sub>17</sub> =prop+ P<sub>18</sub>

*/\*Incident information is propagated to the operational picture that becomes updated\*/*

Cycle8: P<sub>19</sub> =prop+ P<sub>20</sub>

*/\*The sensor system is notified that a new incident related entity should be tracked\*/*

Cycle9: P<sub>21</sub> +adapt= P<sub>22</sub>

*/\*SISC should be notified by the operator when incident related entities have been taken care of by the security staff, that is the entity has been captivated or has disappeared from the facility\*/*

Cycle10: [guard10,5]  $P_{22} > elim - P_{23}$   
*/\*captivated or disappeared incident related entities are eliminated from the Incident-entity-set and from the Incident-entity-relations. If Incident-entity-set becomes empty the cycle terminates, which means that the incident is over, otherwise control is switched to Cycle5 and the incident proceeds in a new loop\*/*  
 Cycle11: [guard11, 0]  $P_{18} = prop + P_{19}$   
*/\*The incident has been terminated (Event status is set to "Normal", Alarm status to "Inactive" ) and all relevant information is saved in the databases for later use and control is transferred to the normal status loop, i.e. Cycle0\*/*

### B. Basic Computation Cycles

The algorithmic computational descriptions of the SIS controller described in terms of Slow Intelligence operators and operational cycles can in a simplified overview version be described as follows:

- Cycle0: handles the situation when the Event status is *Normal* and Alarm status is *Inactive*. During the execution of this cycle these two variables may be switched to 1) *Deviation* and *Unverified* or 2) *Incident* and *Verified*. In case 1) SISC is transferred to Cycle1 and in 2) to Cycle2.
- Cycle1: handles the situation when Event status has been set to *Deviation* and Alarm status to *Unverified*. When the alarm has been verified the state may change to 1) *Normal* and *Inactive* and control switched back to Cycle0 or 2) *Incident* and *Verified* then control to Cycle2.
- Cycle2 – Cycle10: these cycles control the maintenance of the incident.
- Cycle11: the incident has terminated and condition is set to *Normal* and *Inactive* and control is switched to Cycle0.

### C. Details of computation Cycle 2

A more detailed description of the process in Cycle 2 above can be described as:

Cycle2: variables: **Incident-entity-set (IES)**, **New-entities (NE)**, **Incident-entity-relations (IER)**, **Entities-in-Facility (EIF)**, **Observed-entity (OE)**

$P_1$ :  $IES = \{OE_j\}$   
 $EIF = \{OE_{1,...}, \dots, OE_j, \dots, OE_k, \dots, OE_{k+t}, \dots, OE_n\}$   
 $NE = \{\}$   
 $IER = \{\}$   
 $P_5$ :  $IES = \{OE_j\}$   
 $IER = \{OE_j \text{ rel } OE_{1,...}, OE_j \text{ rel } OE_k, \dots, OE_j \text{ rel } OE_{k+b}, \dots, OE_j \text{ rel } OE_n\}$   
 $P_6$ :  $IES = \{OE_j, \dots, OE_k, \dots, OE_{k+t}, \dots\}$   
 $IER = \{OE_j \text{ rel } OE_k, \dots, OE_j \text{ rel } OE_{k+b}, \dots\}$

The formal specification of the computation cycles provides a concise way to describe the SIS controller and also offers the possibility to mathematically derive certain properties such as the termination or non-termination of the SIS controller.

## IV. SITUATION-BASED RECOMMENDATIONS HANDBOOK

In this section an overview of the situation-based Recommendation Handbook will be presented.

### A. Organization of the Handbook

The Situation-based Recommendation Handbook will be engaged in a number of activities, i.e. to respond to the information received from SISC by looking up recommendations in the Handbook aimed at supporting the operator in the occurred situation. The received information can be of any of the three following alternative types:

- An entity including its Event type, properties, context and Event status set to *Deviation* and Alarm status to *Unverified*.
- An entity including its Event type, properties, context and Event status set to *Incident* and Alarm status to *Verified*.
- Entities, their direct or indirect relations, their Event type and Event status set to *Incident* and Alarm status to *Verified*.

A consequence of the above input to the Handbook is a set of recommended actions to be carried out by the operator. The Handbook will include instructions of type *call the police* or *send out a guard to patrol the location*. However, those types of instructions are basically determined by local authorities at the facilities and will for this reason not be dealt with further here. Of importance to the work is the organization of the Handbook and the means to access its entries. The Handbook is basically split into two parts where the first is concerned with unverified alarms and deviating behavior while the second part is entirely focusing on incidents. The two parts are called the *Deviation part* and the *Incident part*. The search criteria of the two parts can simply be expressed as follows:

- The search criteria of the *Deviation part* are the Entity, Event types and Behavior of entity.
- The search criteria of the *Incident part* are the Entity, Event types and Behavior of entity where

- a single entity is in focus resulting in just a single look-up,
- multiple entities are in focus resulting in one look up for each entity.

The first cases are rather trivial. The last concerns multiple entities that may relate to, e.g. a *meeting* which involves at least two entities that may or may not be of different types but nevertheless will need one look-up for each entity so that the operator can handle them both separately and together.

Besides, recommendations to make, for example, phone calls to specific persons or organizations the Handbook must also give recommendations that concern the context of the operational situation. If an antagonist is walking through a forest around the facility a sent out guard cannot follow that person by car. If the antagonist is expected to carry weapons other precautions must be recommended. The list of special recommendations may be made quite long and cannot be completed here but must be seen as a task determined by the security staff at each specific facility.

#### B. Events corresponding to possible incidents

In Table 1 a series of possible events that may cause an incident are described; the number of incidents in the list is not complete and include just a few examples for illustration purposes. If an entity during an incident is acting accordingly its event status will be set to *Deviation* or *Incident*.

Table 1. Event types, their possible relations to other entities.

Event type	relation to other entity or object
Approaching a fence	An entity is approaching and acting unnatural close to a fence
Approaching a prohibited area	An entity is approaching a prohibited area or have been standing there for some time
Object picked up inside facility	Object picked up by an entity inside the facility
Object picked up outside facility	Object picked up by entity outside the facility
Object thrown over fence of facility	Object thrown over fence from outside or inside the facility by entity

The processing of these events may occur either in Cycles 1, 2 or 6 in SIS controller. Besides occurring during an incident as in 2 or 6, each of these events may either be the cause of an incident or a deviation, i.e. when any of these events occur in Cycle 1. For Cycles 2 and 6 verification of the alarm is not necessary because the incident is already

going on, i.e. the alarm state is already set to *Verified*. Of importance to the events described is that they correspond to capabilities of surveillance applications, see e.g. [3], and carried out in conjunction with the sensor system and SISC.

#### V. A SCENARIO

The scenario given here can in short be described as follows:

*A person is observed walking against a fence of a facility. At the fence the person stops and throws a package over the fence and walks away. After a while a second person comes on the inside of the facility and picks up the package and walks away against a prohibited area.*

This short scenario includes, in sequence a number of events and for each one of them SISC generates instances of the Status and Context information relations. This information is then sent to the Handbook that looks up the corresponding recommendations.

##### Event 1

Status information: <person#3,14.45, *Deviation*, *Unverified*, approaching fence >

Context information: <person#3, fence, forest >

Hand book search criteria: Person, *Deviation*, *Behavior of entity*

Recommendations: *send out guard to verify alarm; instruct guard to inform on what is going on; operator should follow track of person approaching fence in operational picture.*

##### Event 2

Status information: <person#3, 14.54, *Incident*, *Verified*, object thrown over fence >

Context information: <person#3, fence, road>

Handbook search criteria: Person, *Incident*, *Behavior of entity*

Recommendations: *call police, set facility in safe mode.*

##### Event 3

Status information: <person#46, 15.23, *Incident*, *Verified*, object picked up inside >

Context information: <person#46, fence, road>

Handbook search criteria: Person, *Incident*, *Behavior of entity*

Recommendations: *the operator is instructed to follow track of person in operational picture; a pair of guards should be sent out to observe the person; Guards instructs to report on what is going on.*

Report from guard: *person outside fence is taken care of.*

##### Event 4

Status information: <person#46, 15.38, *Incident*, *Verified*, approaching prohibited area >

Context information: <person#46, fence, road>  
 Handbook search criteria: Person, *Incident*,  
*Behavior of entity*  
 Recommendations: *the operator is instructed to follow track of person in operational picture.*

#### Event 5

Status information: <person#46, 15.54, *Incident*,  
*Verified*, at prohibited area>  
 Context information: <person#46, prohibited area,  
 road>  
 Handbook search criteria: Person, *Incident*,  
*Behavior of entity*  
 Recommendations: *instruct guards to arrest person at prohibited area.*  
 Report from guard: *person at prohibited area is arrested.*

As soon as the last report has come in the incident has been brought to its end and the state of the surveillance system will be set to *Normal* and *Inactive*. However, this last step may need to include some further activities as the antagonists may have carried out activities whose effects have not yet been discovered and that consequently may cause problems later on. It is consequently necessary to inspect the facility for such perhaps dangerous threats even after the incident has been terminated. This is an activity that must be carried out by the staff of the facility.

## VI. INFORMATION DESCRIPTION

In this section information structures that need to be used by the surveillance system are described.

### A. Event related information

The purpose of event related information is to serve two capabilities of the surveillance system, i.e. to

- find relevant entries to the Handbook
- keep the operational picture updated at a current state.

That is, more or less the same information used to look up entries in the Handbook is also used to keep the operational picture updated. Consequently, Event dependent information used in these two activities corresponds to information acquired by means of the sensor system and in conjunction with the Slow Intelligence system controller that continuously is surveilling the facility and analyzing the acquired information; an activity that can be seen as the screening of a number of ongoing events that may correspond to various types of incidents. Incidents may consequently range from quite harmless behavior of different entities and up to really serious events carried out by terrorists. Furthermore, event related information can also be determined in part from historic events where the underlying data are captured

over long periods in time essentially to allow statistical determination of what is a deviation from normal. Examples of such information could be tracks of observed objects that compared to historic data shows that the entity deviate from what can be considered normal. Other information that may be needed to improve the knowledge of an observed object and its general behavior could be the determination of relations to other entities. This information may be used to find new and relevant entries in the Handbook.

The information that needs to be collected by the sensor system and eventually stored belongs to classes that can be expressed as follows including also possible but not entirely complete value sets.

#### Facility entities

- Facility subarea: {outside facility boundary, inside facility boundary, facility boundary, facility airspace, restricted area, ...}
- Physical installation of facility: {fence, building, road, walk way, gateway, check point ...}
- Facility terrain type: {forest, hill, park, plain, water front, urban area ...}
- Sensor system: {sensor type, sensor location ...}
- Manually controlled sensor: {sensor type, sensor location...}

#### Event entity

- Event location: {coordinates /2D or 3D/}
- Event subarea: {perimeter, outside facility boundary, inside facility boundary, airspace, prohibited area ...}
- Physical installation at event: {fence, building, road, walk way, gateway, check point ...}
- Event terrain type: {forest, hill, park, plain, water front, urban area}

#### Event condition

- State of event: {day, night}
- Time of event: {time}
- Weather condition: {rain, snow, fog, clear sky...}

#### Observed entity type

- Entity type: {Person, Car, Truck, Aircraft ...}
- Person: {antagonist, police, fireman, guard ...}

#### Behavior of entity

- Observed behavior: {walking, running, driving, still, climbing, entering, exiting, hiding ...}
- Estimated direction: {N, NE ..... W, NW}
- Estimated Target: {/facility dependent/}
- Event installation type: {fence, building, roof, road, walk way, gateway, check point ...}
- Event terrain type: {hill, park, plain, water front, urban area, prohibited area... }



### Event situation

- Event status: {*Normal*, *Deviation*, *Incident*}
- Alarm status: {*Inactive*, *Unverified*, *Verified* (false alarm, system failure)}

The above set of classes can be seen as an ontology, see Fig. 2.

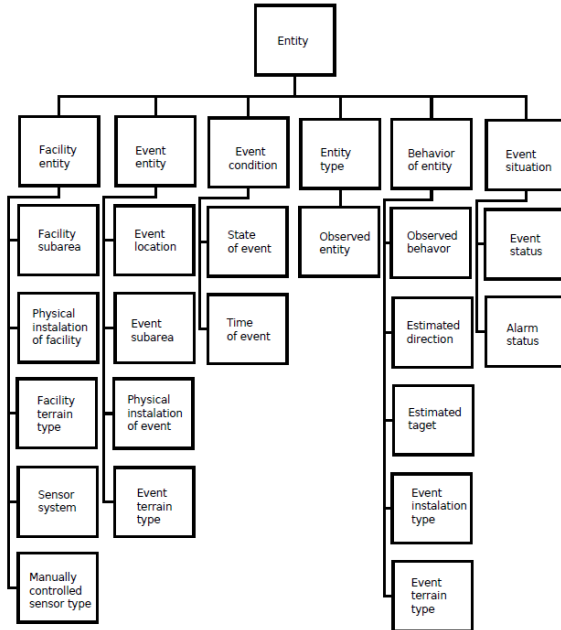


Figure 2. The ontology of Entities.

### B. Acquired and complementary information

Acquired information in this context means information captured by the sensor system that relates to detected entities and their status and properties. That is, information originating from sensor data that have been analyzed, often fused and eventually transferred into SISC for further analyses. Besides this, there is also a need for stationary information such as geographical information, which here is called complementary information and relates in most cases to the context of the facility and where the detected entities may reside from time to time. To thoroughly describe the context in which an entity resides and its status at a specific time during an ongoing event both acquired and complementary information is required. This can be described in terms of two information sets (relations) where some of the information appears in both sets basically for identification. The relations are approximately described as follows:

- Status information: <Entity, Event location, Time of event, Event status, Event type, Alarm status, Observed object type, Observed behavior>

- Context information: <Entity, Event location, Time of event, Event status, Alarm status, Subarea, Event installation, Terrain type>

The first relation, Status information, describes the status of an event at a specific point in time (Time of event) and its location; further information concerns the actual event type that is normal, a deviation or an incident. Data are gathered periodically for every observed entity within the area covered by the sensor system. Altogether, for every observed entity a track of every observed entity can be determined although it seems more economical to just track entities with the event status *Deviation* or *Incident*. Events that are classified as normal will be analyzed for determination of whether the general behavior of the entity is normal; all this will obviously require analyses of massively large data sets. For normal events no alarm is activated, that is the alarm status is set to *Inactive* and for these events the Handbook is not consulted. However, for all observations the Operational picture is updated to give the operator a presentation of the current situation at the facility. If the event is classified as a deviation by the system the Handbook must be consulted, and the operator may be instructed to verify the alarm to determine whether the event is to be classified as an incident. In case of an incident the operator must bring the incident to an end by means of the recommendations from the Handbook and the views of the operational picture.

### C. The event situation

The event situation concerns the status both of observed events and the current alarm status. The relationship between Event status and Alarm status can be described as in Table 2. To be observed here is also that events like threats must be considered as just deviations which can be seen as a situation with unverified alarms. This means that this alarm must be verified before an incident is at hand.

Table 2. Possible event situation

Event status /Alarm status	Inactive	Unverified	Verified
Normal	N	Error	Error
Deviation	Error	D	I
Incident	Error	Error	D

Whenever an error occurs the operator must deal with a system failure; in other words it is a serious event that immediately must be handled by special domain experts or technicians but it is not an

incident or deviation in the usual sense.

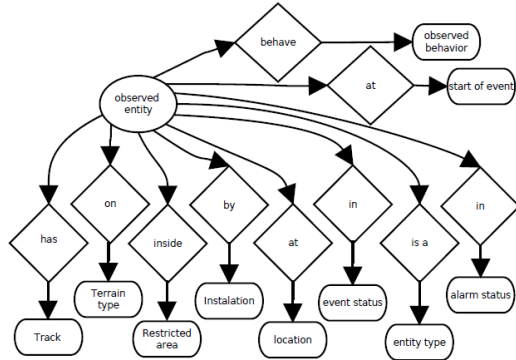


Figure 3. Properties and context of an observed entity.

#### D. Entity relations

All observed entities have relations to the environment in which they are acting but they also relate to the current event and leave tracks that especially during deviations and incidents must be acquired. As a consequence, the state description of any observed object may look like in Fig. 3 where both properties and contextual descriptive terms are included.

Of importance are not only the properties and the context of observed entities but also the relations between entities involved in deviations and incidents. Such relations can be direct as in the first relation in Fig. 4, that is in the simple relation *Entity-i meets Entity-j*. In this relation no incident may be caused unless *Entity-i* is already classified as an intruder. In the second case an incident is determined because of the indirect relation as in the second example in the figure where *Entity-i throws object-k* and *Entity-j picks up Object-k*, where an entity throws an object, which will cause an incident, and then the indirect relation is established when the second entity picks up the thrown object. Which causes the second and indirect relation to be determined as *Entity-j* is considered part of the incident as well. Obviously, such entities, although they are classified as normal, appear to have relations to entities determined as incident related because throwing an object will automatically cause a verified alarm leading to an incident. Consequently, entities that appear to have any kind of relation to an entity involved in an incident must be seen as part of the incident i.e. their Event status must be switched from *Normal* to *Incident* and the crisis management staff must start acting accordingly. This means that the Handbook tells the operator to focus on the new incident related entity as well, which also is indicated in the operational picture that will show its status as *Incident*.

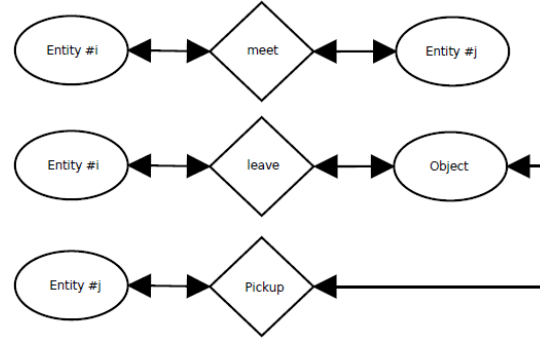


Figure 4. Examples of direct (above) and indirect entity relations (below).

## VII. RELATED WORKS

Relations to other work in this context concerns several different aspects. Basically, they are concerned with Slow Intelligence, various approaches to recommendations systems but also to surveillance systems for protection of critical infrastructures. Of importance to note is also that this is the first attempt on the design of security and surveillance systems using Slow Intelligence for determination of entities involved in incidents at critical infrastructure facilities.

The Slow Intelligence approach was first proposed by Shi-Kuo Chang [1]. The visual specification of component-based Slow Intelligence Systems is described in [4]. This work introduces the visual description of super-components by Petri nets or other UML diagrams. It provides the foundation of the present work. Component-based Slow Intelligence Systems has been applied to many areas, including social influence analysis, topic and trend detection, high dimensional feature selection, image analysis, swimming activity recognition, and most recently pet care systems and energy control systems. In [5] the notion of an abstract machine for computation cycle was introduced. Our current approach is based upon it.

Related surveillance system for various approaches can be found in Goodall [6] where 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 applied to emergency response. Hansson et al. [8] demonstrates the intentions to determine the general context for security systems as a foundation for user and system requirements. In [9] is a project called RESCUE discussed. In this work are a number of aspects that relates to this work discussed e.g. the focus on needs to assess situations for improved awareness. Pozzobon et al. [10] discuss primarily user requirements of surveillance systems

with special emphasis on security in ports. A further relationship to this work concerns basically user requirements needed for determination of user oriented capabilities which plays an important role in systems development. However, this is more in focus to the work discussed in [3].

Content based recommendation systems are generally based on descriptions of various items that may be of interest to a user with a particular profile see e.g. Pazzani and Billsus [11]. However, here the situation differs in that the recommendations can be seen as the result of observed events detected by the sensor system. Consequently, such recommendation systems will be event driven and the user has no other choice than to react to the events and given recommendations. Example of an application of this latter approach is proposed by Laliwala [12] in which an event-driven service-oriented agricultural recommendation system is proposed. Another similar example is given by Kim et al. [13].

## VIII. DISCUSSION AND CONCLUSIONS

In this work a surveillance system for protection of critical infrastructures is proposed. The main focus concerns capabilities to identify entities involved in abnormal behavior that eventually will cause alarms and turn the status of the system status into *Incident*. Further aspects that have been subject to attention here are also determination of incidental events and relations between entities involved in such incidents. The approach taken has been to carry out the determination of such information based on an approach to Slow Intelligence; the outcome is a SIS controller that will use various search criteria to a situation-based recommendation Handbook and the maintenance of an operational picture system. Complementary to this, the required information structures are also described and discussed. Finally, to demonstrate the capabilities discussed a simple scenario is carried out.

The approach taken in this work contributes to Slow Intelligence research in the sense the scenarios describe realistic computation cycles for the SIS controller. Hence, further research must focus on analysis of the computation cycles to determine the properties of SIS controllers such as termination conditions, existence of endless loops and so on.

## REFERENCES

1. S.-K. Chang, "A General Framework for Slow Intelligence Systems", International Journal of Software Engineering and Knowledge Engineering, Volume 20, Number 1, February 2010, 1-16.

2. S.-K. Chang, E. Jungert, A Self-Organizing Approach to Mission Initialization and Control in Emergency Management, Proceedings of the International Conference on Distributed Multimedia Systems, San Francisco, September 6-8, 2007, pp 51-56.

3. E. Jungert, N. Hallberg & N. Wadströmer, A system design for surveillance systems protecting critical infrastructures, Journal of Visual Languages and Computing, December 2014, Vol 25(6), pp 650-657.
4. S.-K. Chang, Y. Wang and Y. 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.
5. S. K. Chang, W. H. Chen, B. Kao, L. Kuang, and Y. Z. Wang, "The design of pet care systems based upon Slow Intelligence principles," Int'l Journal of Software Engineering and Knowledge Engineering, 2014.
6. 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. M. 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, Aug. 2011, pp. 64-69.
9. S. Mehrotra, C. Butts, D. Kalashnikov, N. Venkatasubramanian, R. Rao, G. Chockalingam, R. Eguchi, B. Adams and C. Huyck, "Project Rescue: Challenges in Responding to the Unexpected", in *SPIE*, Vol. 5304, Jan 2004, pp. 179-192.
10. A. Pozzobon, G. Sciutto, and V. Recagno, "Security in ports: The user requirements for surveillance systems", In The of a Wireless Sensor Network Application from End-User Requirements", Proc. Of the 2010 6<sup>th</sup> int. Conf. on Mobile Ad hoc and Sensor Networks (MSN) Dec. 20-22, 2010, pp 168-175. Springer Int. Series in Eng. And Comp. Sci., Vol. 488, 1999, pp 18-26.
11. M. J. Pazzani, D. Billsus, Content-Based Recommendation Systems, Ed(s) P. Brusilovsky, A. Kobsa, W. Nejdl, The Adaptive Web - Methods and Strategies of Web Personalization, Springer Verlag, vol. 4321, 2007, pp 325 - 341.
12. Z- Laliwala, "Semantic and Rule Based Event-driven Service-Oriented Agricultural Recommendation System", 26<sup>th</sup> IEEE Int. conf. on Distrib. Comp. Systems Workshops (ICDCS), July 4-7, 2006.
13. J. K. Kim, H. K. Kim, Y. H. Cho, "A user-oriented contents recommendation in peer-to-peer architecture", Expert Systems with Applications, Jan 2008, Vol. 34(1), pp 300-312.

# Graph Databases Lifecycle Methodology and Tool to Support Index/Store Versioning

Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi, Nadia Rauch

DISIT Lab, Dep. of Information Engineering, University of Florence, Italy

<http://www.disit.dinfo.unifi.it>, {pierfrancesco.bellini, ivan.bruno, paolo.nesi, nadia.rauch}@unifi.it

**Abstract**— Graph databases are taking place in many different applications: smart city, smart cloud, smart education, etc. In most cases, the applications imply the creation of ontologies and the integration of a large set of knowledge to build a knowledge base as an RDF KB store, with ontologies, static data, historical data and real time data. Most of the RDF stores are endowed of inferential engines that materialize some knowledge as triples during indexing or querying. In these cases, deleting concepts may imply the removal and change of many triples, especially if the triples are those modeling the ontological part of the knowledge base, or are referred by many other concepts. For these solutions, the graph database versioning feature is not provided at level of the RDF stores tool, and it is quite complex and time consuming to be addressed as black box approach. In most cases the indexing is a time consuming process, and the rebuilding of the KB may imply manually edited long scripts that are error prone. Therefore, in order to solve these kinds of problems, this paper proposes a lifecycle methodology and a tool supporting versioning of indexes for RDF KB store. The solution proposed has been developed on the basis of a number of knowledge oriented projects as Sii-Mobility (smart city), RESOLUTE (smart city risk assessment), ICARO (smart cloud). Results are reported in terms of time saving and reliability.

**Keywords** — *RDF Knowledge base versioning, graph stores versioning, RDF store management, knowledge base life cycle.*

## I. INTRODUCTION

Large graph databases are getting a strong push in their diffusion for setting up new kind of big data services for smart cities, digital libraries, competence modeling, health care, smart education, etc. This fact is mainly due to their capability in modeling knowledge and thus on creating Knowledge-Based, KB, systems [Grosan and Abraham, 2011]. Graph databases may be implemented as RDF stores (Resource Description Framework) [Klyne and Carrol, 2004], to create interactive services in which reasoning and deductions can be elaborated including inference engines on top of the store. An RDF store is grounded on the concept of triple that puts in relationship two entities. For example, *Carl knows Paolo*, consisting of a subject, a predicate and an object, which in turn are represented with URI. Predicates, as “*knows*”, may be specified by using vocabulary that defines relations. A vocabulary defines the common characteristics of things belonging to classes and their relations. A vocabulary, also called ontology, is defined by using RDFS (RDF Schema, RDF Vocabulary Description Language) or the OWL extension (Ontology Web Language). Recently RDF store have been also addressed as noSQL stores for big data [Bellini

et al., 2013a]. A large set of ontologies and related data sets are now accessible, see for example the large number of LOD (linked open data) accessible and related each other via URI [Berners-Lee, 2006], [Bizer et al., 2011]. RDF stores may be made accessible via an entry point to pose semantic queries formalized for example in SPARQL [Hartig et al., 2009] (SPARQL Protocol and RDF Query Language, recursive definition). Non trivial RDF stores based solutions are typically produced by exploiting **multiple ontologies**, loading data triples and testing/validating the obtained results. This means that they are built by using some ontology building methodology [Noy and McGuinness, 2001], [Lopez, 1999], integrated with a knowledge base development life cycles.

The RDF store may grow over time adding new triples, and may have the capacity to learn if endowed of an **inferential reasoner/engine**, i.e., producing new knowledge that are new triples. Thus, the inferential engine associated with the RDF store materializes new triples during reasoning (for example at the time of indexing or querying). These facts are the main motivations to low performances in indexing, and critical performances in deleting triples of RDF stores as graph databases since they are involved in removing the materialized triples in the store. These features impact on store performances, and thus, in literature, many benchmarks for the evaluation of RDF stores are present. Some of them use real data as from dbPedia, UniProt, WordNet, other use synthetically generated data as LUBM [Guo et al., 2005] (university domain), BSBM [Bizer et al., 2009] (e-commerce domain), SP2Bench [Schmidt et al., 2009] (library domain). More recently, in Linked Data Benchmark Council LDLC EU project, two new benchmarks have been developed: one based on Social Network [Erling et al., 2015] and the second on Semantic Publishing. While LUBM and SP2Bench benchmarks are based on real data, and evaluate only the queries performed after the data load. BSBM and LDLC benchmarks evaluate a mix of insert/update/delete/query workloads. When RDF stores are used as a support for a KB, some of the changes in the RDF store can be destructive for the graph model, such as changes in the triples modeling the ontology on which millions of instances are related. In order to keep the performance acceptable, the RDF store has to be rebuilt from scratch or from some partial version to save time in releasing the new version. Thus, the lifecycle may present multiple cycles in which the RDF store is built incrementally via progressive refinements mediating among: (i) reusing ontological models, (ii) increasing the capability of making

DOI reference number: 10.18293/DMS2015-016



deductions and reasoning on the knowledge base, (iii) maintaining acceptable query performance and rendering performances, (iv) simplifying the design of the front-end services, (v) satisfying the arrival of additional data and models and/or corrections, etc. A commonly agreed lifecycle model to build KBs is not available yet and many researchers have tried to embed KB development steps into some conventional software lifecycle models [Batarseh, Gonzalez, 2013]. In general, development of KB systems is a multistep process and proceeds iteratively, using an evolutionary prototyping strategy. A number of lifecycle models have been proposed specifically for KB systems [Milette 2012].

In the lifecycle model, a change in the ontology may generate the review and regeneration of a wide amount of RDF triples. The problem of ontology versioning as addressed in [Klein et al., 2002], [Noy and Musen, 2004] can be easily applied if the ontology is not used as a basis for creating a large RDF KB store. Moreover, in [Volkel et al., 2005], the versioning of RDF KB has been addressed similarly to the CVS solutions by using commands as: *commit*, *update*, *branch*, *merge*, and *diff*. The differences are computed at semantic level on files of triples. Thus, [Zegins et al., 2007] presented a solution for versioning RDF models assuming the possibility of estimating the delta between two RDF models by performing a set of adds and deletes to a model to transform it to the other. At database level, the key performance aspects of an RDF KB store version management are the storage space and the time to create a new version [Tzitzikas et al., 2008]. Therefore, possible approaches could be to store: (a) each version as an independent triples store [Klein et al., 2002], [Noy and Musen, 2004], [Volkel et al., 2005]; (b) the deltas in terms of triples between two consecutive versions and implementing a computationally expensive and time consuming chain of processes to maintain and apply deltas [Zegins et al., 2007].

**In this paper**, a versioning system for RDF KB proposes to integrate both (a) and (b) solutions. It manages versioning of RDF stores by: (i) keeping trace of the set of triples to build each version, (ii) storing each version and related set of triples, (iii) providing an automated tool for keeping trace of triple files, descriptions for store building and stores, (iv) allowing the versioning of the RDF KB store, (v) reducing the critical manual error prone operations. This approach allows to make indexing versioning for RDF stores that materialize triples at indexing (as OWLIM [http://www.ontotext.com/]) or at querying (as Virtuoso [http://virtuoso.openlinksw.com/]) without influencing the RDF store reconstruction. The resulting time for returning to a previous version and to reconstruction of a new one is satisfactory and viable, since some of the RDF stores are very time consuming in indexing, while other do not allow the deletion of triples. Therefore, the paper presents an RDF KB methodology life-cycle suitable for big data graph databases, and a versioning tool for RDF KB stores that has been developed and tested for SESAME OWLIM and Virtuoso; and thus it can be simply extended to other RDF stores. The solutions have been developed for Km4City project [Bellini et al., 2013b], and adopted for other RDF KB oriented projects as Sii-Mobility Smart City national

project and RESOLUTE H2020 European Commission Project. They are large KB oriented projects in the Smart City, smart cloud, smart railway domains, developed at the DISIT Lab of the University of Florence <http://www.disit.org/6568>.

The paper is organized as follows. Section II presents the RDF Knowledge Base life-cycle model and methodology for development. In Section III, the RDF KB indexing flow and requirements for the RDF Indexing Manager tool are presented. Section IV describes the RDF Index Manager tool, detailing the architecture, and the XML formal model for index descriptors. In Section V, experimental results are reported providing data related to real cases, in terms of time and managed complexity. Conclusions are drawn in Section VI.

## II. A KNOWLEDGE BASE LIFE-CYCLE

Building a RDF KB is a challenging practice that needs a well-defined methodology and lifecycle to keep under control the entire development process. RDF KBs are mainly developed thanks to a cycle approach that allows checking and validating the advances made, and if needed, to make adjustments when a problem is identified. As stated above, the lifecycle proposed in this paper has been derived from the DISIT Lab experience cumulated while developing a number of big data RDF KBs.

The proposed methodology and lifecycle for RDF KB is reported in **Figure 1**. The life-cycle presents 4 vertical pillars and one horizontal block that represents the **RDF Store usage and Maintenance**. The life-cycle spans from the ontology creation to the RDF Store usage on the front-end where also real time data are added.

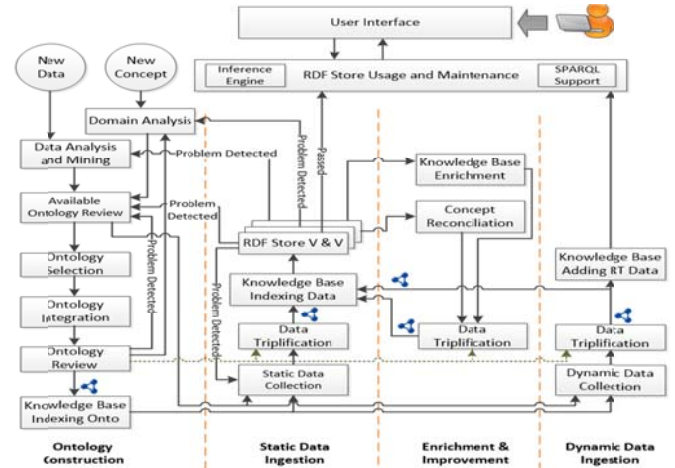


Figure 1. RDF KB Life Cycle Model

The pillars refer to the:

**Ontology construction**, from domain analysis the setup of the RDF Store containing triples of the selected ontologies and possible additional triples to complete the domain model (Knowledge Base O, KB-O). For example, the Km4City ontology reuses: *dcterms* to set of properties and classes for modeling metadata; *foaf* dedicated to relations among people or groups; *schema.org* for a description of people and organizations; *wgs84\_pos* representing latitude and longitude; *GoodRelations* for a description of business entities and their

locations; *OWL-Time* for temporal modeling; *OTN* for transport aspects; *GIS Dictionary*, to represent the spatial component of geographic features; etc. [Bellini et al., 2013b]. The combined ontology is reviewed and possible problems may lead to more or less deep redefinition of the process.

**Static Data Ingestion:** this phase is related to the loading of the data instances of the ontological classes and attributes. Despite their name, static data may change rarely over time, for example, the position of bus stops may be considered static data even if they change seasonally. They come from several sources (static, statistical, historical, etc.), and have to be converted in triples according to the KB-O coming from the previous phase. Then, they are finally indexed by using several sets of triples, maybe thousands. The indexing produces a KB including the former KB-O, plus many data instances; thus, allowing performing the Verification and Validation, V&V, of the RDF KB.

The V&V phase may be the moment in which some problems are detected. They may constrain the expert to: (i) wrong data or incomplete data to need a review of the data mapping to the ontology (restart from the first step of this phase of data collection), (ii) missing ontology aspects and classes, thus leading to the review of the ontology built (returning to Ontology Review), (iii) problems in data collected that may be wrongly mapped to ontology classes (returning to Data Analysis and Mining), (iv) mistake in data mapping that may lead to revise the whole Domain Analysis, and successive steps. If this phase is passed, the RDF Store passes to the phase of RDF Store Usage and Maintenance. Additional static data sets may be added to the KB-O if the ontological model supports them without deletion, otherwise a review is needed.

**Enrichment and Improvement, E&I:** this phase allows solving problems that may be present in the produced RDF Store. E&I processes may take advantage from the access to the partially integrated KB, exploiting for examples solutions of Link Discovering [Ngomo, 2011], [Isele, Bizer, 2013], and/or making tuned semantic queries. Additional processes of E&I may be added to the RDF Store if the model supports them without performing some delete otherwise a model review is needed.

**Dynamic Data Ingestion:** when the RDF store is in use, collected data from real time information (for example, bus delay with respect the arrival time, weather forecast, likes on the user profile, status of sensors, status of cloud processes, etc.) can be added to the RDF Store and saved into the repository of the historical triples. Additional dynamic data sources may be added to the RDF Store if the model supports them without performing some delete otherwise a model review is needed. Please note that dynamic data should not need to validate and verify process since the data to be added in real time are new instances of data already mapped and integrated as historical data.

#### A. Data & Domain Analysis and Ontology Construction

Brief descriptions of major interesting blocks pertaining to the proposed life cycle and methodology (see Figure 1) are now provided.

**Data Analysis and Mining:** Each data set (static or real-time) to be addressed in the RDF KB is analyzed and checked to assess if the information related to each single data field is well described in terms of type, range, and context. The data collected is analyzed to understand the concepts in terms of their structure, relationships and information in domain.

**Domain Analysis:** this step is executed in parallel or in alternative to the above data analysis steps. In this phase, the concepts of the domain addressed by the application are studied to understand concepts, terminology, their relationships, and the general rules that are related to them. Several methodologies are accessible to help the analysts in identifying concept from the literature review of the domain – as well as thumb rules: substantive are classes, verbs are relationship, details are attributes, etc.

**Available Ontology Review:** This phase is very important. Once the major aspects of the domain have been identified. The phase consists of studying other related ontologies at the state of the art to see if they can model the whole identified domain and data concepts or at least a part of them. The realistic solution is to start from one or a set of available ontologies and complete the expected model with some specific classes and relationships. This task, it could be performed every time new static and/or dynamic data kind have to be addressed, or for addressing identified problems. When this ontology review is performed starting from an active domain ontology (in the RDF KB), it may happen that the expert may discover that no changes are needed at ontological level (e.g., a new class is not needed since the concept for hosting data is already in place), thus resulting in a direct jump to the phases of static or dynamic data management.

**Ontology Selection:** on the basis of the actions previously performed on concepts and data against ontologies, it is possible to make a selection of the most suitable ontologies to be taken as seeding concepts. The process of selection has to take into account also the licensing aspects, which impose some constraints. For example, some of licenses of the ontologies do not allow being tuned/modified. If the study has not led to any results, it is always possible to write a specific domain ontology.

**Ontology Integration:** as a result of the previous steps the main ontologies have been identified and thus they have to be integrated/glued with each other. In addition, the missing concepts have to be formalized by completing the fitting of the KB with the domain analysis performed. **Ontology Review:** Once the ontology was created/modified, a first revision took place even without the massive loading of instance. Thanks to tools like *Protégé* [<http://protege.stanford.edu/>], which allow to apply a reasoner to the ontology in order to verify that knowledge is modeled as desired. A number of metrics and criteria may be also applied to verify if the ontology has been developed with common criteria. E.g., [Noy and McGuinness, 2001], [Gómez-Pérez, 2004], [Rector et al., 2004].

**Knowledge Base Indexing Onto, KBIO:** the task in which the RDF Store index containing the selected ontologies, vocabularies, and custom defined concepts are integrated as



triples. They may be some files and some tens of classes. This process usually starts from an empty RDF store and takes a few seconds since the ontologies are comprised of a small number of triples and the RDF is empty; differently from what happens when millions of triples of data sets are indexed and they lead to many materialized triples.

#### B. From Ontology to KB via Data Ingestion, major tasks

**Static Data Collection:** on the basis of the created domain ontology, the analysed data (addressed in task Data Analysis and Mining) have to be processed. Static data are typically obtained from open data, statistical data, private data that do not change over time so rapidly. The process of static data ingestion may be performed by means of parallel and distributed architectures executing processes as ETL (Extract, Transform and Load), Java, microgrid, harvesting, crawling, etc. It may include: file access, REST/WS calls, data mapping, quality improvement, e.g., [Bellini et al., 2013b]. **Data Triplification:** A task in which the data (static, dynamic) are mapped to triples on the basis of the domain ontology model.

**Knowledge Base Enrichment:** task focused on enriching the RDF KB Store by adding links to external LOD. For example, referring from the street title to VIP to its dbPedia definition (from Avenida Winston Churchill, to its page on dbPedia: [http://live.dbpedia.org/page/Winston\\_Churchill](http://live.dbpedia.org/page/Winston_Churchill)). For example for the Km4city KB a tool has been created, that allows to identify famous names inside the KB and search for the same name on dbPedia, to finally create triple *cite/isCitedBy* thanks to the *CITO Ontology* [<http://purl.org/spar/cito>].

**Concept Reconciliation:** task related to solve the lack of coherence among indexed entities referring to the same concept but coming from different data sets. This process is a critical step during the KB realization and helps to create new knowledge and new connections between data that would otherwise remain unconnected. For example, different services located at the same street number, several profile aspects of the same person, different representations of the same part of the brain. This task typically produces a number of triples solving the problem of missing links possible. Triples includes relationships of *owl:sameAs*.

**Dynamic Data Collection:** Dynamic data are subject to a lighter ingestion process with respect to static data. In fact, they are picked up and immediately mapped into RDF triples, in order to speed up as much as possible the process that allows making them available to users adding them to the RDF store (*Knowledge Base Adding RT Data*). At the same time, Real Time triples are stored as Historical Dynamic Data for successive construction of versioned data stored.

**Knowledge Base Indexing Data, KBID:** This task takes in charge a high number of triples coming from different data sets:

- **Static data:** for example one or more file containing a set of triples for each single data set;
- **Historical Dynamic data:** several files and triples for each real time data collection channel. For example, the collected weather forecasts of the past two months, the last 200 measures of traffic flow sensor DG32453165, the data regarding the Cloud Host and VM in the last week;

- **Reconciliated data:** triples connecting concepts and data into the RDF KB;
- **Enrichments data:** triples connecting data entities of the RDF KB to external LOD RDF stores. When the enrichment tasks are performed on real time data, they have to be performed in real time as well. For examples if the enrichment is performed on an Opera Name, or about a VIP person name.

In order to pass from the ontological model to a real RDF KB store, many data sets (static, statistical and historical), should be included / indexed in the RDF KB. Very often, indexing process of large files may take several hours. Often files of triples are linked each other and the order of indexing of these data may become essential. In some cases, the historical data can lead to very huge number of triples, thus compromising / influencing the performance of the whole RDF Store. This implies that the RDF KB has to be periodically polished by removing most of the cumulated historical data. This activity is quite natural for smart city and smart cloud applications. For example in cloud monitoring systems as NAGIOS, data are dense in the close time and sparse in the past.

#### C. RDF Store Verification and Validation, V&V

Once the RDF KB Store containing triples coming from data (static, historic, reconciliation and enrichment) has been produced, it is possible to precede with the validation and verification of the RDF Store vs the ontological definition. Please note that, the RDF store index has to be accessible to perform the following V&V processes via semantic queries and analyzing consistency. They can be automatically performed through a set of validation processes implemented as SILK [Isele, Bizer, 2013] as well as SPARQL processes.

The verification and validation process has the duty to detect inconsistencies and incompleteness: (i) verify if the data indexing has been correctly performed, (ii) detect eventual reconciliations to be performed identifying missing connections, (iii) identify eventual enrichments to be performed, (iv) identify eventual mismatch from data loaded and the ontology (for example counting the triples to be indexed and those indexed in reality), (v) verify if the expected inferences are exploited at the query time, etc. The above mentioned criteria allow identifying different kind of problems that may lead to revise the ontological model, the data ingestion process, etc. etc.

### III. RDF INDEXING FLOW AND REQUIREMENTS

As described in the previous section, there are several reasons for which into the RDF KB life cycle the process may lead to (i) revise the ontology (and thus to revise the data mapping and triplification invalidating the indexing and the materialization of triples); (ii) revise the data ingestion including a new data mapping, quality improvement, reconciliation, enrichment and triplification. As stated in the previous section, the life-cycle model foresees two steps where the Knowledge Base Indexing has to be performed: KBIO, KBID. On the other hand, as pointed out in the introduction, in most of the RDF store models, the versioning is not an internal feature. This is due to

the fact that it cannot be easily performed at level index and stored triples for their complexity in removing them, due to the triples materialization by inference. According to the proposed RDF KB life cycle, the modeling of a chain of connected versions of *indexes/RDF Stores*, with incremental complexity may be very useful to keep under control the evolving index with the aim of saving time by exploiting intermediate versions in generating the RDF Store/index for the successive deployment. For example, in the case of Smart City, the layered versions of the index may include the ontology, static and dynamic data, historical data, etc.

To better describe the process of RDF Index versioning, it is necessary to put in evidence the differences between the “*index*” and “*index descriptor*”. An RDF KB store is in substance an “*index*”, while content can be accessed via URI cited in the triples elements. The index is created by loading the triples into the RDF store, and as a result a binary index is built, maybe materializing additional triples according to the ontological model and the specific RDF store inferential engine adopted. The recipe to create the RDF Store index, that is the collection of atomic files containing triples (including triples of ontologies as well as those related to data sets: static, historical, dynamic), can be called as the “*index descriptor*”, that may be used to generate a script for index generation. The script syntax can be different from an RDF Store to another, since their commands for loading and indexing can be different. This approach implies to have aside each pair “*index*” and “*index descriptor*” also the history of files containing triples with their versions, last update dates, and dependencies from other files. For example, see **Figure 2**, where the reconciliation of triples connecting parking locations (File 1, ver 1.5) with respect to civic numbers depends on the ontology and on the parking area data sets. Thus leading to create a set of triples connected with dashed lines.

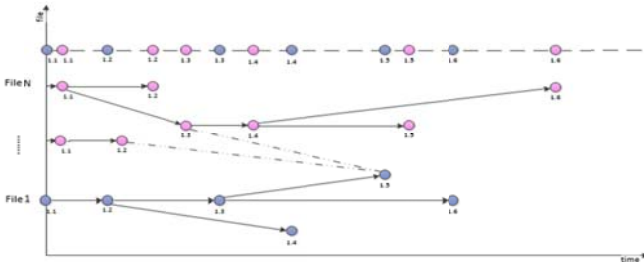


Figure 2. Example of set of file versioning

**Definition.** Let  $F = \{f_1, f_2, f_3, \dots\}$  be the set of triple files that are available for indexing and  $DS = \{ds_1, ds_2, ds_3, \dots\}$  is the set of datasets and ontologies that are available for ingestion. The function  $ds: F \rightarrow DS$  associate the file to the dataset it belongs to, function  $time: F \rightarrow \mathbb{N}$  associate each file with the time when it was created and function  $dep: F \rightarrow \wp(F)$  associate each triple file with a set of files that it depends on (e.g. ontologies),  $\wp(X)$  is the power set of set  $X$ . The  $dep$  function must not introduce a cyclic dependency among files. Moreover, a file should not depend on files created in the future:

$$\forall f \in F, \forall s \in dep(f). time(s) < time(f)$$

**Example**  $DS = \{km4c, otn, roads, services, busses\}$ ,  
 $F = \{kf_1, kf_2, of_1, rf_1, rf_2, sf_1, sf_2, bf_1, bf_2\}$ ,

$ds = \{(kf_1 \rightarrow km4c), (kf_2 \rightarrow km4c), (of_1 \rightarrow otn), (rf_1 \rightarrow roads), (rf_2 \rightarrow roads), (sf_1 \rightarrow services), (sf_2 \rightarrow services), (bf_1 \rightarrow busses), (bf_2 \rightarrow busses)\}$   
 $time = \{(kf_1 \rightarrow 2), (kf_2 \rightarrow 5), (of_1 \rightarrow 1), (rf_1 \rightarrow 3), (rf_2 \rightarrow 8), (sf_1 \rightarrow 2), (sf_2 \rightarrow 8), (bf_1 \rightarrow 3), (bf_2 \rightarrow 8)\}$   
 $dep = \{(kf_1 \rightarrow \{of_1\}), (kf_2 \rightarrow \{of_1\}), (rf_1 \rightarrow \{kf_1\}), (rf_2 \rightarrow \{kf_2\}), (sf_1 \rightarrow \{kf_1\}), (sf_2 \rightarrow \{kf_2\}), (bf_1 \rightarrow \{kf_1\}), (bf_2 \rightarrow \{kf_2\})\}$

**Definition** A subset  $S$  of  $F$  is *indexable* iff

$$\forall f, f' \in S, f \neq f' \rightarrow ds(f) \neq ds(f')$$

Meaning that files need to be associated with different datasets. **Example** the set  $\{kf_1, of_1, rf_1, sf_1\}$  is *indexable* while  $\{kf_1, rf_1, rf_2\}$  is not indexable because  $ds(rf_1) = ds(rf_2) = roads$ .

**Definition** The function  $C: \wp(F) \rightarrow \wp(F)$  associates a subset of  $F$  with closure of the subset with respect to the  $dep$  function. It can be computed using the recursive function:

$$C(S) = \begin{cases} S \cup C(dep(S) \setminus S) & S \neq \emptyset \\ \emptyset & S = \emptyset \end{cases}$$

Where:

$$dep(S) = \bigcup_{s \in S} dep(s)$$

**Example**  $C(\{rf_1, sf_2\}) = \{kf_1, kf_2, of_1, rf_1, sf_2\}$

**Definition** Let  $I = \{i_1, i_2, i_3, \dots\} \cup \{\varepsilon\}$  be the set of indexes produced and  $\varepsilon$  is the empty index. The function  $from: I \rightarrow I$  associates an index with the index it was started from and the function  $files: I \rightarrow \wp(F)$  associate an index with the set of files to be added to the index we are starting from. Consider that the “*from*” function must not introduce a cyclic dependency among indexes.

**Example:**  $I = \{i_1, i_2, i_3, i_4\}$

$from = \{(i_1 \rightarrow \varepsilon), (i_2 \rightarrow i_1), (i_3 \rightarrow i_2), (i_4 \rightarrow i_2)\}$

$files = \{(i_1 \rightarrow \{kf_1\}), (i_2 \rightarrow \{rf_1, bf_1\}), (i_3 \rightarrow \{sf_1\}), (i_4 \rightarrow \{sf_2\})\}$

**Definition** Function  $\phi: I \rightarrow \wp(F)$  provides for each index the set of files that are indexed, it is defined recursively as:

$$\phi(i) = \begin{cases} \phi(from(i)) \cup files(i) & i \neq \varepsilon \\ \emptyset & i = \varepsilon \end{cases}$$

**Example**  $\phi(i_1) = \{kf_1\}$ ,  $\phi(i_2) = \{kf_1, rf_1, bf_1\}$ ,  $\phi(i_3) = \{kf_1, rf_1, bf_1, sf_1\}$ ,  $\phi(i_4) = \{kf_1, rf_1, bf_1, sf_2\}$

**Definition** An index  $i \in I$  is *correct* if  $C(\phi(i))$  is indexable meaning that in the closure of files in the index are not present different versions of files of the same dataset. **Example** the indexes  $i_1, i_2, i_3$  are correct while  $i_4$  is not correct because  $C(\phi(i_4)) = \{kf_1, kf_2, of_1, rf_1, bf_1, sf_2\}$  is not indexable.

**Figure 3** shows possible evolutions of an RDF Store with their corresponding index-descriptors and indexes. The figure wants to highlight that simultaneously can be carried out different versions of pairs: index, index-descriptor, each of which

containing different data. The different index colors indicate that each index may contain different data, according to the evolution with which it has been created. For example, considering the index, index-descriptor pair version labeled 1, including ontologies and vocabularies, we can assume that the pairs number 1.1 could be incrementally generated, starting from version 1 by adding geographic information and bus stops; and version 1.1.1 by adding services. Subsequently the need to create another alternative branch occurred since the bus stops changed positions, and thus version 1.2 was created by adding geographic information and the new bus stops; and from that version 1.2.1 adding again the services. Please note that version 1.2.2, represents an example of index generation by starting from version 1.2 by cloning index and index-descriptor, and adding other data set triples.

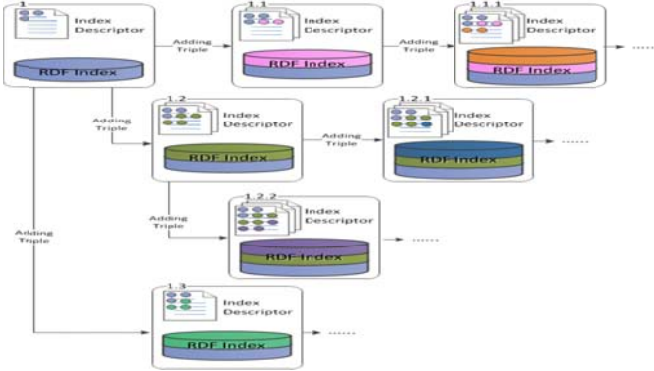


Figure 3. RDF KB store index versioning: reporting index-descriptors and indexed

In this last scenario, an existing index may be extended with new generated data set or updated by including new corrected versions of data and/or ontologies. Since the RDF KB building is an evolving process, it is not possible to predict whether one has to keep a specific previously created version of the index or not. Any small change could be used to generate a new version, while the suggestion is to save versions every time a consolidated point is available similarly to virtual machine snapshots. Moreover, since the triples associated with each single data set are accessible, reconstruction of partial intermediate versions are also possible, saving time in generating triples. Furthermore, each times some ontologies change, most of triples must be generated again, and therefore, for the same dataset, more triples versions could exist.

#### A. Requirements for **RDF Index Manager Tool**

On the basis of the above presented model, the RDF KB indexing versioning activities described can be supported by means of an RDF Index Manager (RIM), that should allow:

- Keep tracing *RDF KB Store Versions*, *RKBSV*, in terms of *files of triples*, *index-description*, and *RDF Index*;
- Maintaining a repository of *RKBSVs* where they could be stored and retrieved;
- Selecting a *RKBSV* from the repository for modification, to examine changes and the history version, to be used as base for building a new version;

- Managing the *index descriptor* as a list of files containing triples;
- Generating a RDF KB index on the basis of an *RKBSV* independently from the RDF store kind automatically, and in particular for *SESAME OWLIM* and *Virtuoso*;
- Monitoring the RDF KB index generation and the feeding state;
- Suggest the closest version of the *RKBSV* with respect to the demanded new index in terms of files of triples;
- Avoiding manually managing the script file of indexing, since it is time consuming and an error prone process.

#### IV. RDF INDEX MANAGER TOOL

The **RDF Index Manager** tool satisfy the above presented requirements, creates and manages *index descriptors*, and files of triples, and generates automatically the corresponding *indexes* independently from the RDF store type. The *index descriptor*, as mentioned before, is a list of ontologies and related data sets described with their triple files and version. The chosen approach with generation and update is to: (i) build the entire index (*build all*) by loading triples when ontologies and related data set change, (ii) extending the index when only new data sets and triples have to be added (*incremental building*), (iii) make a physical copy (*clone*) of a consolidated RDF index when an index descriptor is built starting from an older consolidated descriptor. The big amount of triples to load in the index suggested exploiting the bulk data loading supported by many RDF stores.

The main functionalities provided by the tool are described as following: Setup of a new index descriptor, to create an empty index descriptor; Clone a previous index descriptor to create a new version that it is populated with the same data sets and triples version of the parent with some addition. A clone of the parent RDF index is made and used to build the new store loading the new additional data sets; Copy a previous index with updated versions to create a new version populated with same data sets of parent and new versions of triples. This allows speeding up the creation of an update version of the index descriptor. A new RDF index will be created and loaded from scratch; Edit the index descriptor to add a data set (ontology, static, historical and reconciliations), select triples version; update triples version of a data set; remove a data set; Import/Export the index descriptions as XML representations that could be used for backup/restore and share; RDF Index Generation by producing a scripted procedure (for Windows and Linux) according to the index descriptor and the selected RDF store kind. The procedure may be incremental or for reconstructing the index from scratch; Monitoring the RDF Index Generation by controlling the store feeding as: the queue of data set to be loaded, the data set already in the store, time indicators (time spent, max time to upload a data set, etc..), progression and output of building process; Logging building data related to RDF store building for further access (i.e. statistical and verification analysis).



#### A. Architecture, RDF Index Generation and evolution

The RDF Index Manager is constituted by the following components. The *RDF Store Manager* manages different versions of RDF Stores exploiting the *Version Manager* which provides the triples files version for all the data sets. The *Index Manager Application Server and GUI* which is the user interface for creating, loading and editing the index, building and putting in execution the scripts for RDF store feed, monitoring the whole creation process. It also provides users management, user control access and configuration settings. The *Index Manager API REST Interface* consists of a set of REST calls to be invoked by the indexing script during the RDF store building to keep trace of the indexing process status. The *Index Builder Manager* generates the scripts according to the RDF Store kind. The section contains a list of ontologies/file and each file is described by: an unique identifier corresponding to the name, the reference to the index, the version of triples to use, the operation to perform add, update, remove and commit, and an entity for setting if it was inherited by a cloning (Clone). The historical data differs from other section for the presence of time interval that defines the triples to use (date and time for TripleStart and TripleEnd).

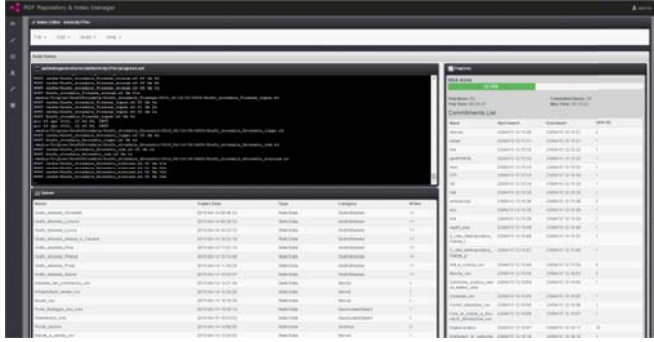


Figure 4. RDF Index Building Monitor

For the RDF Index generation the RDF Index Manager produces a script according to the index descriptor and the RDF store target. The script is structured in the following steps: (i) setup of script, (ii) initialization of RDF store, (iii) bulk uploading of triples into the store, (iv) RDF store finalization, (v) create possible additional indexes as textual indexes, geographical indexes that need additional database commands, and (vi) update index building status.

The RDF Index Manager has been realized as a PHP 5.5.x web application with MySQL support running under Ubuntu. The **Figure 4** shows the Building Monitor View when a batch script is running. This view provides different information panels: the output of script in real-time on top, the queue of data set to insert, the progress and the total time spent for the committed data set. Such information allows also evaluating the time necessary to build a repository using the two RDF Stores.

#### V. EXPERIMENTAL RESULTS

In Table 1, examples of results are reported. The data refer to the comparison of the usage of the RIM and versioning in building a Smart City RDF store. The RDF stores currently

managed are Virtuoso 7.2 as open source RDF store and the commercial OWLIM SE ver. 4.3 and GraphDB 6.1. The measures reported have been performed by means of an incremental building of the RDF Store for the three solutions. The building started with 12 files of triples including ontologies (first column), then each column of the table refers to the added triples/files (street graphs, smart city services, enrichment and reconciliations, historical data of real time data for 1 month). The time estimated for the cases of total indexing include: create, load, finalize; while those for incremental indexing include: clone, load, finalize. The three RDF store kinds have a different behavior. OWLIM and GraphDB create inferred triples at the indexing; this determines a higher number of triples with respect to Virtuoso, i.e., 73.4 wrt 46.2 million; and a higher indexing time. In both cases, the percentage of saved time, for non small RDF stores, is very high, greater than the 22% up to the 97% of saved time. For small stores, Virtuoso can be indexed in shorter time, and thus it could be better to rebuild instead of cloning and versioning.

	Ontologies	+ street graphs	+ smart city Services	+Enrich& Reconciliations	+Historical data 1 month
<b>Indexing process</b>					
Final number of triples	15809	33547501	34462930	34557142	44218719
Final number of Files	12	137	178	185	27794
Added triples with respect to previous version	15809	33531692	915429	94212	9661577
Added Files with respect to previous version	12	125	41	7	27609
<b>OWLIM SE 4.3</b>					
Indexing Time without RIM (s)	18	6536	6198	7516	12093
Indexing Time with RIM (s)	11	6029	514	343	5745
<b>% of saved time, RIM versioning</b>	<b>38,9</b>	<b>7,8</b>	<b>91,7</b>	<b>95,4</b>	<b>52,5</b>
Final Number of triples (including geo + inferred)	16062	57486956	59395432	59486748	73441126
disk space in Mbyte	310	8669	8936	9039	13110
<b>VIRTUOSO 7.2</b>					
Indexing Time without RIM (s)	146	806	964	1000	2487
Indexing Time with RIM (s)	156	833	421	296	1932
<b>% of saved time, RIM versioning</b>	<b>-6,8</b>	<b>-3,3</b>	<b>56,3</b>	<b>70,4</b>	<b>22,3</b>
Final Number of triples (including geo, no inferred)	21628	35452613	36301322	36420445	46232510
disk space in Mbyte	68	1450	1632	1631	2294
<b>GraphDB 6.1</b>					
Indexing Time without RIM (s)	9	7818	7929	7671	12915
Indexing Time with RIM (s)	2	6791	454	214	4849
<b>% of saved time, RIM versioning</b>	<b>77,8</b>	<b>13,1</b>	<b>94,3</b>	<b>97,2</b>	<b>62,45</b>
Final Number of triples (including geo + inferred)	15809	57486415	59394891	59487551	73441929

disk space in Mbyte	96	4276	4466	4643	5714
------------------------	----	------	------	------	------

Table 1 – Saving time using Index Manager with respect to rebuilding. Data collected on Ubuntu 64bit, 16 core x 2 Ghz, 500 Gbyte HD

## VI. CONCLUSIONS

Graph databases are used in many different applications: smart city, smart cloud, smart education, etc., where large RDF KB store are created with ontologies, static data, historical data and real time data. Most of the RDF stores are endowed of inferential engines that materialize some knowledge as triples during indexing or querying. In these cases, the delete of concepts may imply the removal and change of many triples, especially if the triples are those modeling the ontological part of the knowledge base, or are referred by many other concepts. For these solutions, the graph database versioning feature is not provided at level of the RDF stores tool, and it is quite complex and time consuming to be addressed as black box approach. In most cases, the RDF store rebuilt by indexing is time consuming, and may imply manually edited long scripts that are error prone. In order to solve this kind of problem, in this paper, a lifecycle methodology and our RIM tool for RDF KB store versioning are proposed. The results have shown that saving time up to 95% are possible depending on the number of triples, files and cases to be indexed.

## ACKNOWLEDGMENT

The authors would like to thank to the coworkers that have contributed to the experiments in the several projects, and in particular to Km4City: Giacomo Martelli, Mariano Di Claudio. Thanks also to Ontotext for providing a trial version of their tools.

## REFERENCES

- [Batarseh, Gonzalez, 2013] Batarseh, Feras A., and Avelino J. Gonzalez. "Incremental lifecycle validation of knowledge-based systems through CommonKADS." *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol.43, n.3, 2013, pp.643-654.
- [Bellini et al., 2013a] P. Bellini, M. Di Claudio, P. Nesi, N. Rauch, "Tassonomy and Review of Big Data Solutions Navigation", as Chapter 2 in "Big Data Computing", Ed. Rajendra Akerkar, Western Norway Research Institute, Norway, Chapman and Hall/CRC press, ISBN 978-1-46-657837-1, 2013
- [Bellini et al., 2013b] P. Bellini, M. Benigni, R. Billero, P. Nesi and N. Rauch, "Km4City Ontology Bulding vs Data Harvesting and Cleaning for Smart-city Services", *International Journal of Visual Language and Computing*, Elsevier, <http://dx.doi.org/10.1016/j.jvlc.2014.10.023>, 2013
- [Berners-Lee, 2006] T. Berners-Lee, "Linked Data", <http://www.w3.org/DesignIssues/LinkedData.html>, 2006.
- [Bizer et al., 2009] C. Bizer, A. Schultz. "The Berlin SPARQL Benchmark". *International Journal on Semantic Web & Information Systems*, Vol. 5, Issue 2, Pages 1-24, 2009
- [Bizer et al., 2011] Bizer, C., Jentzsch, A., Cyganiak, R.: State of the LOD cloud. <http://lod-cloud.net/state/> Retrieved July 5, 2014.
- [Erling et al., 2015] O. Erling, A. Averbuch, J.L. LarribaPey, Hassan Chafi, Andrey Gubichev, Arnau Prat, Minh-Duc Pham, Peter Boncz, The LDBC Social Network Benchmark: Interactive Workload. *Proceedings of SIGMOD 2015*, Melbourne.
- [Gómez-Pérez, 2004] Gómez-Pérez, A. *Ontology Evaluation. Handbook on Ontologies*. S. Staab and R. Studer Editors. Springer. International Handbooks on Information Systems. Pp: 251 – 274. 2004.
- [Grosan and Abraham, 2011] Grosan, C., and A. Abraham. *Intelligent Systems: A Modern Approach*, Springer-Verlag, Berlin, 2011.
- [Guo et al., 2005] Y. Guo, Z. Pan, and J. Heflin. "Lubm: A benchmark for owl knowledge base systems". *J. Web Semantics*, 3(2-3):158–182, 2005.
- [Hartig et al., 2009] 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.
- [Isele, Bizer, 2013] 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
- [Klein et al., 2002] M. Klein, D. Fensel, A. Kiryakov, and D. Ognyanov. "Ontology versioning and change detection on the web". In *Proc. of the 13th European Conf. on Knowledge Engineering and Knowledge Management (EKAW02)*, pages 197–212. Springer, 2002.
- [Klyne and Carroll, 2004] G. Klyne, J. Carroll, "Resource Description Framework (RDF): Concepts and Abstract Syntax - W3C Recommendation", 2004
- [Lopez, 1999] M. Fernandez Lopez, "Overview of Methodologies for Building Ontologies", in: *IJCAI99 Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends*, Stockholm, 1999.
- [Milette 2012] L. Milette, *Improving the Knowledge-Based Expert System Lifecycle*, UNF report, 2012.
- [Ngomo, 2011] Ngomo, A. C. N., & Auer, S. Limes-a time-efficient approach for large-scale link discovery on the web of data. *integration*, 15, 3. (2011).
- [Noy and McGuinness, 2001] Noy, Natalya F., and Deborah L. McGuinness. "Ontology development 101: A guide to creating your first ontology." Technical Report SMI-2001-0880, Stanford Medical Informatics. 2001.
- [Noy and Musen, 2004] N. F. Noy and M. A. Musen. "Ontology versioning in an ontology management framework". *IEEE Intelligent Systems*, 19(4):6–13, 2004.
- [Rector et al., 2004] Rector, A., Drummond, N., Horridge, M., Rogers, J., Knublauch, H., Stevens, R., Wang, H., Wroe, C. "Owl pizzas: Practical experience of teaching owl-dl: Common errors and common patterns". In *Proc. of EKAW 2004*, pp: 63 – 81. Springer. 2004.
- [Schmidt et al., 2009] M. Schmidt, T. Hornung, G. Lausen, and C. Pinkel. "Sp2bench: A sparql performance benchmark". In *ICDE*, pages 222–233, 2009.
- [Tzitzikas et al., 2008] Tzitzikas, Yannis; Theoharis, Yannis; Andreou, Dimitris, *On Storage Policies for Semantic Web Repositories That Support Versioning*, pp.705-719, LNCS 5021 *The Semantic Web: Research and Applications*, Springer, 2008
- [Volkel et al., 2005] M. Volkel, W. Winkler, Y. Sure, S. R. Kruk, and M. Synak. "SemVersion: A Versioning System for RDF and Ontologies". In *Proc. of the 2nd European Semantic Web Conf., ESWC'05., Heraklion, Crete, May 29 June 1 2005*.
- [Zegins et al., 2007] D. Zeginis, Y. Tzitzikas, and V. Christophides. "On the Foundations of Computing Deltas Between RDF Models". In *Proc. of the 6th Intern. Semantic Web Conf., ISWC/ASWC'07*, pages 637–651, Busan, Korea, November 2007.

# AVANZI

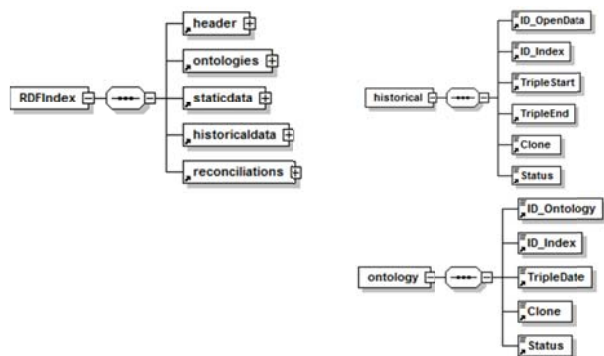


Figure 5. XML Schema of the index-descriptor, a part.

The XML schema of the index-descriptor data model is reported in **Figure 5**.

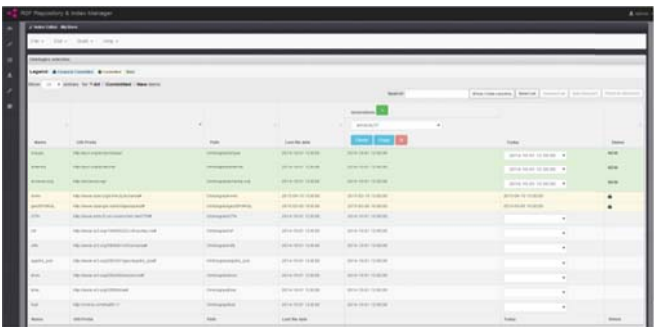


Figure 6. RDF Index Editor: Ontologies data set view

The **Figure 6** shows the Index Editor View for the selection of ontologies. The status of added data set is labeled as “NEW” while for the already committed is marked with a “lock”.

It is structured in five sections: *Header*, *Ontologies*, *Static Data*, *Historical Data* and *Reconciliations*. The Header section contains: the unique identifier of the index (ID), the name of the RDF store (RepositoryID), the path of the batch script for building the store (scriptPath), the reference to the parent (ParentID) when the index is cloned or copied, the chosen RDF store type (Type), a textual description (Description) to trace changes, the version number (Version), the time for building the RDF store (Generation Start and Generation End), the time spent to edit index descriptor (Session Start and Session End), the creation mode (Building\_Mode) and the network address of used index manager (SessionIP). Ontologies, Static Data, Historical Data and Reconciliations are modeled as a collection of data. Their definition is quite similar.

Ontologies allow formalizing the fundamental models of the knowledge and provide machine-executable semantics of data to make information understandable for the computer, thus assisting people to search, extract, interpret and process information. For example, by defining main entities of an application domain and their general relationships as: is-a, is-part-of, is-an-instance-of, etc., and specific relationships newly defined.

These solutions can be regard as high level services for manipulating versions in the graph model. They may be suitable to update triples from a local repository without deleting all triples, and assuming that existing triples can be removed by reference to the context; then new triples can be added in the context. These solutions can be viable in absence of inference and for non-bigdata applications. Consequently, in event of wide set of triples the cost of database analysis, triple identification and delete, reconstructing the consistency could be very high. In RDF store supporting inference strategies (at insertion and/or search time), delete operations are expensive and take as much as few minutes: deletion of a set of triples has also to remove the triples materialized by the inference. And, in some cases, triples materialized at the indexing may be not known without dumping the whole RDF store, since most of the RDF stores do not provide any support for it.

Many different RDF stores [W3C], [Haslhofer et al., 2011] exist and they are classified as: In-memory (triples are stored in main memory), Native store (Persistent storage systems with custom DBs), Non-native store (Persistent storage systems run on third party DBs). RDF stores are further classified on the basis of mapping triples as “subject, predicate, object” (SPO) onto storage-level tables, how they index triples, and how they process complex queries. Moreover, no explicit on native versioning support can be found in common RDF store as: 4store, 5store, Kowari, Mulgara, ARC, OWLIM, Jena TDB/SDB, Redstore, Blazegraph, AllegroGraph and Virtuoso [Haslhofer et al., 2011]. [Haslhofer et al., 2011] B. Haslhofer, E. M. Roochi, B. Schandl, and S. Zander. “Europeana RDF Store Report”. In University of Vienna, Technical Report, 2011. [http://eprints.cs.univie.ac.at/2833/1/europeana\\_ts\\_report.pdf](http://eprints.cs.univie.ac.at/2833/1/europeana_ts_report.pdf).

With the formal validation it is also possible to (i) verify if given a certain set of datasets all the depending former triples set have been loaded on the basis of their dependencies, (ii) get from the system, which is the closest version to be reused to proceeds to reach an identified index recipe.

This entails that index descriptor, index and triples could be built a lot of time. Maintaining their saved versions and tracking the history of assessment and changes became a serious problem. In addition, working with high volumes of triples, the time to populate the index may became a factor to take into account: according to the chosen RDF engine, in fact, the loading times of triples could take several hours [Rohloff et



al., 2007], [Bizer, Schultz, 2009], especially if the engine has not been properly configured and optimized or an inference strategy has been activated at the load-time.

[Rohloff et al., 2007] Rohloff, K., Dean, M., Emmons, I., Ryder, D., & Sumner, J. (2007, January). An evaluation of triple-store technologies for large data stores. In *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops* (pp. 1105-1114). Springer Berlin Heidelberg.

[Bizer, Schultz, 2009] Bizer, C., & Schultz, A. (2009). The berlin sparql benchmark.

The **typical domains** that present evolving models and data that are justified for graph databases can be smart cities, competence storages, or health care. In these cases, the data model can change over time in the system since new concepts may be added, for example coming from new data sets in the smart city, new competence sectors and relationships model that may change radically the former knowledge in competence modeling, in health, etc. (different data with low or no semantic interoperability among the concepts coming from different sources and the current version of the model). These facts, may lead to the revision of the ontological model, and at each change a process of verification & validation, V&V, is needed to avoid putting in production an RDF store with inconsistencies and/or incompleteness. These aspects are related to the big data concepts of variety, variability, etc. [Bellini et al., 2013a]. The variability of data is related to the frequency of data update, and it allows distinguishing static from dynamic data. Static data are rarely updated, such as once per month/year, as opposed to the dynamic data which are updated: from once a day up to every minute or more, to arrive at real time data; thus static, quasi-static, real time data and historical data.

The setup of the script consists of the declaration of all parameters involved in the RDF store creation process such as: location of triples, settings for the used RDF store (url of API REST interfaces, enabling inference at load-time), settings for logging, set the start time and output. The initialization of the RDF store executes all commands necessary to clone a parent store (when required), create a new store or cleaning an existing one. More commands could be necessary according to type of the RDF store to manage. The bulk data loading inserts into the RDF store all files of triples for each data set included in the index descriptor splitting them in smaller ones if too big. The finalization of the RDF store executes all commands required by the RDF store type such as: activate the ontology, run inference after load, create specific relationships (i.e. geo-spatial relationships), flush the memory and save the store on disk, etc... In last step, the script executes all commands to update the index building status and mark the status of data set in the index descriptor as committed.

[Bellandi, .....] OSIM

[4] Hogan, A., Harth, A., Passant, A., Decker, S., Polleres, A. Weaving the Pedantic Web. Linked Data on the Web Workshop LDOW2010 at WWW2010 (2010).

[5] Archer, P., Goedertier, S., and Loutas, N. D7.1.3 – Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC. Deliverable. December 17, 2012.

[6] Heath, T., Bizer, C.: *Linked data: Evolving the Web into a global data space* (1st edition). Morgan & Claypool (2011).

[Hoplin & Erdman 1990] Hoplin, H., and S. Erdman. "Expert Systems: An Expanded Field of Vision for Business," *Proceedings of the 1990 ACM SIGBDP conference on Trends and directions in expert systems*, SIGBDP (1990), pp. 1-16.

[SPARQL] Prud'hommeaux, E., Seaborne, A., *SPARQL Query Language for RDF*, <http://www.w3.org/TR/2004/WD-rdf-sparql-query-20041012/>

[W3C] W3C Large Triple Stores at <http://www.w3.org/wiki/LargeTripleStores>

[Mosqueira00] E. Mosqueira-Rey and V. Moret-Bonillo, "Validation of intelligent systems: A critical study and a tool," *Expert Syst. Appl.*, vol. 18, no. 1, pp. 1–16, Jan. 2000.

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.

Dresden, Germany. A.Kiryakov , D. Ognyanov, D. Manov, OWLIM – a pragmatic semantic repository for OWL, *Proceedings of the 2005 international conference on Web Information Systems Engineering*, November 20-22, 2005, New York, NY [doi>10.1007/11581116\_19]

Klein, M., Fensel, D., "Ontology versioning on the Semantic Web", *Lecture Note in Computer Science*, Springer Berlin Heidelberg, ISBN: 978-3-540-44268-4, pp: 247-259, 2002.

N. F. Noy and M. A. Musen. "Ontology versioning in an ontology management framework". *IEEE Intelligent Systems*, 19(4):6–13, 2004.

Volkel, M., Groza, T., "Semversion: An RDF-based Ontology Versioning System", *IADIS International Conference WWW/Internet 2006*, ISBN: 972-8924-19-4, Vol. VI.2, pp. 195-202, 2006.

# Robust Radial Distortion Estimation Using Good Circular Arcs

Xiaohui Zhang, Weibin Liu  
Institute of Information Science  
Beijing Jiaotong University  
Beijing 100044, China  
e-mail: wblu@bjtu.edu.cn

Weiwei Xing  
School of Software Engineering  
Beijing Jiaotong University  
Beijing 100044, China

**Abstract**—It is a common problem, radial distortion of off-the-shelf cameras, especially those low-cost ones and wide-angle ones. And the most direct method to judge whether radial distortion occurs in an image is the straightness of those lines in the image, for the straight line in the image should be straight in the ideal image under the pin-hole camera model. In this paper, we present a new line-based approach to eliminate the radial distortion which makes the line distorted in the image. It is based on the fact the straight lines in the real world project to circular arcs under the single parameter division model. Compared with those former line-based methods, the method in this paper work well and be easy to find the good circular arcs, which is valid to eliminate the interference of other curves. Experiments are provided for both synthetic and real images and the results show that our method can remove the radial distortion from images validly and robustly.

**Keywords**-radial distortion; division model; circle fitting; good circular arcs;

## I. INTRODUCTION

An ideal pin-hole camera model is used by most computer vision algorithms, for example, 3D reconstruction, quantitative measurement, recognition and tracking of objects, etc. Its basic assumption is that the 3D straight lines mapping into the image plane are still straight lines. However, in reality small or large amounts of distortion are introduced by most lenses, which bend straight lines in the real world into curves. This may introduce severe problems in the preceding vision algorithms, which making distortion correction is a must.

The imperfection of the lens and the misalignment of the optical system lead to the distortion. And radial distortion is one of kinds of distortion, but it is considered as the predominate one among all possible lens distortion [1 2]. The lens introduces barrel distortion at short focal lengths while it introduces pincushion distortion at longer focal lengths. The polynomial model presented by D.C. Brown [3] has been widely applied for an excellent trade of between complexity and accuracy. Another model widely used lately is the division model presented by

Fitzgibbon [4]. And the single-parameter radial distortion model is enough for most lenses and any more elaborated modeling would not help, but also would cause numerical instability [1].

In general, methods for correcting radial distortion can be divided into three categories [5 6]. The first method is multiple view auto-calibration [4 7 8]. No knowledge of the scene and no special pattern is required, but it is not suitable for the distorted image from an unknown source. The second one is point correspondence [1 9]. They identify image points using a known pattern and estimate the distortion parameters as part of the internal parameters of the camera. Hence the results are highly reliable and accurate, but it needs to get multiple images from different views. The last one is plumb-line based method [5 10 11], which assumes straight lines in the real world should be still straight on the image plane. The biggest advantage is that it can correct the distortion only using one image, while the disadvantage is that it needs sufficient lines in the image scene.

What we most expect to the correcting method is easy to remove the distortion automatic and robust from one unknown source image, which is simple and does not need special pattern. Therefore, plumb-based method is the only choice, which can satisfy all the demands. And Wang et al. [10] provides a simple method using the single-parameter division model under the principle that a straight line in the distorted image is a circular arc. The biggest advantage of Wang's method is that it can estimate the distortion center and the parameter of the single-parameter divide model simultaneously. And another advantage of Wang's method is that it can estimate the distortion center and the distortion parameter of the division model only using few straight lines. In principle, three straight lines are enough for computing the distortion center and distortion parameter, and if the distortion center is the image center, it just needs one or two straight lines. Hence, Wang's method can avoid other plumb-line based disadvantage which needs sufficient lines in the image scene. However, Wang's method is not an automatic one that it requires to extract the straight lines manually and the correction results are not robust which vary with the different straight lines.

Then, Bukhari and Dailey proposes an automatic method based on Wang's method which can extract the straight lines automatically and it uses as more straight lines as possible. However, the correcting results are still not robust. And the reason is that they do not solve the problem, selecting out good circular arcs from the curves detected from the distorted image. These line-based methods depend on extraction of long,

This research is partially supported by National Natural Science Foundation of China (No. 61370127, No.61100143, No.61473031, No.61472030), 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). The opinions expressed are solely those of the authors and not the sponsors.  
{ Corresponding author: Weibin Liu, wblu@bjtu.edu.cn }

smoothly curved edges and get thrown easily if a portion of such curves originate from non-linear scene structures [12].

Therefore, in this paper, we analyze all the curves which appear in the distortion images and classify the curve into two kinds, good curves and bad curves. Good curves are the curves projected by the long straight line of the real world, which is crucial to correct the distorted images in the line based methods, especially those automatic ones. Good curves are the edges of the artificial objects, e.g., buildings, signboards, roads, and so on, which are consisted of straight lines. Contrasting to the good curves, bad curves are mainly consisted of three kinds of curves identifying from the edge image. One kind is the curves are both curves either in the distorted image or in the ideal image which means that the curves are still curves, not straight lines, in the ideal images. And one kind is the short curves which might be the curves generating from the lines due to the distortion but is too short to be estimated accurately. Those curves have a common feature that they are too short to be used to estimate the distortion center and the distortion parameter. The last kind the straight lines passing through the distortion center is still straight lines in the distorted images. The reason is that the distortion is mainly radial distortion which makes the pixels move along the radial direction. Therefore, the lines passing through the distortion center are still straight lines either in the distortion image or in the ideal image.

We proposed a non-iterative method to solve this problem selecting out the good curves from all the curves that we extract from the distortion image. And then, we use the good curves to correct the distortion thus avoiding the impact of the bad curves. Our contribution is to make the process fully automatic and robust and it can eliminate the interference of those bad curves very well. The results from the experiments on the synthetic and real image show that the proposed method is simple and valid.

The organization of the remainder of this paper is as follows: Section 2 reviews how to estimate the distortion parameters of the division model and drives an invariant for those points of those lines in the distorted image. Section 3 presents the details of our method. We firstly describe the procedure to select the possible circular arcs from the edge image and estimate their three parameters. Secondly, we find good circular arc using the invariant. Experiments on synthetic and real image are presented in section 4. Section 5 we perform a direct comparison of our method with that of Bukhari-Dailey method [11]. Finally, some conclusions are drawn.

## II. ESTIMATE THE PARAMETER OF DIVISION MODEL AND THE INVARIANT

In this section, we review the division model used in this paper and show how to estimate the parameters of this model. Then, we derive an invariant for the points on the curves in the distorted image.

### A. Division model

The so-called division model, introduced by Fitzgibbon [4], is

$$r_u = \frac{r_d}{1 + \lambda_1 r_d^2 + \lambda_2 r_d^4 + \dots} \quad (1)$$

Where  $(x_u, y_u)$  and  $(x_d, y_d)$  are the corresponding points of the undistorted image and the distorted image respectively.  $r_u$  and  $r_d$  are the Euclidean distances of the undistorted point the distorted point to the distortion center  $(x_0, y_0)$ .  $\lambda_i$  is the parameter of the model which present the radial distortion.

The advantage of the division model is that it requires fewer distortion parameters than the polynomial model [13 14] for the case of severe distortion [10] and it can estimate the distortion center at the same time. For most cameras, many works [10 11] showed that only the first order radial distortion parameter is sufficient. It can be formulated as:

$$r_u = \frac{r_d}{1 + \lambda_1 r_d^2} \quad (2)$$

And we can write it in the following form

$$\begin{aligned} x_u &= x_0 + \frac{x_d - x_0}{1 + \lambda_1 r_d^2} \\ y_u &= y_0 + \frac{y_d - y_0}{1 + \lambda_1 r_d^2} \end{aligned} \quad (3)$$

Where,  $r_d^2 = x_d^2 + y_d^2$

Another advantage of the division model is that we can easy get the inverse of the single parameter division model. Hence, the pixel coordinates of the distortion image pixels can be presented by the pixel coordinates of the undistortion image pixels. Thus, we can get all the pixel values of the undistortion image using the inverse to find the pixel values of corresponding coordinates in the distortion image. Due to the computing results of the pixel coordinate of the distortion image pixels are not integers, we use simple bilinear interpolation in all of the experiments reported on in this paper.

In order to invert the single-parameter division model [11], we first square the (3) to obtain

$$r_u^2 = \frac{r_d^2}{(1 + \lambda_1 r_d^2)^2} \quad (4)$$

Where,  $r_u^2 = (x_u - x_0)^2 + (y_u - y_0)^2$

Simplifying the (4), we can get

$$r_d^2 - \frac{1}{\lambda_1 r_u} r_d + \frac{1}{\lambda_1} = 0 \quad (5)$$

For the positive  $\lambda_1$ , when the distortion is pincushion distortion, given  $0 < r_u^2 < \frac{1}{4\lambda_1}$ , (5) has two positive real roots. We use the smaller one. For negative  $\lambda_1$ , when the distortion is barrel distortion, given any  $r_u^2 > 0$ , there are two real solution. We use the positive one. Thus,  $r_d$  can be presented by  $r_u$ . Then, the image coordinates  $(x_d, y_d)$  can be obtained as the following formula

$$\begin{aligned} x_d &= x_0 + \left(\frac{r_d}{r_u}\right)(x_u - x_0) \\ y_d &= y_0 + \left(\frac{r_d}{r_u}\right)(y_u - y_0) \end{aligned} \quad (6)$$

### B. Estimating distortion parameters using the line points from distorted image

Wang et al. [10] has demonstrated that the straight lines in the real world project to circular arcs under the single parameter division model. And Wang et al. use the slope-y-intercept equation from of a line. Similarly, Bukhari and Dailey [11] obtain the same conclusion using the general equation form of a line. For its advantage, we use the general equation form of a line. And it can be written as:

$$ax_u + by_u + c = 0 \quad (7)$$

Using (3) and (7), we are easy to obtain the circle equation

$$x_d^2 + y_d^2 + Dx_d + Ey_d + F = 0 \quad (8)$$

Where

$$\begin{aligned} D &= \frac{a}{c\lambda} + 2x_0 \\ E &= \frac{b}{c\lambda} + 2y_0 \\ F &= x_0^2 + y_0^2 - \frac{a}{c\lambda}x_0 - \frac{b}{c\lambda}y_0 + \frac{1}{\lambda} \end{aligned} \quad (9)$$

According to the relation of D, E, and F, we obtain the following equation from (9)

$$x_0^2 + y_0^2 + Dx_0 + Ey_0 + F - \frac{1}{\lambda} = 0 \quad (10)$$

Under (8), we can estimate a group of parameter ( $D, E, F$ ) by circle fitting method using points belonging to a “straight line” which extracted from the distorted image. Consequently, we can use three groups of parameter ( $D_i, E_i, F_i$ ) $_{i=1,2,3}$  to compute the coordinates ( $x_0, y_0$ ) of the distorted center, that is

$$\begin{aligned} (D_1 - D_2)x_0 + (E_1 - E_2)y_0 + (F_1 - F_2) &= 0 \\ (D_2 - D_3)x_0 + (E_2 - E_3)y_0 + (F_2 - F_3) &= 0 \\ (D_3 - D_1)x_0 + (E_3 - E_1)y_0 + (F_3 - F_1) &= 0 \end{aligned} \quad (11)$$

Then we can obtain the radial distortion parameter

$$\frac{1}{\lambda} = x_0^2 + y_0^2 + Dx_0 + Ey_0 + F \quad (12)$$

### C. The invariant for circular arcs in the distorted image

Let ( $x_c, y_c$ ) and  $R_c$  are the center coordinates and radius of a circle by fitting points which belong to a “straight line” extracted from the distorted image. Using (8), we have

$$\begin{aligned} x_c &= -\frac{D}{2} \\ y_c &= -\frac{E}{2} \\ R_c &= \sqrt{\frac{D^2 + E^2 + 4F}{4}} \end{aligned} \quad (13)$$

---

Algorithm 1: Choosing good circle arcs and estimate parameters.

---

**Input:**

Arcs parameters set  $\{(x_c, y_c)_i\}$  and  $R_c^i$   $i = 1, 2, 3, \dots, \text{NumberOfArcs}$   
 $[\min, \max]$  are the range of  $C_{lg}$  and  $T$  is the interval for counting the number of  $(\frac{\text{width}}{2}, \frac{\text{height}}{2})$  is the image center

**Output:**

$\lambda, x_0, y_0$  are the distortion parameters

**Begin:**

Compute the constant using Eq. (14) for all candidate circular arcs

**If**  $\text{NumberOfArcs} \geq 3$

Divide the  $[\min, \max]$  into equal intervals, each adjacent interval overlaps half interval, and for each interval, count the number of constant those fall into the interval. Find the interval with maximal constant values support. Then, compute the mean value or mid-value of those constant as the ideal value denoted as  $C_m$ . Chose the circular arcs which constant values fall into the area  $[C_m - T/2, C_m + T/2]$  as good circular arcs

**End**

Estimate  $\lambda, x_0, y_0$  using good circular arcs

**End**

---

After reformulating the equation (12), we obtain

$$\frac{1}{\lambda} = (x_0 - x_c)^2 + (y_0 - y_c)^2 - R_c^2 \quad (14)$$

From the equation (14), we can know that the difference of square of Euclidian- distance of the center of the distorted image and the center of the fitting circular arcs with square of radius of the fitting circular arcs is an invariant to all the “straight lines” in the distorted image. And we can use this invariant to find the good circular arcs, which is valid to eliminate the interference of other curves.

## III. ROBUST ESTIMATION METHOD

In this section, we describe the details of our automatic and robust method to correct the radial distortion using the single parameter division model.

### A. The main procedure of our method

To sum up, the whole process to remove the radial distortion includes four steps and is presented as follows:

1. Extract image edges in the distorted image for detecting circular arcs.
2. Identify circular arcs from the image edges and estimate their parameters for each arc, the coordinates of the circular arc center and the circular arcs radius are included.
3. Find good circular arcs for computing the parameter of radial distortion.
4. Compute the distortion parameter and correct the distorted image using the single parameter division model.

For the first step, we employ the Canny Detector [15] to extract the image edges and link adjacent edge pixels remaining

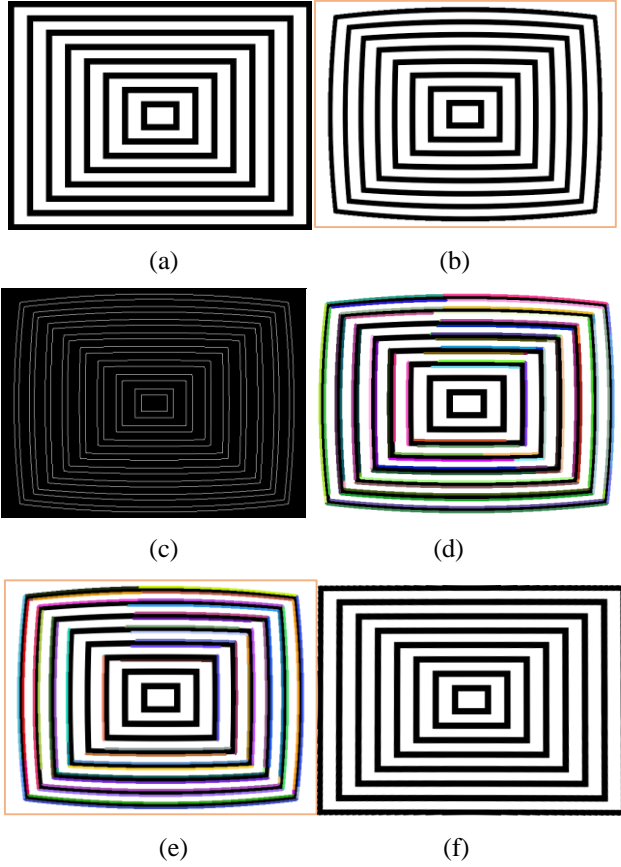


Figure 1. Undistorted processes on the synthetic image. (a) The ideal synthetic image with no distortion. (b) The synthetic image with distortion parameter  $\lambda_{true} = -1.0 \times 10^{-6}$ . (c) The result of canny edge detection. (d) The synthetic image with detected circular arcs. (e) The synthetic image with detected good circular arcs. (f) Undistorted image of the synthetic image. The size of the interval is 0.6.

contours [11]. And we discard those short contours and remain the long ones for more reliable information they provide and when fitting the short and therefore straight edges in circles, the estimated parameters are known to be unstable [16]. Hence, a threshold is set. If the contours whose number of pixels is less than the threshold are discarded for they are too short to be used. For the second step, a modified RANSAC method is used to detect circular arcs not overlapping with other arcs in the same contour that have more support [11]. The termination criterion of the modified algorithm is that to stop once the probability that an arc of minimal length has not yet been found is small. In order to refine the estimated circle parameters, we use the Levenberg-Marquardt (LM) iterative nonlinear least squares method [17] to estimate them. In the next step, we introduce a voting process to filter the good circular arcs, which is presented in detail in the flowing subsection. We can select out the good circle arcs from the circular arcs that extracted from the distortion image in the previous step. In the last step, (11) and (12) are used to compute the distortion parameter and distortion center by using the circle parameters which are found in the preceding step. And then, (6) is used to get the undistortion image. Thus, we implement a process of correcting distorted images automatically. In order to get robust results, the third step is very important.

### B. Choosing the good circular arcs

According to the (14), we can know that good circular arcs have the same constant. Inversely, we can screen out the good circular arcs by a voting process for the constant with maximal support inspired by Hough transformation [18]. And the details are presented in Algorithm 1. Due to the distortion center is unknown, we use the image center to replace it in the algorithm 1 for that the distortion center is nearby of the image center in most of distorted images [17]. Therefore, compute results have deviation with the ideal value but still fall into a small range nearby the ideal value, showed in Fig. 1. And we also conclude from (14) that the ideal constant is reciprocal of the distortion parameter. In reality, the distortion parameter is very small, generally less than 0.00001, hence the constant is very large. So we transform the constant into a special logarithm domain. The transformation relation is presented as follows:

$$C_{lg} = \begin{cases} \lg(c) & \text{if } c > 0 \\ 0 & \text{if } c = 0 \\ -\lg(-c) & \text{if } c < 0 \end{cases} \quad (15)$$

$$c = \frac{1}{\lambda} \quad (16)$$

Where  $c$  is the compute results of (14), and  $C_{lg}$  is value in the special logarithm domain. And the region  $-15 \leq C_{lg} \leq 15$  is enough, that is  $|\lambda| \geq 10^{-15}$ .

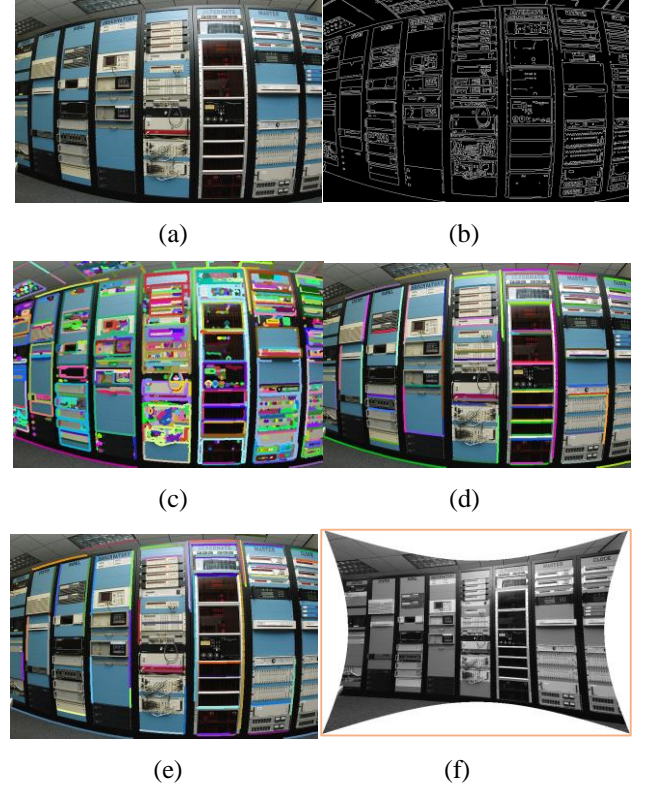


Figure 2. Undistorted processes on the real image. (a) The real distorted image. (b) The result of canny edge detection of the real image. (c) The result of liked contours. (d) The real image with detected circular arcs. (e) The real image with detected good circular arcs. (f) Undistorted image of the real image. The size of the interval is 0.6.

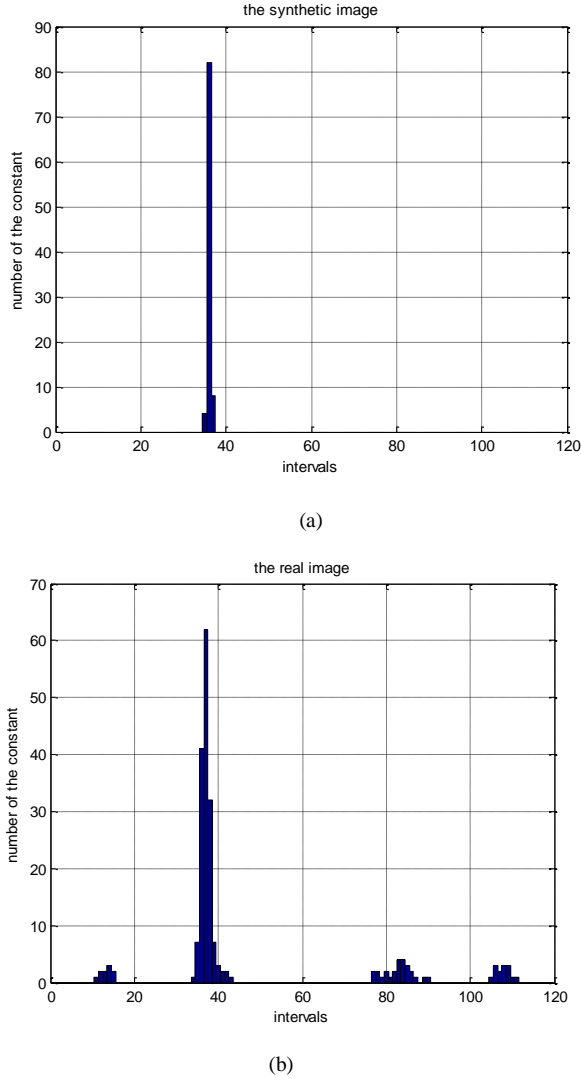


Figure 3. The distribution of the constant values in special logarithm space. (a) Distribution of the constant values of the synthetic image. (b) Distribution of the constant values of the real image.

#### IV. EXPERIMENTS AND RESULT

In this section, we present a detailed study of our method on synthetic and real image data. We distort the same original image (see Fig. 1(a)) for all the synthetic images by using particular ground truth values of the distortion parameters and the division model. And the size of all the synthetic images is 640x480. The minimum length of the detected lines is 100 pixels.

##### A. Experiments on synthetic and real images

In order to test and verify our method, we performed experiments on synthetic and real images. The synthetic image with known distortion parameter  $\lambda_{true} = -1.0 \times 10^{-6}$  and distortion center (320, 240) is shown in the Fig. 1(b). And Fig. 2(a) is a real image of 800x531 obtained from a publicly available database [19].

The Fig. 1(c) is the canny detection result of the synthetic image and in the Fig. 1(d), the detected curves are represented by using a different color to identify them. The synthetic image has sufficient “straight lines” which are the basis for correcting the distortion. The Fig. 1(e) shows the results of the good arcs which are selected out from the detected curves presented in the Fig. 1(d). The correcting result of the synthetic image is presented in the Fig. 1(f). It is obviously that the proposed method can undistort the synthetic image very well.

Fig. 2 are the undistortion process of the real image. Similarly, the total process have five parts which are described in detail in the previous chapter. The results of canny detection, contours linking, identifying circular arcs and finding good circular arcs are presented in the Fig. 2(b)-Fig. 2(e) respectively. Fig. 2(f) that the proposed method also has a good correcting result on the real image.

As observed in Fig. 3(a) the constant values are gathered into a small range, about two or three intervals, while in the Fig. 3(b), most of the constant values are into a small range but some of them fall into other intervals. It is mainly because the circular arcs detected from the synthetic image are all good ones, while the circular arcs detected from the real image include the bad ones, as explained in Section 3. The constant values of good and bad circular arcs are distributed into different intervals, hence we are easy to separate them in the special logarithm space.

##### B. The influence of the size of interval

As described in the proposed method, the special logarithm is divided into equal size intervals and the number of constant values are counted which fall into the same interval. Therefore, the size of the interval has a significant impact on the good circular arcs we selected. In this subsection, we discuss the influence of the size of the interval on the synthetic and real image. The size of the intervals is set to 0.2, 0.4, 0.6, 0.8 and 1.0 respectively in the experiments. The synthetic image we use is the distorted image of  $\lambda_{true} = -1.0 \times 10^{-6}$ , showed in the Fig.1 (b). And the real image we use is the Fig. 2(a). The results of the experiments are presented in the Fig. 4.

In the Fig. 4, first row are distributions of the constant values of the synthetic image under the different size of intervals and the second row are the corresponding synthetic undistorted images. From the third, we can know that the distribution of constant values are nearly the same concentrated into only a small range, two or three intervals, which illustrates that the assumption of the proposed method is right. The reason why the size of the intervals has little influence on the synthetic image is that the circular arcs detecting from the image are very good. Hence, correcting results in the second row are all good. Third row: distribution of the constant values of the real image under the different size of intervals. And the fourth row: corresponding undistorted real images. From correcting results in the fourth row, we can know that the results turn a little bad when the size of intervals is too big or too small. We can know the reasons from the third row that the constants values fall into more intervals when the size of the interval is too small. So the interval we use to select the good circular cannot include all the good circular arcs we want. While the interval would contain the bad ones when the size is too big. From the results, the size



of the interval should not too big or too small and should be set between 0.4 and 0.8. In additional experiments, we set the size is 0.6.

### C. Experiments on image with varying distortion centers

In Algorithm 1 we compute the constant values using the image center instead of the distortion center, hence we test our method on synthetic images with different distortion center. In the following experiment, to find the influence of the distortion center, the centers of the synthetic images are: (320, 240), (290, 210), (260, 180) and (230, 150). And the distortion parameter is  $\lambda_{true} = -1.0 \times 10^{-6}$ .

The results are presented in Fig. 5. In the first row, from left to right, distorted images with different centers are shown. The second row illustrates the distribution of constant values of the distorted images with different distortion center. As observed, when the distance of distortion center with the image center is greater, the distribution of the constant values is more scattered, that is, the constant values are distributed in more intervals, but the maximum is still in the same interval. Hence, the proposed method is still suitable for the case that the distortion center is not in the center of the image. And the undistorted results in the third also demonstrate that the distortion is good to eliminate.

### D. Experiments on image with different distortion parameter

In order to verify the performance of the proposed method, we take a series experiments with varied distortion parameter  $\lambda$ . The distortion centers of the synthetic images fix at  $(x_0, y_0) = (320, 240)$ . And the distortion parameters  $\lambda_i$  are respectively:  $-1.0 \times 10^{-5}$ ,  $-1.0 \times 10^{-6}$ ,  $-1.0 \times 10^{-7}$  and  $-1.0 \times 10^{-8}$ .

Fig. 6 shows some results of the experiments, first row are the distorted images at different levels of  $\lambda$ , second row are the distributions of the constant values of the synthetic images with different distortion parameter, and the third row are corresponding undistorted images. From the row of the Fig. 6, the absolute value of the distortion parameter is bigger. The distortion is more serious. And in general, the value of the distortion parameter is still very small, even though the distortion is very serious. When  $\lambda = -1.0 \times 10^{-8}$ , the distortion is very small which cannot be watched out with human eyes. Hence, range of the special logarithm, we set the previous chapter, is enough. From the second row of Fig. 6, with the varying of the distortion parameter, the distribution of the constant values is different which are concentrated into different intervals.

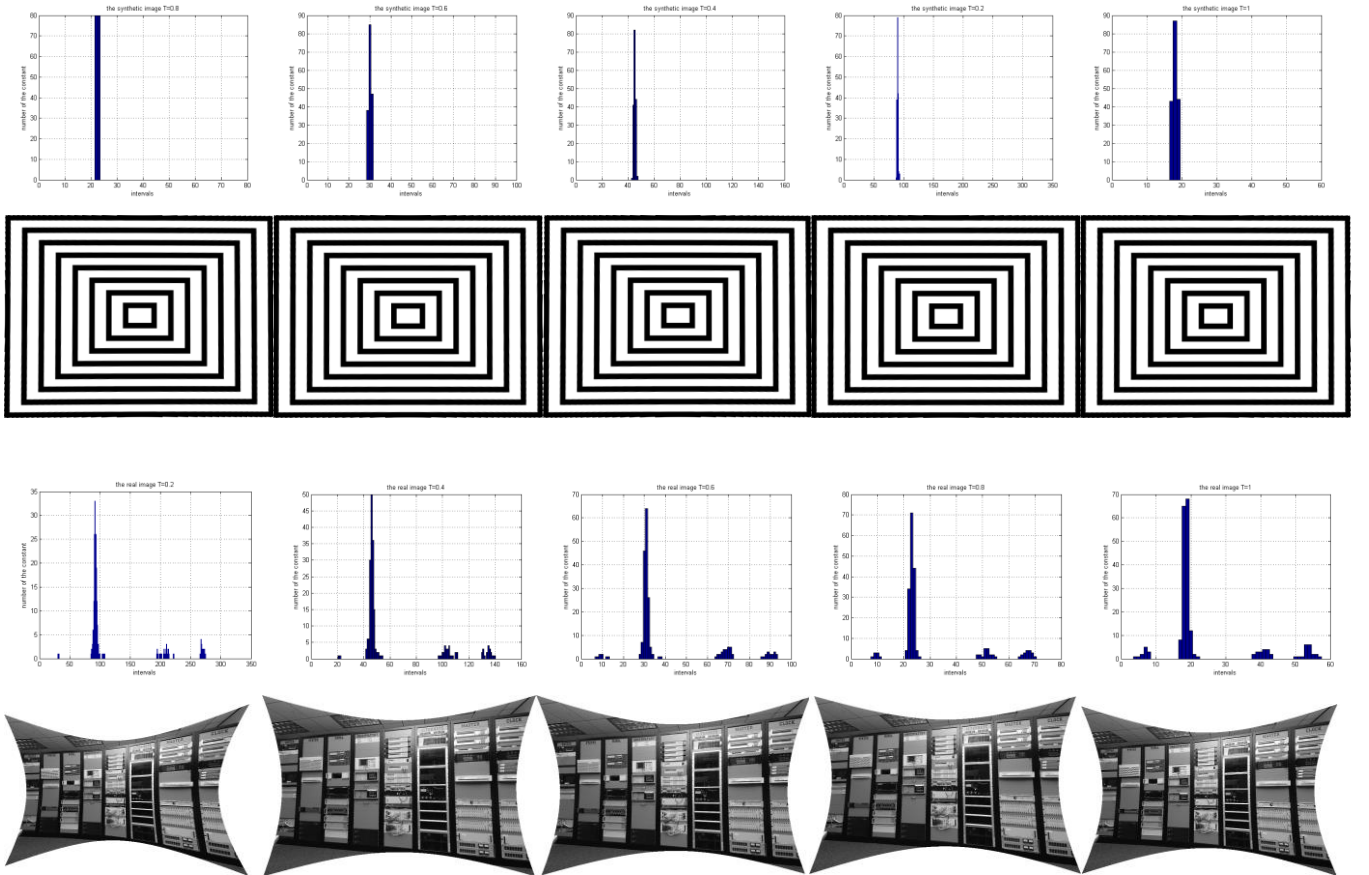


Figure 4. Undistortion of synthetic image and real image in different size of interval. Each column presents the image with the same interval size, and the size of intervals are 0.2, 0.4, 0.6, 0.8 and 1.0. The first row and the second row are the results of the synthetic images, and the third row and the forth row are the results of the real images. First row: distribution of the constant values of synthetic image under the different size of intervals. Second row: corresponding undistorted synthetic images. Third row: distribution of the constant values of real image under the different size of intervals. And fourth row: corresponding undistorted real images.

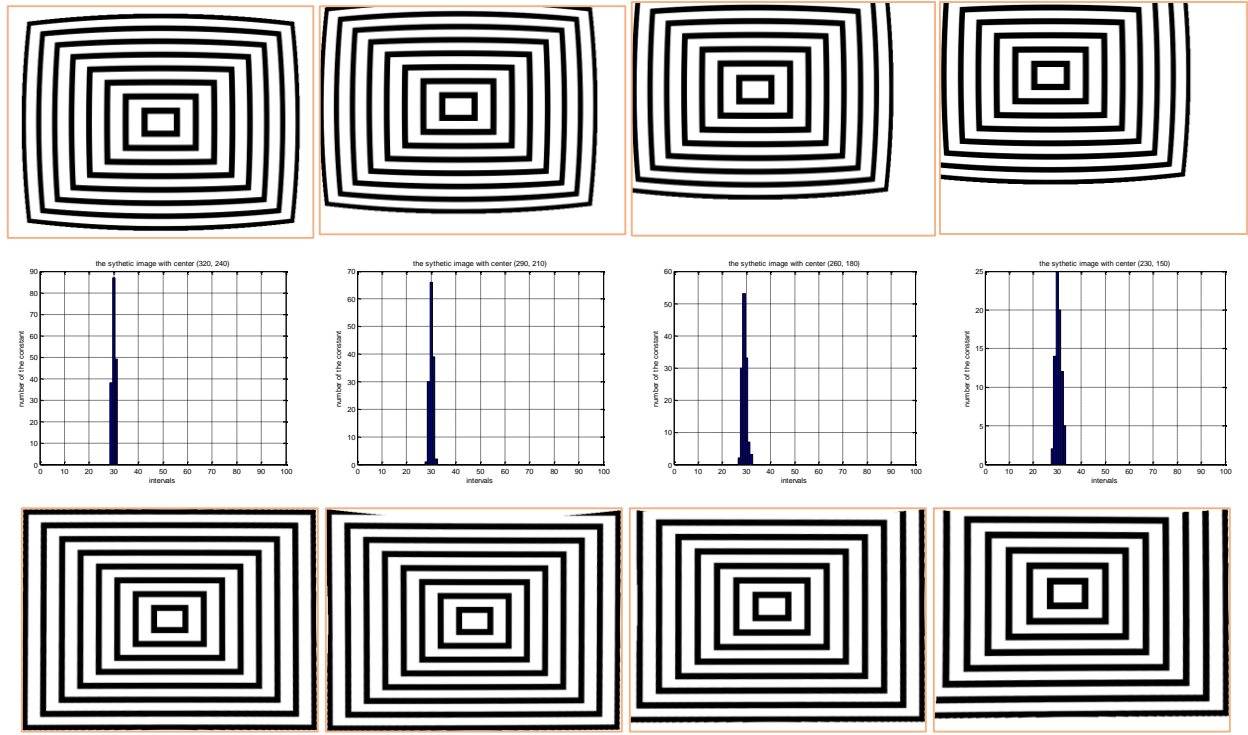


Figure 5. Correcting synthetic images with different distortion center. The image size is 640x480 and the columns from left to right the corresponding centers are: (320,240), (290,210), (260,180) and (230,150). First row: distorted images with different distortion centers. Second row: the distribution of the constant values of the corresponding distorted images. Third row: corresponding undistorted images.

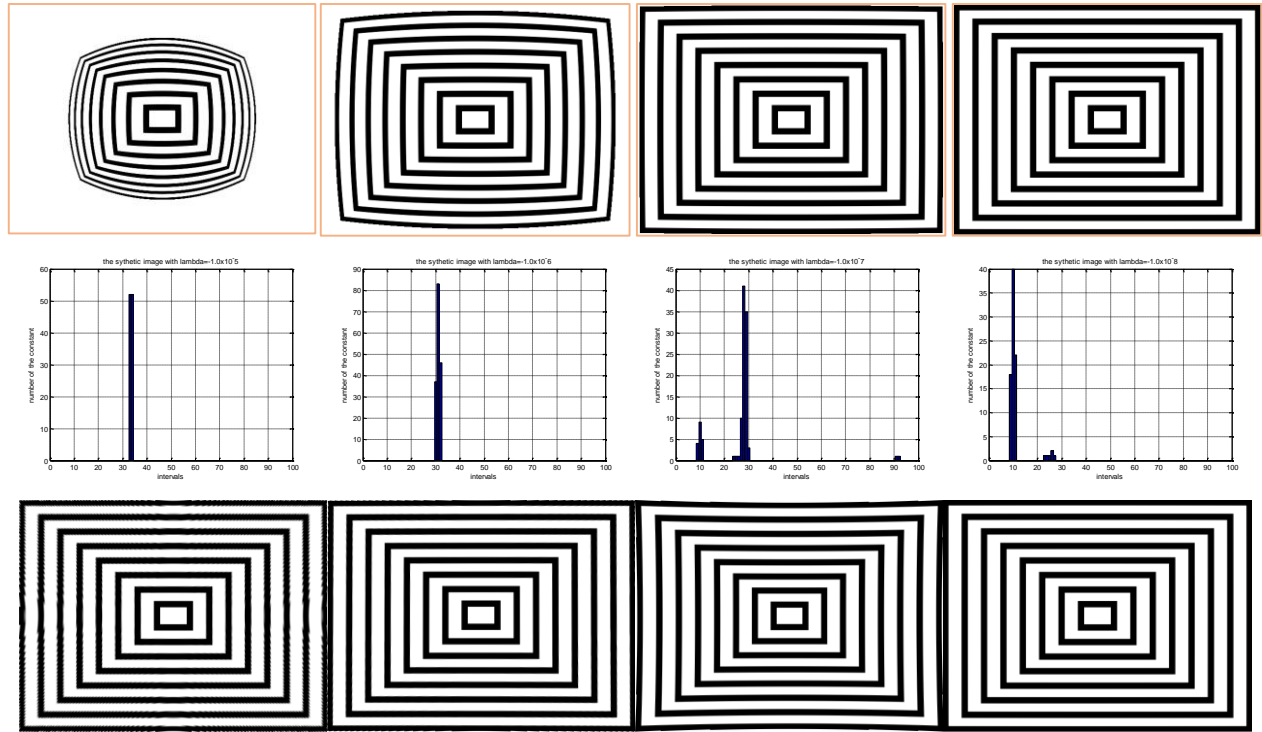


Figure 6. Undistortion of synthetic images with different distortion parameters. Image size is 640x480 and distortion center is (320,240). First row: distorted images at different levels of lambda. Second row: distribution of the constant values of the synthetic images with different distortion parameter. Third row: corresponding undistorted images



Figure 7. Lens distortion correction for a real image:(a) Original image, (b) detected candidate arcs, (c) choose the good arcs using the Bukhari-Dailey method, (d) choose the good arcs using the proposed method, (e) undistorted image using the Bukhari-Dailey method, and (f) undistorted image using the proposed method.

When the distortion parameter is small, there are several constant values falling into other intervals which are far away from the intervals most constant values concentrated. The reason is that there is some bad circular arcs in the curves we extracting from the edge image and the source of the bad

circular arcs are almost straight curves which parameter cannot be estimated accurately. From correcting results in the third row, we can know that the results of the small distortion images are not as well as that of the serious ones. It mainly the reason is that

the circular parameter cannot be estimated accurately for the length of the curve is relatively small with the full circle.

## V. COMPARISON WITH BUKHARI-DAILEY METHOD

Fig. 7 presents the results of a real image with 422x311, from a publicly available database, shown in Fig. 7(a). And Fig. 7(b) shows the arcs detected results which are identified with different colors. Fig. 7(c) presents the results of the good arcs selected by the Bukhari-Dailey method, whereas Fig. 7(d) presents the results of good arcs selected by the proposed the bad circular arcs, while the proposed method can. For instance, the curves of the window and door in the image are bad arcs which should be removed and the short curves is also should be eliminated. Fig. 7(e) and Fig. 7(f) are corresponding correcting

results. It is obvious that the proposed method gets a better result when there are bad circular arcs in the distorted image.

Fig. 8 presents the results for another real image of 720x515, from a publicly available database [20], which has many short straight line in the distorted image, see Fig. 8(a) and Fig. 8(b). The estimation results of distortion parameter in Fig. 8(c) indicate that the varying of distortion parameter of the proposed method is smaller. And the estimation location of distortion centers of the proposed method is more concentrated than that of the Bukhari-Dailey method. Hence, the proposed method is more robust.

Table 1 shows some quantitative results which illustrate the time costs for different number of arcs. Minimum pixel number of arcs, average number of arcs and average CPU time for Fig. 8(a) are computed using the Bukhari-Dailey method and the

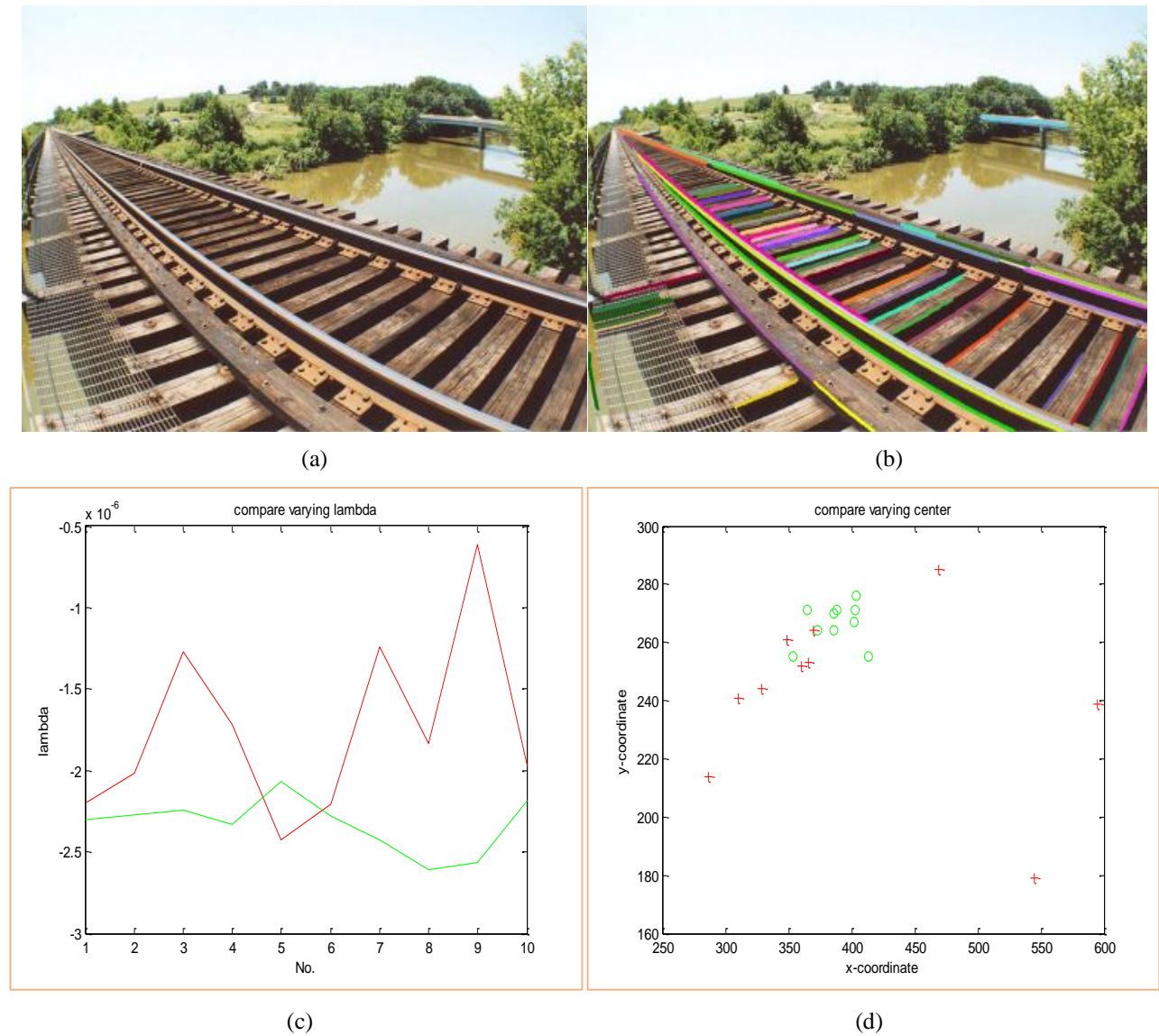


Figure 8. Compare the variation of distortion parameter in 10 runs:(a) Original image, (b) detected candidate arcs, (c)magnitude of lambda (the green line are results of the proposed method while the red line are that of the Bukhari-Dailey method ). (d) location of distortion center(the green symbol “o” presented the distortion center using the proposed method and the red symbol “+” are the distortion center using the Bukhari-Dailey method)



TABLE I.

<b>Fig. 8(a)</b>	Minimum length of arcs	180	160	140	120	100	80
<b>The Bukhari-Dailey</b>	Average number of arcs	12.2	17.7	23.9	37.3	59	110.6
	Average CPU time(MS)	1357.3	1835.5	3495.5	4480.3	4504.5	4691.9
<b>The Proposed method</b>	Average number of arcs	11.2	18.4	23.2	37.9	61.8	109.4
	Average CPU time (MS)	7.3	7.6	7.8	10.9	15.2	19.8

proposed method respectively. Each test runs 10 times, and the CPU time includes selecting good circular arc and estimating the distortion parameter. As observed, it is obviously the proposed method just takes a little time which is far less than that of the Bukhari-Dailey method. It is primarily because the proposed method is simple and non-iterative, while the Bukhari-Dailey method requires an iterative process. With increasing number of the circular arcs, the time costs of the proposed method increases but are still very small, while that of Bukhari-Dailey method increases fast. The proposed method of choosing good circular arcs is faster than that of the Bukhari-Dailey.

## VI. CONCLUSION

In this paper, a method to identify good circular arcs from the detected distorted curves which are vital to line-based method of estimation of distortion parameters, especially the fully automatic ones. It is based on that the constant values are the same for all the good circular arcs. The algorithm is simple, robust and non-iterative. Once good circular arcs are determined, the parameters of distortion can be estimated accurately. Therefore, the proposed method recognizing the good circular arcs and only using the good circular arcs to correct the radial distortion can eliminate the interference of the bad curves. We have presented a variety of experiments on synthetic and real images which show that the proposed method allows removing the radial distortion automatically and robustly.

## REFERENCES

- [1] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No.11, pp. 1330-1334, 2000.
- [2] Z. Kukulova and T. Pajdla, "A minimal solution to radial distortion autocalibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 11, pp. 2410-2422, 2011.
- [3] T. A. Clarke and G. J. Fryer, "The development of camera calibration methods and models," *The Photogrammetric Record*, Vol.16, No. 91, pp. 51-66, 1998.
- [4] A. W. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, 2001, pp. I-25-I-32.
- [5] R. Strand and E. Hayman, "Correcting Radial Distortion by Circle Fitting," *British Machine Vision Conference*, 2005.
- [6] F. Bukhari and M. N. Dailey, "Robust radial distortion estimation from a single image," *The 6th International Conference on Advances in Visual Computing*, Las Vegas, NV, USA, 2010, pp. 11-20.
- [7] G. P. Stein, "Lens distortion calibration using point correspondences," *The 1997 Conference on Computer Vision and Pattern Recognition*, 1997, pp. 602-608.
- [8] R. Hartley and S. B. Kang, "Parameter-free radial distortion correction with center of distortion estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 8, pp. 1309-1321, 2007.
- [9] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, Vol. 3, No. 4, pp. 323-344, 1987.
- [10] A. Wang, T. Qiu, and L. Shao, "A simple method of radial distortion correction with centre of distortion estimation," *Journal of Mathematical Imaging and Vision*, Vol. 35, No. 3, pp. 165-172, 2009.
- [11] F. Bukhari and M. N. Dailey, "Automatic radial distortion estimation from a single image," *Journal of Mathematical Imaging and Vision*, Vol. 45, No. 1, pp. 31-45, 2013.
- [12] H. Wildenauer and B. Micusik, "Closed form solution for radial distortion estimation from a single vanishing point," *British Machine Vision Conference*, 2013.
- [13] C. B. Duane, "Close-range camera calibration," *Photogrammetric Engineering*, Vol. 37, No. 8, pp. 855-866, 1971.
- [14] F. Devernay and O. Faugeras, "Straight lines have to be straight," *Machine Vision and Applications*, Vol. 13, No. 1, pp. 14-24, 2001.
- [15] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 88, No.6, pp. 679-698, 1986.
- [16] N. Chernov, *Circular and linear regression: Fitting circles and lines by least squares*. Monographs on Statistic and Applied Probability, CRC Press/Taylor & Francis, Boca Ration, 2010.
- [17] N. Chernov and C. Lesort, "Least squares fitting of circles, *Journal of Mathematical Imaging and Vision*," vol. 23, No. 3, pp. 239-252, 2005.
- [18] P. V. C. Hough, "Method and means for recognizing complex patterns," *U.S. Patent*, 3069654, 1962.
- [19] File: Usno-amc.jpg. <http://commons.wikimedia.org/wiki/File:Usno-amc.jpg>.
- [20] R. Oleson: Full-circle examples. [http://rick\\_oleson.tripod.com/index-105.html](http://rick_oleson.tripod.com/index-105.html).

# Differential Evolutionary Algorithm Based on Multiple Vector Metrics for Semantic Similarity Assessment in Continuous Vector Space

Yuanyuan Cai, Wei Lu, Xiaoping Che, Kailun Shi  
School of Software Engineering  
Beijing Jiaotong University  
Beijing, China

Email: {yycai, luwei, xpche, shikailun}@bjtu.edu.cn

## Abstract

*Automatic service discovery in heterogeneous environment is becoming one of the challenging problems for applications in semantic web, wireless sensor networks, etc. It is mainly due to the lack of accurate semantic similarity assessment between profile attributes of user request and web services. Generally, lexical semantic resources consist of corpus and domain knowledge. To improve similarity measures in terms of accuracy, various hybrid methods have been proposed to either integrate different semantic resources or combine various similarity methods based on a single resource. In this work, we propose a novel approach which combines vector similarity metrics in a continuous vector space to evaluate semantic similarity between concepts. This approach takes advantage of both corpus and knowledge base by constructing diverse vector space models. Specifically, we use differential evolutionary (DE) algorithm which is an powerful population-based stochastic search strategy for obtaining optimal value of the combination. Our approach has been validated against a variety of vector-based similarity approaches on multiple benchmark datasets. The empirical results demonstrate that our approach outperforms the state-of-the-art approaches. The results also indicate the continuous vectors are efficient for evaluating semantic similarity, since they have outstanding expressiveness to latent semantic features of words. Moreover, the robustness of our approach is presented by the steady measure results under different hyper-parameters of neural network.*

*Keywords-differential evolutionary; semantic similarity; continuous vector space; vector similarity metrics*

## 1 Introduction

The vast number of information and heterogeneous resources distributed on the web have made the semantic analysis and semantic interoperability more challenging, especially in some fields such as semantic web, natural language processing (NLP) and social network. Semantic similarity measurement for concepts, which measures the degree of similarity or dissimilarity between two concepts, enables the precise service discovery and information inquiry. For example, a user who is querying the *bank* service can obtain results consisting of the words *deposit* and *interests* rather than *slope* and *river*. Hence, the semantic similarity measurement for concepts has been an attractive research content and also an important component in the related applications, such as automated service discovery [27], text classification [15] and emotion mining [4].

Existing approaches to measuring semantic similarity between concepts can be divided into corpus-based and knowledge-based approaches in terms of the semantic resources available. Corpus-based approaches primarily map a given corpus into a vector space [37] to compute the similarity between lexicon vectors. The words close together in the vector space tend to be semantically similar or occur in similar contexts. In these approaches, semantic features of words derive from the distributional properties of words in statistic corpus, which consist of the distribution and the frequency of lexical context. Corpus-based approaches are limited to the distributional VSM based on lexical co-occurrence statistics in corpus, since the vectors are modeled by “bag of words” which scratch the surface of words without reflecting sufficient semantic association of words. To explicitly decode implied semantic information from corpus into the distributional vector space, some related works leverage dimension reduction technologies such as Latent Semantic Analysis (LSA) [12], Latent Dirichlet Allocation (LDA) [8] and distributional information simi-



larity [20]. However, these works still use discrete vectors which lack the powerful expression capability of latent semantic and syntactic information. Therefore, rare and polysomous words are often poorly estimated.

Knowledge-based approaches take advantage of pre-existing knowledge bases such as thesauri and WordNet ontology [24] to measure semantic similarity. In terms of semantic properties used in semantic computations, WordNet-based measures can be roughly classified into path-based, information content (IC)-based, feature-based and hybrid measures. The path-based measures and the IC-based measures mainly exploit the path difference and IC difference between concepts, while the feature-based measures rely on constructing concept vectors based on intrinsic properties of concepts and computing the similarity between vectors. As the feature-based approaches, gloss overlaps [6] and the cosine similarity between gloss vectors [28] can be directly used to measure semantic similarity. Liu et al. took local densities as the intrinsic properties of concepts and computed the cosine similarity of concept vectors for measuring semantic similarity between concepts [21].

To capture different aspects of semantic similarity between concepts, a variety of combined strategies are proposed, in terms of different measures and heterogeneous semantic resources. Yih and Qazvinian incorporated different vector measurements based on the heterogeneous lexical sources such as Wikipedia, web search engine, thesaurus and WordNet [35]. Alves et al. proposed a regression function where lexical similarity, syntactic similarity, semantic similarity and distributional similarity are input as independent variables [2]. Similarly, Bär et al. introduced a linear regression model integrating multiple content similarity values at the aspects of string, semantic, structure, etc [7]. Chaves-González and MartíNez-Gil combined WordNet-based semantic similarity measures using a meta-heuristic algorithm to find a optimized solution [9]. Mihalcea et al. focused on the corpus-based cosine similarity and WordNet-based similarity [5]. In their approach, the distributed word vectors were linearly aggregated into diverse level representation related to phrase, sentence and paragraph. These hybrid approaches integrate different vector space models or different similarity methods with a single resource. However, few measures focus on the combination of vector similarity metrics for semantic similarity measurement.

This work contributes to integrating various vector similarity metrics such as cosine distance and Euclidean distance using a differential evolutionary (DE) algorithm. We assume that different metrics can induce varying degrees of semantic similarity between concepts. E.g., the cosine distance determinates the angle distance between two vectors (directional similarity) in the vector space, whereas the Euclidean distance evaluates straight-line distance between

two vectors (magnitude similarity). Hence, in this work, fine-grained semantic similarities from different aspects are provided by a variety of metrics to optimize the similarity measurement. We use a DE algorithm to combine different vector-based similarity measures which rely on either corpus or WordNet. Furthermore, inspired by the application of distributed word representation from deep learning [22], we measure semantic similarity in the continuous vector space which reveals latent semantics. In addition, we conduct an additional experimentation to study the effects of various similarity metrics and hyper-parameters of neural network on the results of semantic comparison, since some systematical investigations indicated that the vector-based similarity approaches highly depend on the quality of VSM construction.

The rest of this paper is organized as follows: the related works are presented in Section 2. The problem and similarity metrics we used in this work are summarized in Section 3. Our methodology and experimental results on several evaluation criteria are discussed in Section 4. Conclusions and future work are given in Section 5.

## 2 Related works

Previous semantic similarity measures take advantage of domain ontology or corpus to compute the similarity between words. Ontology-based measures focus on exploring structure properties of ontology in semantic similarity computation, while corpus-based measures are based on the similarity of discrete vectors and improved by the technologies of dimensionality reduction. As an alternative of discrete vector model, the continuous word representation derived from deep learning has significantly benefited the vector-based semantic similarity measurement recently [36]. Continuous word representation, namely distributed word embedding, is a real-valued vector whose each dimension represents a latent semantic feature of words. In the continuous VSM, the words are encoded within a low-dimension vectors via unsupervised neural network training, which can better understand the significance and syntactic structure of words in a corpus text. With the powerful expressiveness of latent semantics, the continuous VSMs contribute to the outstanding performances of semantic disambiguation and analogy reasoning as well as other tasks [18]. Specially, according to Mikolov [23], the continuous word representations are independent across languages in terms of analogy relationship of word pairs.

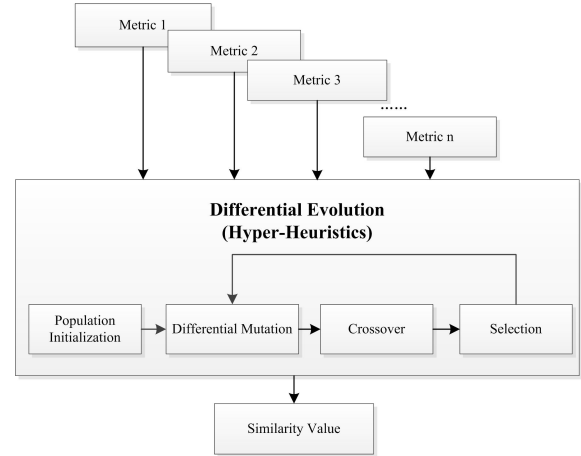
Similarity metric or distance metric is an important part of vector similarity measures. When evaluating the semantic similarity of concepts, most works perform with a single computational metric, such as vector overlaps, cosine distance and Euclidean distance [1]. Based on the cosine similarity of vectors, Faruqi and Dyer evaluated the concept

similarity and the diversity of continuous word embeddings derived from different natural networks [13]. Pennington et al. learned distributed vectors from unsupervised global log-bilinear regression model with matrix factorization, and took the cosine value of the vectors as concept similarity [29]. However, a single metric could not capture all the aspects of semantic similarity and suit all types of input data. In addition, some works focus on studying effects of various similarity metrics on semantic similarity measurement. These studies contribute to the integration of different computational metrics. As an instance, Kiela and Clark studied the computational metric, data source, dimensionality reduction strategy, term weighting scheme and the parameters of vectors including window size and feature granularity in similarity tasks [19]. However, their evaluation concentrated on the distributional vector models. As regards continuous distributed vector model, Hill et al. demonstrated that the larger training windows work better for measuring similarity of abstract words than concrete words, and vice versa [17]. Chen et al. found that the lower dimensions of word embeddings significantly drop the accuracy of the classifiers across all the publicly available word embeddings [10]. Inspired by these work, we focus on the continuous vector space. Differing from other studies on similarity measures, we take advantage of vector similarity metrics.

Instead of proposing a new vector similarity metric, our study aims to improve the evaluation results obtained in single metric by combining multiple vector similarity functions. Hence, we propose a combination strategy to assessing semantic similarity based on the differential evolutionary (DE) algorithm. The algorithm of DE [34] is a population-based stochastic search strategy for solving global optimization problems. It derives from evolutionary algorithm (EA) and has multiple variants according to the strategy for generation of new candidate members [11, 26]. These variants have been proved applicable for continuous function optimization in a large number of research domains such as heat transfer [3].

### 3 Semantic similarity measurement based on differential evolutionary algorithm

In this section, we define the problem and research object on similarity evaluation, and describe the proposed hybrid measure which incorporates the heterogeneous similarity metrics for vector via differential evolutionary algorithm. In this work, the differential evolution algorithm is used for addressing the problem of the incorporation of various metrics, since it offers competitive solutions for evaluating the different aspects of semantic similarity. It iteratively assigns each similarity metric a specific weight. Fig. 1 illustrates the DE algorithm in our work. It performs with the similarity values provided by various vector-based metric-



**Figure 1. Illustrative workflow of the differential evolution (DE) algorithm.**

s. All the metrics evenly contribute to evaluate the degree of semantic similarity between two concepts at the beginning of the differential evolution. Then the metric which provides the most similar results to the human judgement is offered the highest weight after automatic evolution process consists of initialization, mutation, crossover and selection.

#### 3.1 Problem definition

There are given two concepts  $C_1$  and  $C_2$ , the problem is to determine the degree of their semantic similarity. The vector-based semantic similarity calculation not only depends on the quality of vector but also involves the vector distance metric. Hence, we adopt various similarity metrics with low-dimensional continuous vectors. Each metric focuses on different lexical semantic relations between concepts consist of synonymy, hypernymy, hyponym and even antonymy, as well as co-occurrence relation [38], which respectively provide a certain degree of semantic similarity. Based on the combination strategy, we realize the integration of different metrics to capture semantic relations and determine semantic similarity between vectors. Formally, we define the two concepts as vector  $X$  and  $Y$ .

#### 3.2 Vector similarity metrics

There exist numbers of metrics for vector similarity computation. Table 1 summarizes the similarity metrics explored in our work for two concept vectors. The first column indicates the general type of metrics and the second column gives their formalized definition. And the third column presents a brief explanation of the metrics.

From the perspective of vector direction, cosine metric measures how similar two vectors are. On the contrary, Eu-

**Table 1. Similarity metrics between n-dimensional vector  $X$  and  $Y$ .**

Similarity measure	Function definition	Description
Cosine	$\frac{X \cdot Y}{ X  \cdot  Y }$	Cosine similarity computes cosine value of the vectorial angle in vector space
Euclidean	$\frac{1}{1+ X-Y }$	Euclidean distance evaluates the absolute length of the line segment which connects the terminal points of two vector
Manhattan	$\frac{1}{1+\sum_{i=1}^n  X_i - Y_i }$	Also known as the Cityblock distance, which is only possible to travel directly along pixel grid lines when going from one pixel to the other
Chebyshev	$\frac{1}{1+\max_i  X_i - Y_i }$	Chebyshev distance evaluates the maximum of the absolute distances in each dimension of vectors
Correlation	$\frac{(X-\bar{X}) \cdot (Y-\bar{Y})}{ X  \cdot  Y }$	Correlation distance evaluates the degree of linear correlation between vectors
Tanimoto	$\frac{X \cdot Y}{ X  +  Y  - X \cdot Y}$	Tanimoto similarity measures the degree of shared features between two vectors

clidean distance which is sensitive to the absolute difference of individual numerical features provides us the magnitude of the difference between two vectors. Other distance measures such as Manhattan distance and Chebyshev distance evaluate the sum or the maximum of differences on the features of vectors. Correlation distance contributes to revealing the linear association between two vectors. The Tanimoto coefficient is used to measure matching degree of the features between two vectors.

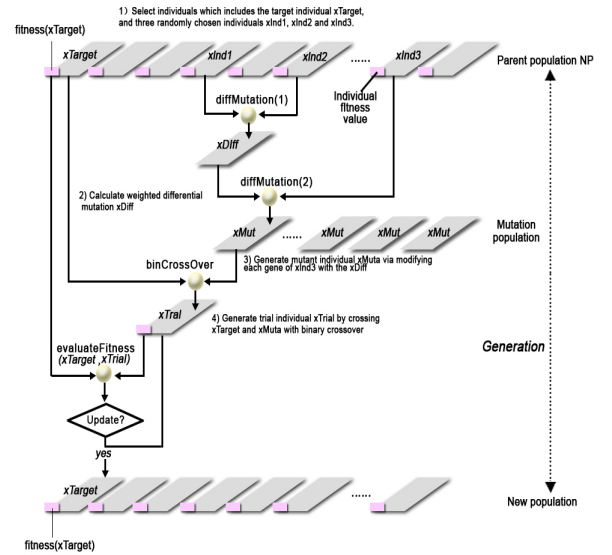
### 3.3 Differential evolution algorithm

The hyper-heuristics DE algorithm works as a solution for the global optimization of the combination of vector metrics. It holds a population with the size of  $NP$  and defines each member of the population as a candidate solution that a vector of weighting coefficients. In the evolution process, new individuals are generated due to the difference between the chosen individuals (see Fig. 2). Table 2 profiles the individuals in population, where each dimensionality of the individuals represents a similarity metric  $M_k$  whose similarity result weighted by and the coefficient  $w(M_k)$ .

**Table 2. Individual profile.**

Metric 1	Metric 2	Metric 3	...	Metric N
$w(M_1)$	$w(M_2)$	$w(M_3)$	...	$w(M_N)$

One individual in a population is represented as a vector like  $\vec{I} = [w(M_1), w(M_2), \dots, w(M_N)]$  where each element  $w(M_k) \in [\text{MIN}, \text{MAX}]$  is a real number. To some extent, the task of DE algorithm is a search for a vector  $\vec{I}^*$  to optimize the objective function of the given problem. DE performs the evolution of NP individuals  $\vec{I}_{ik}$  with N dimensions ( $i=1, 2, \dots, NP$ ;  $k=1, 2, \dots, N$ ) in a vast search space. It

**Figure 2. Profile of the rand/1/bin differential evolution (DE) algorithm.**

consists of three basic operations that mutation, crossover and selection. Among the existing variants of DE algorithm, we choose the strategy *rand/1/bin* [34] in this work, in terms of the scheme of mutation and crossover as well as selection. The notation *rand/1/bin* indicates how the mutation and crossover operators work. That is, the DE algorithm selects individuals at random, then adopts binomial crossover (bin) and a unique difference vector (/1/) to generate the mutation of the random individual (rand) in the parent population. Fig. 2 illustrates the *rand/1/bin* strategy, and its configures are detailed in Section 4. This strategy

starts with the random generation of the population through assigning a random weights to each gene of the individual.

The main process of the DE algorithm initiates after calculating fitness for the whole population. DE algorithm selects the individuals consisting of target individual  $\vec{I}_t$ , and three randomly chosen individuals  $\vec{I}_{r1}$ ,  $\vec{I}_{r2}$ ,  $\vec{I}_{r3}$ . Then the weighted differential mutation  $\delta\vec{I}$  is calculated according to the expression that  $\delta\vec{I} \leftarrow F \cdot (\vec{I}_{r1} - \vec{I}_{r2})$ , where the mutation factor  $F$  scales the effect of the pairs of chosen individuals on the calculation of the mutation value. Then the mutant individual  $\vec{I}_m$  is produced via modifying each gene of  $\vec{I}_{r3}$  with the  $\delta\vec{I}$ , which is formalized as  $\vec{I}_m \leftarrow \vec{I}_{r3} + \delta\vec{I}$ . DE exploits binary crossover operation to obtain the trial individual and so that keeps the diversity of population. The trial individual vector  $\vec{I}_{tr}$  is generated via crossing  $\vec{I}_t$  and  $\vec{I}_m$  with the binary crossover scheme as the expression that  $\vec{I}_{tr} \leftarrow \text{binCrossover}(\vec{I}_t, \vec{I}_m, P)$ . The crossover probability,  $P \in [0, 1]$ , controls the effect of parents on the generation of offsprings. The process of DE algorithm is ended at comparing  $\vec{I}_t$  against the new individual  $\vec{I}_{tr}$  in terms of fitness and determining whether replace it with the  $\vec{I}_{tr}$  accordingly. The better individual will be saved in the position of original  $\vec{I}_t$  which is described as,

$$\tilde{\mathbf{I}}_t = \begin{cases} \vec{I}_{tr} & \text{if } f(\vec{I}_{tr}) \leq f(\vec{I}_t) \\ \vec{I}_t & \text{otherwise} \end{cases} \quad (1)$$

where  $f(\vec{I})$  is the objective function of vector  $\vec{I}$  to be minimized. For each individual, the above process is repeated parallelly with the max iteration (i.e., generations) of G during evolution. Finally, the individual  $\vec{I}^*$  with the best fitness is returned as the optimized result of the DE algorithm.

In this work, Pearson correlation coefficient [33] is taken as the fitness of each individual to evaluate the quality of each individual. This correlation,  $\rho_{xy}$ , is calculated as follows:

$$\rho_{xy} = \frac{\text{Cov}(x, y)}{\sqrt{D(x)}\sqrt{D(y)}} = \frac{E(xy) - E(x)E(y)}{\sqrt{D(x)}\sqrt{D(y)}} \quad (2)$$

where the numerator is covariance of variable  $x$  and variable  $y$ ,  $E(x)$  refers to the expectation of variable  $x$ . The denominator is the product of the standard deviations of variable  $x$  and variable  $y$ .

The correlation is used to compare computational results of various similarity methods with the human judgments for word pairs. It is a floating point value between -1 (extreme negative correlation) and +1 (extreme positive correlation) which indicates the degree of linear dependence between the computational methods and human opinion. The nearer the value of correlation is to any of the extreme values (-1 or +1), the stronger is the correlation between the variables and the higher is the performance of the method. If the Pearson correlation of a method gets near to 0, it indicates the method results in poor performance. In terms

of Pearson correlation, we compare the performance of our combination strategy and other methods for semantic similarity measurement. Besides, the parameters of DE algorithm consisting of  $NP$ ,  $F$ ,  $P$  and  $G$  need to be fixed as constants. In the following Section 4, we give the concrete values conducted in our experiments.

## 4 Experiments and results

In this section we demonstrate the experiments which conduct the combination of various vector similarity metrics on different benchmarks and discuss the results. In order to measure semantic similarity between concepts in continuous feature vector space, we learn continuous distributed concept vectors by training neural network model.

### 4.1 Methodology

We use the tool word2vec<sup>1</sup> to implement CBOW neural network model since its effectiveness and simplicity. We formalize a refined vocabulary as  $V$ . For a word  $w$  in  $V$ , the CBOW model averages the set of its context  $c_t = \{w_{t-k}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+k}\}$  which consists of  $k$  words to the left and right at projection layer. The training objective of CBOW is to maximize the log probability of the target word  $w$ , formally,

$$Obj = \frac{1}{T} \sum_{t=1}^T \sum_{(-k \leq j \leq k, j \neq 0)} \log p(w_t | w_{t+j}) \quad (3)$$

where  $w_t$  is a given target word,  $w_{t+j}$  is the surrounding words in context, and  $k$  is the context window size. The inner summation spans from -k to +k to compute the log probability of correctly predicting the central word  $w_t$  given all the context words  $w_{t+j}$ . The conditional probability  $p(w_t | w_{t+j})$  is defined in the following softmax function:

$$p(w_t | w_{t+j}) = \frac{\exp(\text{vec}'(w_t)^\top \text{vec}(w_{t+j}))}{\sum_{w=1}^V \exp(\text{vec}'(w)^\top \text{vec}(w_{t+j}))} \quad (4)$$

where  $\text{vec}(w)$  and  $\text{vec}'(w)$  refer to the input vector and output vector of word  $w$ .

Three unlabeled corpora are fed as input of the CBOW model, including Wikipedia<sup>2</sup> (3,483,254 word types and  $10^9$  tokens), BNC<sup>3</sup> (346,592 word types,  $10^7$  tokens) and Brown Corpus<sup>4</sup> (14,783 types,  $10^5$  tokens). Once the input corpora are available, pre-processing of corpus is conducted firstly, including data cleaning, tokenization, abbreviation removal, stop-word removal, etc. Named entities and

<sup>1</sup><http://word2vec.googlecode.com/svn/trunk/>

<sup>2</sup><http://dumps.wikimedia.org/enwiki/20140903/>

<sup>3</sup><http://www.ota.ox.ac.uk/desc/2554>

<sup>4</sup>[http://nltk.googlecode.com/svn/trunk/nltk\\_data/index.xml/](http://nltk.googlecode.com/svn/trunk/nltk_data/index.xml/)

special terms that contain uppercase letters are taken as abbreviations and removed from the corpus since they may significantly impact the training precision. In most studies on NLP, stop words are considered useful for handling syntax information, such as progressive relationship and transition relation. However, we consider that this work mainly focuses on the expression ability of word vectors, whereas stop words which occur frequently disturb the sense-group of sentences due to they have little real meaning. Therefore, the stop words are removed to avoid over-training and make the remaining lexical meaning clearly represented. Therefore, we get a vocabulary of over 0.8 billion tokens after processing the raw corpora in advance.

Based on the generated continuous vectors, different similarity results between concepts are computed by various vector metrics. These results are input into the DE algorithm to obtain an optimized value. Table 3 summarizes the configuration settings of the DE algorithm in this work, which provides more competitive results based on the *rand/1/bin* strategy than other variants of DE algorithm<sup>5</sup>.

**Table 3. Optimal parameters.**

Parameter	Value
Population size, $NP$	10*N
Mutation factor, $F$	0.5
Crossover probability, $P$	0.1
Max generations, $G$	1000
Max, Min	+10, -10

## 4.2 Benchmark datasets

7 benchmarks are conducted in our experiments for results verify, including WS-353, WS-sim, WS-rel, RG-65, MC-30, YP-130 and MTurk-287. These datasets are widely used in word similarity studies to compare the semantic similarity methods with human judgements. The **WS-353** dataset [14] contains of 353 word pairs of English words with similarity rating by humans. The degree of similarity of each pair is assessed on a scale of 0-10 by 13-16 human subjects, where the mean is used as the final score. WS-353 was further divided into two subsets [1] that similar pairs (**WS-sim**) and related pairs (**WS-rel**) in terms of the degree of similarity between word pairs. The **RG-65** [32] contains 65 pairs of words assessed on a 0-4 scale by 51 human subjects. The **MC-30** dataset [25], 30 word pairs from RG-65, are reassessed by 38 subjects and a small portion of WS-353. Although these datasets contain overlapping word pairs, their similarity scores are different since they are given by different human judges in the diverse experiments.

<sup>5</sup><http://www1.icsi.berkeley.edu/storn/code.html>

In addition, the WS-353 contains the words within various part-of-speeches whereas others merely contain nouns. We also evaluate our model on the **MTurk-287** benchmark [31] which consists of 287 word pairs evaluated by 10 subjects on a scale of 1 to 5 for each and crowdsourced from Amazon Mechanical Turk. To specifically emphasize the effect on verb, the **YP-130** dataset [39] that contains 130 verb pairs was created and judged by human as well.

## 4.3 Result discussion

We conduct three kinds of experiments to evaluate the proposed approach described in Section 3. Firstly, we compare our DE-based approach with two different sets of similarity metric (vector-based metrics and WordNet-based metrics) on the RG-65 benchmark dataset. Next, we implement our approach on multiple benchmark datasets. Finally, we investigate the parameters of CBOW model which include dimension and window size to demonstrate the robustness of our approach and the effect of these parameters on the similarity measurement of concepts.

### 4.3.1 Experiments with different metrics on RG dataset

Our approach is compared against two sets of metrics on RG dataset. Firstly, we evaluate various similarity metrics based on the continuous vectors extracted from corpus. Table 4 presents the Pearson correlation between the computational results and human ratings on RG dataset, where the top lists the performance of individual similarity metrics and the bottom shows the result of our DE-based approach. The experimental results demonstrate that our approach improves the accuracy of existing corpus-based vector similarity metrics and achieves a result of 0.894 with the dimension of 500 and window size of 7. While the result of cosine metric which is considered as most effective in most of previous literatures achieves 0.805.

**Table 4. Pearson correlation between computational vector metrics and human ratings on RG dataset.**

Similarity method	Correlation
Chebyshev	0.660
Tanimoto	0.785
Manhattan	0.788
Euclidean	0.794
Correlation	0.805
Cosine	0.805
<b>Ours (6 metrics)</b>	<b>0.894</b>

In order to take full advantage of the semantic information from both WordNet and corpus, we further integrate two additional gloss-based methods into the DE strategy. As mentioned in Section 1, WordNet-based similarity methods contain four categories that path-based, IC-based, feature-based and hybrid methods. In this experiment, we focus on the feature-based methods where the feature properties of WordNet are used to construct concept vectors. Therefore, beside the vector metrics presented in Table 4, our approach combines extended gloss overlap [6] and cosine similarity of gloss vector [28]. For comparison, we choose some hybrid methods which tend to be superior to other WordNet-based methods since they adequately employ various semantic information from WordNet.

**Table 5. Pearson correlation between WordNet-based similarity methods and human ratings on RG dataset.**

Similarity method	Correlation
Extended gloss overlap[6]	0.350
Gloss vector[28]	0.797
Liu [21]	0.810
Pirro[30]	0.872
Gao[16]	0.885
<b>Ours (8 metrics)</b>	<b>0.903</b>

Table 5 indicates that our DE-based combination better aligns with human judgement in contrast with the individual feature-based methods and hybrid methods in the studies related to WordNet. The results also show that continuous vectors learned from corpus seem to supply more precise semantic than the gloss vector extracted from WordNet. Moreover, although having relatively high performance as well as our approach, the hybrid method proposed by Gao [16] requires parameters to be settled.

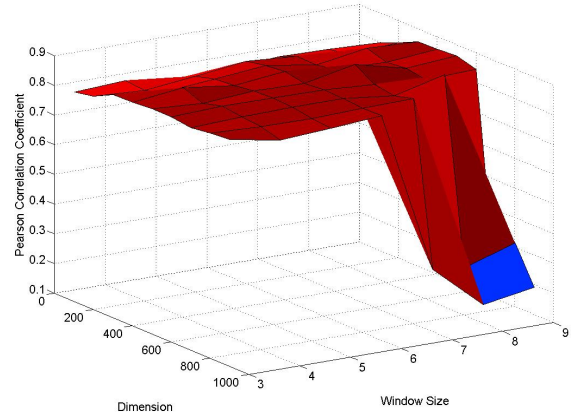
#### 4.3.2 Experiments with different datasets

Table 6 summarizes the results of state-of-the-art similarity methods on 7 benchmarks, such as WS-353, YP-130, etc. While outperforming our approach on the WS-353, WS-sim and WS-rel dataset, the approach of Yih [35] needs more heterogeneous semantic sources (web search, Wikipedia, Bloomsbury and WordNet) to turn out averaged cosine similarity score. Based on both web corpus and WordNet, Agirre et al. [1] conduct a supervised combination of several similarity methods, which obtains a higher result than ours on the RG-65 dataset. However, their approach has to train a SVM to turn parameters and needs a mass of training data. Unlike some approaches [31] that perform well on some datasets but poorly on others, our approach is more robust

since it holds high performance on the additional MTurk-287 dataset and YP-130 dataset. In order to further evaluate the quality of the continuous real-value vectors learned via neural network training, we perform our DE-based approach across different parameter settings.

#### 4.3.3 Experiments with different parameters of CBOW model

In our study, the quality of concept vector depends on the hyper-parameters of CBOW model. To further indicate the robustness of our approach, we estimate the window size of training and the dimensionality size of vector. The window size is set to 3 up to 9. The dimensionality which reveals the feature granularity of vectors ranges from 100 to 900 with a step length of 100. According to the results on RG



**Figure 3. The performances of our method under different settings of dimensionality and window.**

dataset shown in Fig. 3, our approach keeps steady across different dimensionality and window sizes, which implies the continuous vector representations used in our approach remain stable expression of semantic features. However, the curved surface suffers a drastic decline near the point with dimensionality 900 and window 9 due to the overfitting resulted by excessive training.

## 5 Conclusions

This work proposes a differential evolutionary based approach to measure the semantic similarity in a continuous vector space. The differential evolutionary algorithm is used to leverage the results derived from different vector-based similarity metrics and find a optimal combination strategy of the metrics. The continuous vectors which reveal



**Table 6. The performance of state-of-the-art methods on multiple datasets.**

Similarity method	RG-65	MC-30	WS-353	WS-sim	WS-rel	MTurk-287	YP-130
Yih [35]	0.89	0.89	<b>0.81</b>	<b>0.87</b>	<b>0.77</b>	0.68	NA*
Radinsky [31]	NA	NA	0.80	NA	NA	0.63	NA
Agirre [1]	<b>0.96</b>	0.92	0.78	0.83	0.72	NA	NA
Ours (8 metrics)	0.90	<b>0.93</b>	0.76	0.83	0.70	<b>0.71</b>	<b>0.75</b>

\* N/A means empty value.

latent semantic features of words are explored to improve the vector similarity computation. The experiment results demonstrate our combined approach outperforms other similarity methods on multiple benchmark datasets and has the robustness under different training parameters. In future works, we will present an WordNet-constrained neural network model to further improve the quality the distributed vectors and the accuracy of the semantic similarity measurement between concepts.

## Acknowledgment

This work is supported in part by National Natural Science Foundation of China (No.61272353, 61370128, and 61428201), Program for New Century Excellent Talents in University (NCET-13-0659), Beijing Higher Education Young Elite Teacher Project (YETP0583).

## References

- [1] E. Agirre, E. Alfonseca, K. Hall, J. Kravalova, M. Paşca, and A. Soroa. A study on similarity and relatedness using distributional and wordnet-based approaches. In *Proceedings of the conference of The North American Chapter of the Association for Computational Linguistics - Human Language Technologies*, pages 19–27, June 2009.
- [2] A. O. Alves, A. Ferrugento, M. Lourenço, and F. Rodrigues. Asap: Automatic semantic alignment for phrases. In *the 8th International Workshop on Semantic Evaluation (SemEval)*, pages 104–108, Dublin, Ireland, August 2014.
- [3] B. V. Babu and S. A. Munawar. Differential evolution strategies for optimal design of shell-and-tube heat exchangers. *Chemical Engineering Science*, 62:3720–3739, 2007.
- [4] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)*, pages 2200–2204. European Language Resources Association, May 2010.
- [5] C. Banea, D. Chen, R. Mihalcea, C. Cardie, and J. Wiebe. Simcompass: Using deep learning word embeddings to assess cross-level similarity. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval)*, pages 560–565, August 2014.
- [6] S. Banerjee and T. Pedersen. Extended gloss overlaps as a measure of semantic relatedness. In *International Joint*

- Conference on Artificial Intelligence*, volume 3, pages 805–810, August 2003.
- [7] D. Bär, C. Biemann, I. Gurevych, and T. Zesch. Ukp: Computing semantic textual similarity by combining multiple content similarity measures. In *Proceedings of the 1st Joint Conference on Lexical and Computational Semantics*, pages 435–440. Association for Computational Linguistics, June 2012.
- [8] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3:993–1022, January 2003.
- [9] J. M. Chaves-González and J. MartíNez-Gil. Evolutionary algorithm based on different semantic similarity functions for synonym recognition in the biomedical domain. *Knowledge-Based Systems*, 37:62–69, January 2013.
- [10] Y. Chen, B. Perozzi, R. Al-Rfou, and S. Skiena. The expressive power of word embeddings. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, June 2013.
- [11] S. Das and P. N. Suganthan. Differential evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation*, 15(1):4–31, 2011.
- [12] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*, 41(6):391–407, September 1990.
- [13] M. Faruqi and C. Dyer. Community evaluation and exchange of word vectors at wordvectors.org. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, June 2014.
- [14] L. Finkelstein, E. Gabilovich, Y. Matias, E. Rivlin, Z. Solan, G. Wolfman, and E. Ruppín. Placing search in context: The concept revisited. *ACM Transactions on Information Systems*, 20(1):116–131, January 2002.
- [15] G. Forman. An extensive empirical study of feature selection metrics for text classification. *Journal of machine learning research*, 3:1289–1305, March 2003.
- [16] J.-B. Gao, B.-W. Zhang, and X.-H. Chen. A wordnet-based semantic similarity measurement combining edge-counting and information content theory. *Engineering Applications of Artificial Intelligence*, 39:80–88, 2015.
- [17] F. Hill, D. Kiela, and A. Korhonen. Concreteness and corpora: A theoretical and practical analysis. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 75–83. Association for Computational Linguistics, August 2013.

- [18] E. H. Huang, R. Socher, C. D. Manning, and A. Y. Ng. Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 873–882, Jeju Island, South Korea, July 2012.
- [19] D. Kiela and S. Clark. A systematic study of semantic vector space model parameters. In *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC) at EACL*, pages 21–30. Association for Computational Linguistics, April 2014.
- [20] D. Lin. An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning (ICML)*, pages 296–304, Madison, Wisconsin, USA, July 1998.
- [21] H. Z. Liu, H. Bao, and D. Xu. Concept vector for semantic similarity and relatedness based on wordnet structure. *The Journal of Systems and Software*, 85(2):370–381, August 2012.
- [22] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In *the International Conference on Learning Representations (ICLR) Workshop*, Scottsdale, Arizona, USA, May 2013.
- [23] T. Mikolov, W. tau Yih, and G. Zweig. Linguistic regularities in continuous space word representations. In *the conference of North American Chapter of the Association for Computational Linguistics - Human Language Technologies*, pages 746–751, Atlanta, GA, USA, June 2013.
- [24] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, November 1995.
- [25] G. A. Miller and W. G. Charles. Contextual correlates of semantic similarity. *Language and cognitive processes*, 6(1):1–28, 1991.
- [26] E. nn Mezura-Montes, J. Velázquez-Reyes, and C. A. C. Coello. A comparative study of differential evolution variants for global optimization. In *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, GECCO '06*, pages 485–492, New York, NY, USA, 2006. ACM.
- [27] A. V. Paliwal, B. Shafiq, J. Vaidya, H. Xiong, and N. Adam. Semantics-based automated service discovery. *IEEE Transactions on Services Computing*, 5(2):260–275, May 2012.
- [28] S. Patwardhan and T. Pedersen. Using wordnet-based context vectors to estimate the semantic relatedness of concepts. In *Proceedings of the EACL Workshop on Making Sense of Sense-Bringing Computational Linguistics and Psycholinguistics Together*, pages 1–8. Citeseer, March 2006.
- [29] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. Association for Computational Linguistics, October 2014.
- [30] G. Pirró. A semantic similarity metric combining features and intrinsic information content. *Data & Knowledge Engineering*, 68(11):1289–1308, 2009.
- [31] K. Radinsky, E. Agichtein, E. Gabrilovich, and S. Markovitch. A word at a time: computing word relatedness using temporal semantic analysis. In *Proceedings of the 20th international conference on world wide web*, pages 337–346. ACM, March 2011.
- [32] H. Rubenstein and J. B. Goodenough. Contextual correlates of synonymy. *Communications of the ACM*, 8(10):627–633, October 1965.
- [33] J. S. Simonoff. *Smoothing methods in statistics*. Springer, 1996.
- [34] R. Storn and K. Price. Differential evolution- a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.
- [35] W. tau Yih and V. Qazvinian. Measuring word relatedness using heterogeneous vector space models. In *the conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies*, pages 616–620. Association for Computational Linguistics, June 2012.
- [36] 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*, pages 384–394. Association for Computational Linguistics, July 2010.
- [37] P. D. Turney and P. Pantel. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1):141–188, February 2010.
- [38] E. M. Voorhees. Query expansion using lexical-semantic relations. In *Proceedings of the 17th Annual International ACM SIGIR Conference*, pages 61–69. Springer, January 1994.
- [39] D. Yang and D. M. Powers. Verb similarity on the taxonomy of wordnet. In *Proceedings of the 3rd International WordNet Conference (GWC)*, pages 121–128, January 2006.

# PhysQSR: Improving Reasoning in Three Dimensions and Time With Image Processing and Physics

Nathan W. Eløe<sup>1</sup> and Jennifer L. Leopold<sup>2</sup>

<sup>1</sup>Department of Computer Science and Information Systems, Northwest Missouri State University,  
nathan.eloe@gmail.com

<sup>2</sup>Department of Computer Science, Missouri University of Science and Technology,  
leopoldj@mst.edu

## Abstract

*Qualitative Spatial Reasoning (QSR) is an exceptionally powerful tool in the fields of computer cognition and automated computer reasoning. Recent results have shown the potential feasibility of pairing image processing techniques with basic principles of physics that humans inherently understand in order to allow the computer to extrapolate additional information about the environment in which it exists. Initial results showed that, while using the tenets of conservation of mass, conservation of energy, and inertia allowed the computer to gain more information than was initially apparent, noise in the perceived input data resulted in the software erroneously reasoning about the state of the system. Hence improving the image processing techniques used in analyzing the data should ameliorate the errors in reasoning. In this paper, the authors investigate this claim, and present a system that allows a more precise and correct computational view of the environment.*

**Keywords** Qualitative Spatial Reasoning, Image Processing, Object Segmentation

## 1 Introduction

Human perception of an environment is incredibly complex to replicate in a computational system. Alone, the fact that the perception of the environment is not necessarily consistent between two different observers leads to the conclusion that the results reported by a computer may be verified by one person and invalidated by another. The best that can be done is to have the computer deduce all

objectively correct information: that which can be mathematically proven to be true.

Previous results [9] have shown that, in the absence of full three-dimensional knowledge of an environment, it is feasible to use stereoscopic images to obtain information about relative depths and shapes of objects from the computer's observation point. Augmenting this data with laws that govern the physical world allows the computer to learn more about the world. These physical laws were chosen based on their relation to how humans perceive the world. For example, the laws of Conservation of Mass and Energy can directly be seen as a mathematical way to describe Object Permanence [2].

In the authors' previous work, these results showed that noise in the input data led directly to incorrect or even impossible output. The observations of objects when fully visible were used to extrapolate positions of the objects when obscured; noise in the observed positions had a deleterious effect on the extrapolated positions, resulting in potential incorrect assumptions about the system. Herein we discuss the effects a more robust object segmentation algorithm has on the quality of data used in computational reasoning.

## 2 Background and Related Work

### 2.1 Image Processing and Disparity

Image processing is an important field in computer and robotic vision. A significant amount of research in this area has been devoted to finding computationally efficient algorithms; images are inherently two-dimensional, which implies that most naive algorithms are at best  $O(m \times n)$  in their

computational complexity for an  $m \times n$  image. The persistence of high resolution images (full high definition already common and 4k resolution is beginning to emerge) means that these algorithms will be computationally expensive. Many image formats are 4-channel (giving an  $m \times n \times 4$  data structure size to hold RGBA or HSVA (Hue Saturation Value Alpha) information, two popular information formats), which only serves to increase the amount of computation needed for a single image.

Disparity [3, 10, 7] and the parallax effect are two concepts exploited in image processing to mimic human perception of depth; objects closer to the observer appear larger than more distant objects. Thus, by determining the parallax between occurrences of an object in each of a pair of stereo images, the relative distance from the cameras to the object can be determined. Disparity also has been used to estimate the motion of objects [7]. It is an invaluable tool in determining spatial information from multiple observations of the same scene.

Object segmentation in image processing is a task that has a large number of established approaches. These methods range from thresholding mechanisms like Otsu's method [11], to clustering algorithms, to region-growing methods. For the purposes of this research, a combination of an edge detection mechanism and a heavy modification of a watershed style algorithm is used to segment objects when all objects are visible. A more explicit description of this object segmentation method is included in Section 3.

## 2.2 Qualitative Spatial Reasoning (QSR)

Qualitative Spatial Reasoning (QSR) has varying applications in Geographic Information Systems (GIS), visual programming language semantics, and digital image analysis [13, 6, 12, 15]. Systems for spatial reasoning over a set of objects have evolved in both expressive power and complexity. The design of each system focuses on certain criteria, including efficiency of computation, ease of human comprehension, and expressive power.

The spatial reasoning system chosen for this investigation is VRCC-3D+ [16], an expansion and implementation of the RCC-3D [1] system designed by Albath et al. As opposed to other RCC systems (most of which have no implementation), the relations in VRCC-3D+ express both connectivity (in 3D) and obscuration. Obscuration will change from viewpoint to viewpoint, but connectivity is a global property that can be used to discern new information at every perspective in the system.

For this work, the authors focus on the obscuration element of the VRCC-3D+ relation. The connectivity portion of the relation will become important as the system is expanded to handle an arbitrary number of cameras and vantage points. VRCC-3D+ identifies four basic kinds of

obscuration: no obscuration (*nObs*), partial obscuration (*pObs*), complete obscuration (*cObs*), and equal obscuration (*eObs*). The system further breaks each base obscuration into four different classes: regular obscuration (object A obscures object B), converse obscuration (object A is obscured by object B), equal obscuration (object A and object B obscure each other equally), and mutual obscuration (objects A and B obscure each other). At this point in the investigation, this further classification is unimportant; it only matters if obscuration is present between two objects, not which object is being obscured.

## 2.3 PhysQSR: QSR with Image Processing and Physics

The system described in [9] has been enhanced to improve the ease with which the system can be used and extended. A thorough explanation of the enhancements made can be found in [8]. Briefly, the implementation of PhysQSR has moved to using a detector system on each set of frame pairs. These detectors are hot-pluggable based on what kind of information they require to run. Initially, only a motion detector and a collision detector had been implemented. This paper describes the construction of an object detector that is used to more accurately track objects as they move through the environment.

The basic process of analyzing the video footage remains mostly unchanged; while becoming more modular through the use of detectors, each frame pair goes through image analysis (e.g. disparity calculations), obscuration analysis, and object analysis (object position, either calculated or estimated). Previously, this process was very procedural; the shift to detectors allows this to become a more modular process. A pair of frames is grabbed (from each of the left and right cameras), and passed through detectors. These detectors perform tasks such as object detection, obscuration detection, collision detection, and motion detection.

## 3 Object Segmentation in PhysQSR

The initial exploration of augmenting QSR with Physics and Image Processing used a very simplistic object segmentation mechanism. The initial testing video input (see Figure 1) were generated in Blender. Because the objects were known to be spheres of two well defined colors, a naive HSV (Hue Saturation Value) filtering/thresholding method was used to generate object masks. This method was satisfactory for one video pair, but when analyzing a more complex scene in which the objects collided, changes in lighting caused a large amount of noise in the calculated and estimated positions of the objects (Figure 2). While the general observed motion of the objects is correct, the noise frequently creates problems where, in some cases, the

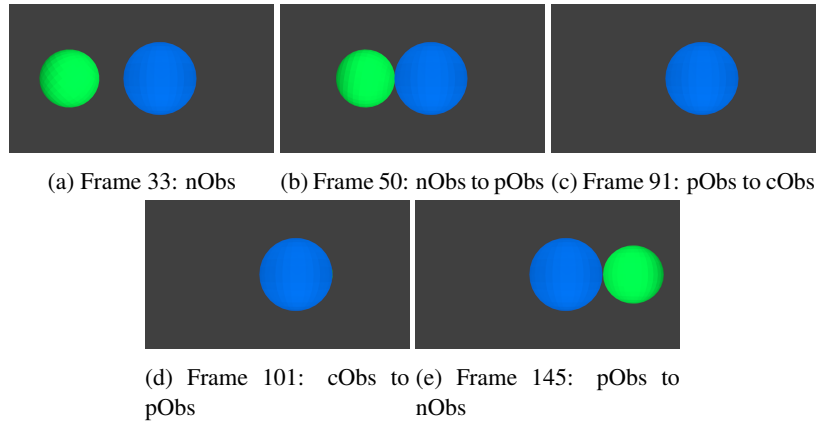


Figure 1: Images from analyzed video: as seen from the left camera. The green sphere is further from the cameras than the blue sphere, and as such appears smaller.

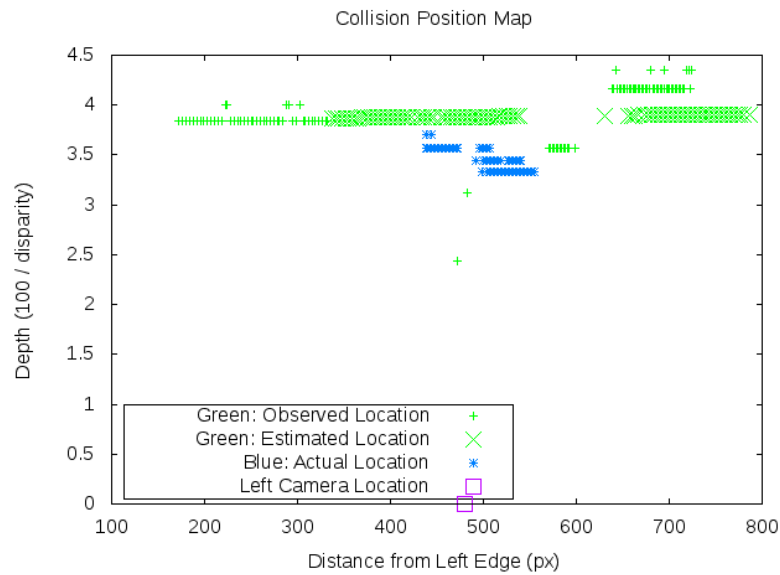


Figure 2: Observed and estimated object positions when collision is present in input.

green ball (which, in the video, passes behind the blue ball) is determined to be in front of the blue ball, a physical impossibility.

A significant portion of the noise in this data can be directly attributed to the effects of lighting in the scene. Figure 3 clearly demonstrates how lighting in the scene can artificially introduce noise into the data when using HSV segmentation. Note that the shape of the mask between the two frames shown changes with the shape of the shadows cast by the blue object. It is interesting to note that these particular images were taken when the lower bound of the saturation value for the mask was set to 20. When set to 100 (the lower bound used in experimentation), both the leading and trailing edge of the masks were practically unrecognizable, which in turn would lead to an unpredictable lateral position in the image. As the position of the trailing edge along the horizontal is used to calculate both the distance from the left edge of the image and the disparity, the oscillating behavior in calculated positions that were initially observed in [9] is understandable.

As such, a more powerful mechanism of object segmentation was deemed necessary. Instead of using HSV masks solely, a combination of edge detection and a watershed like method was used to segment the objects. First, a Canny edge detector [5] was applied to the frame. The thresholds for the hysteresis procedure were set at 150 and 200, and all other parameters were left as the default values provided by the OpenCV [4] Python bindings. The edges resulting from the detector were used to identify closed contours. These contours were then used with a modified Watershed Transform algorithm [14]. In this method the areas enclosed by contours were flooded, designating individual objects. In this way, the objects were segmented. All that remains is to determine which object is which; in this case, the color of the objects is the defining feature, so the HSV masks are still used to identify objects uniquely.

The identified centers of these objects are then used as the locations of the objects for the calculation of disparity and distance from image edge. By using edge detection before the object is filtered based on color, the effects of lighting can be ameliorated. Figure 4 shows how this method provides a much more precise view of where objects are positioned in the image.

## 4 Results

First, consider the effect this has on the observed position of the objects. Using the new object detection mechanism results in the positions observed in Figure 5. The observed positions demonstrate significantly less noise than the original, as shown in Figure 6. The lateral shift in the position (distance from image edge) is easy to explain. In the previous object tracking method, the trailing edge of the

object was used to determine where it existed. Using the new watershed/edge tracking method, the center of mass of the observed object is used. As such, the lateral shift is only a change in the point of interest on the object; the fact that the magnitude of the shift is roughly constant throughout the calculated positions corroborates this.

Applying this new object segmentation algorithm to the video originally analyzed in [9] yields some interesting results. Figure 7 shows the calculated positions of the objects in the scene when applying the legacy HSV masking object segmentation algorithm, modified only to take into account the transition of the code to detectors instead of the original procedural mechanism for video analysis, and identical experimentation parameters to the video representing collision. Transitioning to the edge detection segmentation algorithm yielded the positions seen in Figure 8.

It is fascinating that this transition exhibits both improvement and regression. The most marked improvement can be seen in the portion of the scene where the blue ball obscures the green ball. There was significant noise in the data such that the horizontal position of the green ball was calculated to be inhabiting the blue ball's space. Not only that, but at times, the green ball was calculated to be in front of the blue ball, a physical impossibility. However, the noise in the calculated depth at the locations where the green ball was fully visible exhibits less noise overall when using the HSV segmentation than when using the Watershed method. In both cases, however, the noise in the calculated depth cause the fit polynomial to trend closer to the camera. This leads the investigators to believe that perhaps using a polynomial fit line that accounts for all previously known information for the location of the object may not be an optimal way to extrapolate the position of the object.

## 5 Conclusions

Earlier work showed the feasibility of using image processing techniques on limited visual data (stereoscopic images of a scene) in conjunction with physical properties to learn more about the environment. The identified weakness in the system was primarily the overly simplistic object segmentation mechanism, as it led to noise in the data that caused inconsistencies in the computer system's conclusions. Before expanding the system to work on real-life video input (in real time), the effects of this limitation were deemed critical to examine before continuing with the implementation of this system.

As theorized, using a more robust image tracking system immediately provides better data for the computer to reason with. The system is more robust with respect to variations in lighting and color shifts, and the result is immediately cleaner data. Any regressions noted in the transition did not lead to reasoning that was more incorrect. Indeed, the



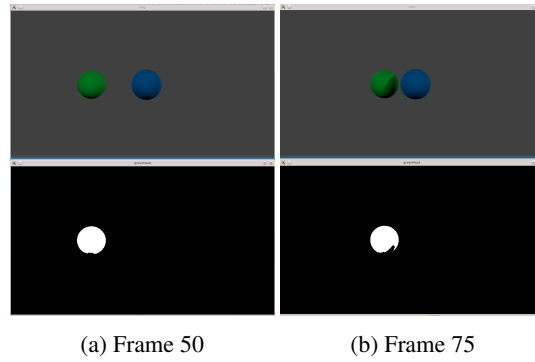


Figure 3: The effect of lighting with HSV object segmentation. Only the segmentation of the green object is shown. The white area in each bottom image represents the area marked by the segmentation as belonging to the green object.

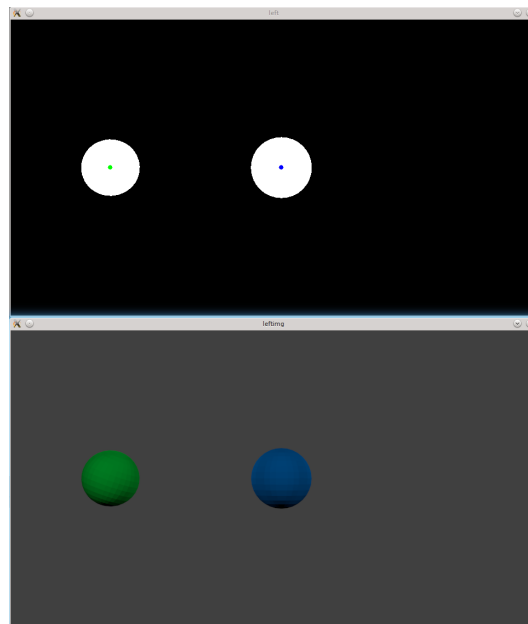


Figure 4: Tracking objects with edge detection and modified Watershed. Each white area is a segmented object in the original image, which are identified with a colored dot corresponding to the original object's color.

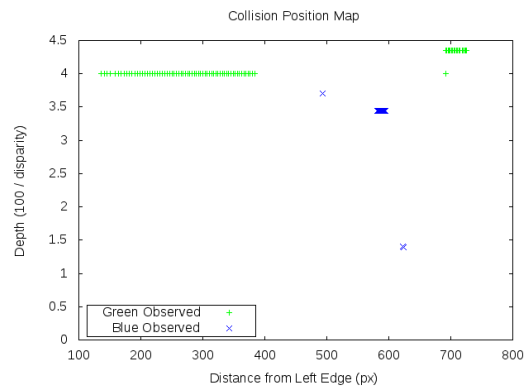


Figure 5: Observed and estimated object positions when collision present in input, new object segmentation.

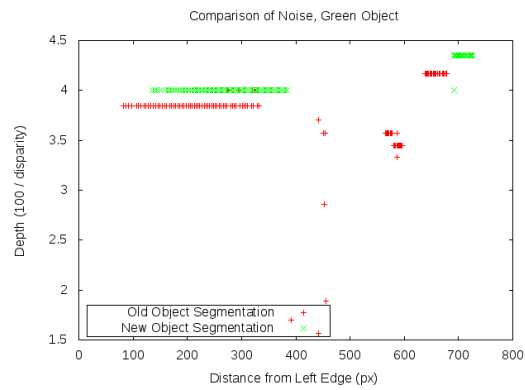


Figure 6: Comparison of noise with new and old object segmentation methods.

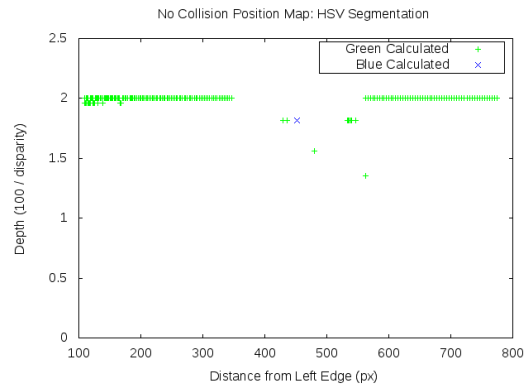


Figure 7: Comparison of noise with new and old object segmentation methods.

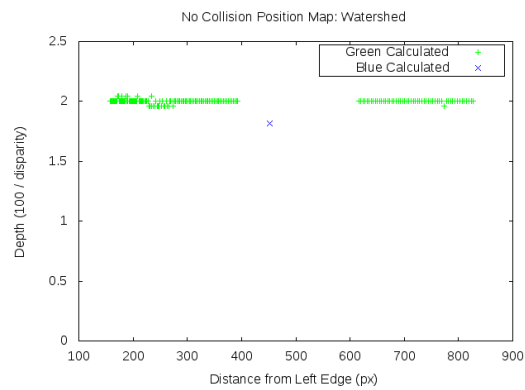


Figure 8: Comparison of noise with new and old object segmentation methods.

additional noise observed in the scenario with no collision resulted in very little to no change in the estimated behavior of the object when it could not be fully observed. This suggests that the system is more sensitive to the magnitude of the noise in the data instead of the amount of the noise.

## 6 Future Work

Immediate future work on this project will focus on two areas: using the system to analyze real world video data, and further suppression of noise in the input data. Analyzing any real world video data will introduce more noise into the system as imperfect synchronization, mismatched camera sensors, and data transmission over data buses causes degradation. As such, noise suppression will be important to the future success of this project.

Furthermore, investigation of the mechanism used to extrapolate the position of occluded objects will be required. The presence of any noise in the data causes the estimated positions to exhibit erroneous behavior. As it may be impossible to completely eliminate all noise from input data (or even be able to identify noise), more robust object position estimation algorithms will provide a more powerful system that produces reliable results.

## References

- [1] 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*, pages 74–79, Nov. 2010.
- [2] R. Baillargeon. Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants. *Cognition*, 23(1):21–41, 1986.
- [3] S. T. Barnard. Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(4):333–340, July 1980.
- [4] G. Bradski. *Dr. Dobb's Journal of Software Tools*, 2000.
- [5] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698, 1986.
- [6] A. G. Cohn, B. Bennett, J. Gooday, and M. M. Gotts. Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. *Geoinformatica*, 1(3):275–316, Oct. 1997.
- [7] D. Demirdjian and T. Darrell. Motion Estimation from Disparity Images. In *Proceedings of the Eighth IEEE International Conference on Computer Vision ICCV*, volume 1, pages 213–218, 2001.
- [8] N. Eløe, J. L. Leopold, and C. L. Sabharwal. Spatial Temporal Reasoning Using Image Processing, Physics, and Qualitative Spatial Reasoning. *International Journal of Software Engineering and Knowledge Engineering*. In Review.
- [9] N. Eløe, J. L. Leopold, C. L. Sabharwal, and Z. Yin. Spatial Temporal Reasoning Using QSR, Physics, and Image Processing. In *Distributed Multimedia Systems '13*, pages 14–19, Aug 2013.
- [10] K. Mühlmann, D. Maier, J. Hesser, and R. Männer. Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation. *International Journal of Computer Vision*, 47(1–3):79–88, Apr. 2002.
- [11] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.
- [12] D. A. Randell and M. Witkowski. Using Occlusion Calculi to Interpret Digital Images. In *Proceedings of the 17th European Conference on Artificial Intelligence*, pages 432–436, Aug. 2006.
- [13] J. Renz. *Qualitative Spatial Reasoning with Topological Information*. Springer-Verlag, 2002.
- [14] J. B. Roerdink and A. Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta informaticae*, 41(1):187–228, 2000.
- [15] P. Rost, L. Hotz, and S. von Riegen. Supporting Mobile Robot's Tasks through Qualitative Spatial Reasoning. In *Proceedings of the Ninth International Conference on Informatics in Control, Automation and Robotics (ICINCO 2012)*, volume 2, pages 394–399, 2012.
- [16] C. L. Sabharwal, J. L. Leopold, and N. Eløe. 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))*, pages 307–311, Aug. 2011.

# TEco: an integration model to augment the Web with a trust area for inter-pares interactions.

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

*We propose an integrated and modular model called TEco. It is a Web-based trust area in which, through the integration of various systems, users interact with a greater degree of trust. In particular, users: own a Trusted Digital Identity to authenticate keeping anonymity (when required); establish Inter-Pares Interactions based on contracted agreements and knowing each other's reputation; can be the owners of the information they produce and protect their privacy. We discuss the feasibility of the model, the compatibility with the current Web and the things to do for putting it into practice.*

**Keywords:** *Trust area, TEco, Privacy, Anonymity, Single Sign-On, Trust and Reputation, User-centric.*

## 1 Introduction

Despite its enormous success and indisputable usefulness, the Web is not exempt from problems that should be addressed to make it even more secure and reliable. In fact, some issues, such as the uncertainty of the identities, the almost complete lack of privacy and of guarantees on the reliability of the counterparts, i.e. the lack of trust among people, may limit its potential development [1]. Other issues are the lack of control and ownership of the information regarding a person or a company; the lack of specific information about service providers (e.g., reliability, quality, punctuality, etc. I.e. their reputation); the exploitation of anonymity to perform malicious actions [2].

In recent years many studies have focused on the development of new protocols and methodologies to allow unambiguous identification of the user, the ability to keep anonymity, to protect privacy and to authenticate once to access many services (Single Sign-On) [3, 4]. However, the aforementioned issues were almost always addressed individually.

Generally, organizations (i.e. companies, academics, etc.) authenticate users and grant them roles through Identity Providers. The Identity Management systems, instead, adopting a user-centric paradigm, rely on the user rather than on the service provider for the control of digital identities [5, 4, 6]. The control of their own identities allows users to decide which information to share with others and under which conditions.

Based on the above reasons, our objective is the design of a comprehensive framework aimed at providing a trust area in the Web that combines the online and offline world smoothly and seamlessly, including the best solutions in a single model.

Our integrated and modular model is called *TEco* (acronym of Trust Ecosystem). Here, *ecosystem* means (see [7, 8]) a loosely coupled, domain clustered environment where each species conserves the environment, is proactive and responsive for its own benefits. In our case, species are the entities (e.g., users and online services) which preserve the environment and comply with fixed rules, are proactive and responsive as each of them, using a reward-punishment mechanism (feedback), contribute to the success of the system and, consequently, to their own benefit.

Digital ecosystems, as emphasized in [9], “can play the role of a unification ‘umbrella’ over significant, challenging and visionary computing approaches that emerge in parallel”. In this sense TEco will also act as a “field of comparison” and facilitate scientific communication in the sector.

The remainder of this paper is organized as follows: the next section summarizes some works related to ours; in Section 3 we introduce TEco; in Section 4 we discuss some practical issues related to the implementation of the model; lastly, in Section 5 we draw some conclusions and outline future work.

## 2 Related work

In the literature, we can find non-integrated solutions for:

- Identity Management systems (IdMs) and Single Sign-On (SSO);
- Trust and Reputation Management systems (TRMs);
- Anonymity and privacy protection;

Nevertheless, to the best of our knowledge, there are no studies in the literature that have faced all the aforementioned problems in a comprehensive and systematic view.

In their survey on IdMs [6], Torres et al. point out that the use of IdMs, which also enable SSO, may help to solve the new challenges related to security and privacy protection. Conversely, authors in [10] highlight some of their weaknesses, especially the impossibility for a user to decide which personal information to share with every service provider or to obtain information on their reliability. To address these issues, the authors propose techniques to integrate IdMs with Reputation Management Systems, which provide information on the past behavior of the service providers [2].

Other studies, driven by the emerging of new technologies, are focused on the next generation Internet, termed **Future Internet** [11]. Nevertheless, they do not provide a common view on what the Future Internet is and mostly consider its network infrastructure, termed **Future Network** [6].

Despite the importance of many of the problems faced, such as infrastructural ones, we believe that even other aspects deserve attention, such as the relationship between digital identities and reputation and other little investigated sectors: the respect of user rights and the possibility for users to keep control of their data.

It is worth noting that Microsoft introduced a *Trust Ecosystem*, more narrowly defined as an environment that engenders trust and accountability between people and businesses [12]. In that system, users have several *Windows CardSpace* to access a service provider without having to authenticate [13]. Despite the similar name, our model includes more features and differs substantially from the one introduced by Microsoft, as we will show in the following.

### 3 Trust Ecosystem

The TEco system can be accessed by users (individuals and legal persons) and online services. All of them are considered as “entities” which interact with each other “at par” with no distinction between client and server, user and provider, services and humans. Following a user-centric paradigm, TEco was built by integrating different innovative systems to provide the following features:

- **Trusted Digital Identity:** every digital identity corresponds to an individual who is identified with “certainty” still keeping anonymity and privacy;
- **Content Management:** users are the owners of the information they produce and can manage such information autonomously;
- **Reputation Management:** it is possible to obtain reliable and updated reputation information about all users;
- **Interaction Agreement:** interactions are always based on a contract agreed between the parties, that have equal bargaining power;

The coexistence of these features makes TEco a trust area. In fact, users can mutually trust, as they are all identifiable, their reputation is known and while interacting, they can bargain conditions with law effectiveness. Furthermore, depending on their needs and the demands of others, users can decide which information to disseminate, protecting their privacy or maintaining complete anonymity.

#### 3.1 Trusted Digital Identity

In the current Web, each entity has a **Digital Identity**, which can be defined as the digital representation of the information known about a specific individual or organization [3]. In TEco, in addition, each digital identity corresponds to an entity in the offline world whose identity is verified with certainty. To this end, an entity is required to register at TEco providing its own unique identifier. For individuals, this can be the identifier used by the governments for tracking their citizens as the National Identification Number. For corporate bodies (companies, organizations, associations, etc.) it can be their VAT number. To complete the registration to TEco it is therefore necessary that an entity proves to be the owner of the provided identifier. For instance, individuals could complete the registration at the Municipal Registry Office and legal persons at the Registry of Companies. Once the registration is completed, the entity will possess a **Trusted digital Identity (TId)** in TEco. The TId will correspond to an account associated with all the information available of the requester and its identifier. The requester is the only owner of the access credentials for that account. As the TEco is only accessible to certified digital identities, online services must have a TId too. In this case, the owner of the domain name must certify the association between provider’s TId and URL of the service (used as its unique identifier). The whole registration process is handled by the **Identity Management Systems (IDMs)** which assign and manage identities and belong to a **Federated Identity Management (FIdM)**. In general, a Federation can be defined as the set of agreements, policies,

standards and technologies to achieve its objective [3, 5]. The purpose of FIdM, is to allow entities belonging to different IdMs to be identified from all others, regardless the used authentication system (e.g. Kantara Initiative <sup>1</sup>, Liberty Alliance <sup>2</sup>, Shibboleth [14], Kerberos [15], etc.).

The IdMs are the only ones to know the association between offline world entities and their TId. For this reason, an entity must possess a **Web Alter Ego (WAE)** to interact in TEco, i.e. an alternative identity to present itself to others. Based on his/her needs, an individual can create different WAEs (e.g., as a researcher, as a chess player, etc.), choosing for each WAE which information to show among those associated to his/her own TId. Each WAE is completely independent from the others and is seen by counterparts as a separate entity. In fact, a counterpart cannot relate all the WAEs belonging to the same identity. This safeguards an entity's privacy, since it can use one of its WAEs without worrying that its true identity is revealed or that one of its WAEs is associated to others (in the following, we will see how this can be guaranteed through the use of temporary identifiers).

A registered entity to access TEco must logon at the IdM which manages its TId through the planned identification procedure (e.g. based on username/password, biometric data, etc.). Then, the entity receives from the IdM the list of all its own temporary identifiers, referred to as *TempWAEs*, specifically generated. Each of them uniquely identifies a specific WAE and allows the entity to interact within TEco without logging on to any specific service. This enables an SSO authentication. While the entity is "connected" to TEco, the *TempWAEs* are regenerated and sent back by the IdM to the entity periodically according to predefined security criteria or upon an entity's explicit request. It should be noted that the regeneration of the identifiers does not require a new logon. The *TempWAEs*' validity expires as the entity "disconnects", by logging out after an indefinite time. Besides identity management, TEco also provides a reputation system based on several **Reputation Management Systems (RMSs)**, each responsible to collect, aggregate and disseminate data on the reputation of the entities [2]. The RMSs belong to a **Federated Reputation Management System (FRMS)**, which manages their interaction. The integration of FRMS and FIdM provides the users with a high level of mutual trust. In fact, they are encouraged to take appropriate behavior because they know they are identified with certainty and their past behavior is known to all. The greater mutual trust increases the *social capital*, intended as the richness of the interactions between members, which itself affects the reputation system encouraging an active and honest participation and thus increasing its effectiveness [16]. The FIdM assigns each entity a refer-

ence RMS which is also involved in managing the reputation of all its WAEs. At the end of an interaction, an entity is required to leave an anonymous feedback on the counterparts to its reference RMS. The latter, in turn, according to the times and rules set by the federation, sends the feedback to the reference RMS of the recipient entity. An entity can request the reputation of the other entities to its own reference RMS, which obtains it through the federation. It is worth recalling that, being independent, each WAE has its own reputation independently from others. Since a good reputation requires time, this reduces the proliferation of WAEs (see "newbies" in [17]).

### 3.2 Inter Pares Interaction

In the current Web, the users share information and request services by establishing interactions. In TEco, any interaction is always based on a contract agreed between the parties. We refer to the interaction as **Inter Pares Interaction** (in the following referred to only as "*TEco Interaction*") and to the contract as **Negotiated Interaction Agreement** (in the following referred to only as "negotiated agreement"). The negotiated agreement is composed of two parts: the first, preliminary and fixed, contains the principles and general conditions that oversee any interaction in TEco (e.g., to respect owners' constraints on the data, not to maliciously alter reputation, etc.). The second part is subject to negotiation and contains a list of **Agreement's Terms** (in the following referred to only as "term"), i.e. constraints and preferences established in a formal language, that the parties agree to comply with. If some constraints in the *negotiated agreement* are not respected by one of the parts, as terms of a contract with the force of law, can be asserted in judicial offices. Since, as stated in [17], an entity interacts with the others in a given context and assuming a specific role, an **Interaction Context/Role (ICR)** in the *negotiated agreement* will also be mandatorily negotiated. For instance, the consultation of a website is a typical "interaction" between end-user and website owner, where the ICR for the user is "content visualization/reader".

The *negotiated agreement* is established through a phase of **Negotiation of the Agreement** (in the following referred to only as "negotiation"), in which each party sends the other its contract proposal, called **Interaction Agreement** (in the following referred to only as "agreement"), composed of the list of *terms* that a party intends to include in the second part. During negotiation, each *term* can be modified or accepted to reach the *negotiated agreement* in its final form. If all parties agree, the *negotiated agreement* can be changed at any time. Clearly, an entity that does not conclude the phase of *negotiation* can not take part in the interaction.

The *terms, agreements e negotiated agreements* are

<sup>1</sup>[www.kantarainitiative.org](http://www.kantarainitiative.org)

<sup>2</sup>[www.projectliberty.org](http://www.projectliberty.org)



defined through a formal language. This allows the entity to participate to the *TEco interaction* through a **Web Agent**, which suggests or takes decisions on the basis of its acquired experience (self-learning), on the type of entity (e.g., individual) and on the context/role (e.g., e-learning/instructor). For instance, in the case of an individual, human intervention may be required during bargaining. In the context/role e-commerce/seller, the *negotiation* phase of the seller is automatically handled by the Web Agent and the human intervention is not required, unless expressly prescribed by the seller. It should also be pointed out that an *agreement* can be defined by including only standard *terms* that are stored in an archive at the FIDM which also manages an archive of default *agreements*. A new *agreement* is created by choosing the *terms* from a list of standard ones through an appropriate GUI. In order to simplify the *negotiation* phase, while logging on to TEco, the entities receive (similarly to WAEs) lists of predefined *terms* and *agreements* from the FIDM. This way, they can set an *agreement* for each WAE choosing it from the default ones. For instance, the entity could select a WAE called “*Web surfing*” associated to an *agreement* called “*High Privacy*”, requiring counterparties not to request private information such as the *home address*.

Figure 1 shows how two entities establish a *TEco interaction* (the schema can be extended to more than two entities). We use the following notation: **TempWAE** for the temporary identifier of a entity’s Web Alter Ego; **PermWAE** for the permanent one. It is worth recalling that permanent identifiers are never disclosed to entities. As shown in the figure, an interaction is composed of the following steps:

- Step 1. *Ann* requests an interaction to a service provider (SP) providing the WAE (*TempWAEa*) with which she intends to identify herself and her *agreement*;
- Step 2. The SP requests to its reference RMS (*RMSp* in the figure) the reputation associated to *TempWAEa* and related to the context/role (*ICRa*) provided by *Ann* in her *agreement*;
- Step 3. *RMSp* requests to the FIDM the permanent identifier (*PermWAEa*) associated to *TempWAEa*;
- Step 4. Once obtained the *PermWAEa*, *RMSp* checks if it has the reputation associated to *PermWAEa* in the context/role *ICRa*. If not, *RMSp* requests it to the FRMS.
- Step 5. Then, *RMSp* returns to SP the reputation of *TempWAEa* in *ICRa*. It is worth noting that SP receives the reputation of *Ann*’s WAE knowing only her temporary identifier.

- Step 6. SP decides, based on the received reputation, whether to accept TEco interaction request. If so, SP sends *Ann* the WAE with which it intends to interact (*TempWAEp*) and its own *agreement*. Otherwise, it sends a message of rejection and abandons the interaction.
- Steps 7-11. The same actions performed in Steps 2 - 6 on *SP*’s side are now executed on *Ann*’s side. Step 11 opens the negotiation phase which ends with the negotiated agreement.

The *TEco interaction* was schematically shown in sequential steps in order to facilitate the exposure but, actually, some steps may be performed in parallel (e.g., the negotiation phase). As previously mentioned, the parties may express a feedback on counterparts at the end of the interaction.

Nevertheless, to prevent malicious attacks and improve the reputation system, TEco adopts some important countermeasures already described in [17]. After the negotiated agreement is established and before starting a *TEco interaction*, an entity’s Web Agent sends to its *reference RMS* a list of pairs ICR-WAE, each referred to an entity it is going to interact with. The RMS, in turn, sends back to the entity the **Interaction Token** with which the interaction will be uniquely identified for a predetermined time interval. This token allows the RMS to accept only feedbacks to and from entities that indeed took part to the interaction and to make sure that the interaction indeed took place. Therefore, every feedback must include the *interaction token* and the TempWAEs of both the judging and the judged entities. This assures that a feedback is expressed once for each entity involved in an interaction.

Furthermore, the RMS could release encrypted reputation data with date and time of encryption. This ensures data integrity and authenticity. This also speeds up the reputation retrieval, since entities may store the encrypted reputation data and share them with counterparts without querying the FRMS. Counterparts may decide whether to query the FRMS on the basis of both the certification date and the reputation of the entity (too old data may be untrustworthy). We remark that the presence of a contract having the force of law strongly discourages illicit practices, as they can be prosecuted.

### 3.3 Content Management Framework

As mentioned before, one of the objectives of TEco is to ensure that the entities are direct owners of the information they produce. To this aim, an important role is played by the **Content Management Framework (CMF)**, which manages all data (text, multimedia, WAE’s attributes, etc.) related to the entities. Whenever a new content is created

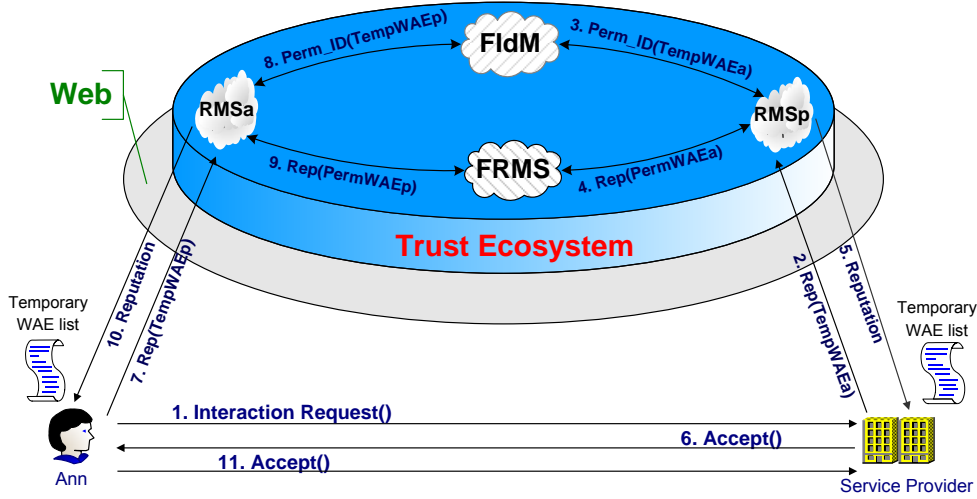


Figure 1: Interaction between entities.

at a service provider, the CMF automatically creates a link between the content and the producing entity. These data are physically stored at the SP or in personal cloud storage systems, local hard disks, etc. In any case, they are property of the entity that produced them, which can decide which *access rights* to grant to other entities (reading, modification, deletion, duplication, disclosure, etc.). Furthermore, the time validity of each *privilege* can also be established. Therefore, contrary to what normally happens in the current Web where the users are deprived of these rights, in TEco the users have management and responsibility of their own data. The CMF ensures that the rights set by the content owner, with any changes, are disclosed to the other entities and that they respect such rights. To this end, the CMS tracks the use of contents by entities and, in case of infringement of privileges, it requests the FRMS to lower the reputation of the infringing entity and, in extreme cases, that it is excluded from TEco. Another duty of the CMF is to “certify” with legal value the publication of a content on the Web. This feature is strongly felt by many users in the current Web. Let us consider, for example, the case of a university that has issued a call for a research grant. The CMF must certify: 1) that the URL of the content is accessible at any time (**Where**); 2) the date and time (timestamp) of the publication (**When**); 3) the integrity of the content (**What**); 4) the authenticity of the publication, i.e., that it comes directly and truly by the entity (**Who**). If the content is modified after publication, the CMF certifies the four **W** for all the previous versions, which are still stored (*versioning*) and made public. As for RMSs, the CMFs are also part of a federation, called **Federated Content Management Frameworks (FCMF)**, which manages their interaction.

## 4 Discussion

The TEco system is an incremental model which enhances the current Web without replacing it. This is one of its strengths as it requires no upheavals in infrastructures. Furthermore, it does not compel users to adapt to new rules or new software. TEco can be developed in parallel with the Web, leaving the users free to choose between a deregulated area and a trust area, exactly as in the offline world. To make it applicable it is necessary that the systems described so far, federated Identity Management system (IdM), federated Reputation Management System (RMS) and federated Content Management Framework (CMF) are implemented and integrated bearing in mind the characteristics described in this section. To obtain the permanent identifier associated to a TempWAE (see Fig. 1 - Steps 3 and 8) from the federated IdM, the federated RMS must use a specific communication protocol that may be similar to the protocols used in Internet to resolve domain names. This protocol must ensure that the association between permanent and temporary identifiers is known only to the two federations. The implementation of federated RMS also requires that an ontology of the Contexts/Roles and, for each of them, the Main Features are identified, as explained in [17], to which the reader should refer for further details.

It is necessary to define a formal language for the specification of Agreement Terms and Interaction Agreements and to define Standard Agreement Terms and some predefined Interaction Agreements. It is also crucial for the success of TEco the implementation of an efficient Web Agent that facilitates entities in all activities related to TEco. In particular, it could include a plugin which works during Web navigation (e.g., as done in [18]). This plugin would allow

an entity to request a *TEco Interaction* by simply entering the address of the website in the browser and specifying the alter ego s/he intends to use. It will then be the Web Agent to handle the request by interacting with the service providers (see Fig. 1 - Steps 1 and 6). A *TEco Interaction* will be established if and only if the service provider, which also owns a TId, accepts the request. In both cases, the navigation would continue normally, except that a *TEco Interaction* will enable all the benefits of TEco (negotiated agreement, SSO, reputation, etc.) and a browser icon will indicate that the transaction is performed in the trust area (as in https). A protocol for negotiation of the agreement is also necessary to allow the Web Agents to perform it autonomously.

As already mentioned, the federated CMF has to manage all the contents and information related to a TId. This may be done by associating a tuple *[name, url\_of\_value, is\_certified]* to each content, where: *name* represents the attribute name, which can be standard (e.g. *date\_of\_birth*) or user-defined (e.g. *preferred\_wine*); *url\_of\_value* indicates where the attribute value is located (e.g. at a Municipal Registry Office, a link to a *Google+* post, etc.); *is\_certified* indicates whether the attribute is declared by the entity or the url is referred to a *certified* value. Whenever a new attribute of an entity is declared, a new record will be added in the CMF. For instance, following the achievement of the PhD in computer science, a new record like *[PhD, www.unisa.it/PascuccioFA/CSPHD, true]* will be added for the corresponding TId. To simplify the handling of content for an entity, its Web Agent could support the user during the creation of new contents. For instance, when a user publishes in a blog, his/her Web Agent suggests a default repository (the user can chose another one) in which to store the data and then sends a link to the content to the blog. If the content is already present in the CMF, the user can simply choose the link without rewriting the text. This would be totally transparent to the user, who would only compose the content through an appropriate GUI, while all other activities would be carried out independently by the Web Agent.

## 5 Conclusions and Future Work

In this work we discussed some critical issues related to the current Web and proposed an overall solution called TEco, which defines a trust area in the Web, where users can move and safely interact with a greater degree of mutual trust. We showed how in TEco entities can: be identified through a Trusted Digital Identity; keep anonymity and protect their privacy through the use of a Web alter ego; perform Single Sign-On authentication; establish inter pares interactions tying counterparts to comply with specific and agreed conditions; know the reputation of counterparts and have complete control of their data. We also discussed how

it can be implemented through the integration of some existing and new systems and how this enhances the current Web without upheavals.

The work is still preliminary. In the future we will continue to work on TEco taking into account the contributions received by the scientific community. In addition, we will develop the communication protocols among all subsystems and the formal languages to define the Agreement Terms and the Interaction Agreements. Lastly, we will develop a prototypical Web Agent with a basic expertise to enable the testing of TEco.

## References

- [1] S. Srinivasan and R. Barker. Global analysis of security and trust perceptions in web design for e-commerce. *Int. J. of Inf. Security and Privacy*, 6(1):1–13, 2012.
- [2] F. Hendriks, K. Bubendorfer, and R. Chard. Reputation systems: A survey and taxonomy. *Journal of Parallel and Distributed Computing*, 75:184–197, 2015.
- [3] E. Bertino, F. Paci, and N. Shang. Digital identity protection - concepts and issues. *ARES '09*, pages lxi–lxxviii.
- [4] G. D. Tormo, G. L. Millán, and G. M. Pérez. Definition of an advanced identity management infrastructure. *Int. Jour. of Infor. Security*, 12(3):173–200, 2012.
- [5] Y. Cao and L. Yang. A survey of identity management technology. *ICITIS '10*, pages 287–293.
- [6] J. Torres, M. Nogueira, and G. Pujolle. A survey on identity management for the future network. *IEEE Comm. Surveys and Tutorials*, 15(2):787–802, 2013.
- [7] H. Boley and E. Chang. Digital ecosystems: Principles and semantics. In *DEST '07*, pages 398–403.
- [8] E. Chang and M. West. Digital Ecosystems A Next Generation of the Collaborative Environment. *iiWAS '06*.
- [9] E. Pournaras and S.J. Miah. From metaphor towards paradigm - a computing roadmap of digital ecosystems. *DEST '12*, pages 1–6.
- [10] G. D. Tormo, F. G. Mármol, and G. M. Pérez. Towards the integration of reputation management in openid. *Computer Standards and Interfaces*, 2013.
- [11] S. Paul, J. Pan, and R. Jain. Architectures for the future networks and the next generation internet: A survey. *Computer Communications*, 34(1):2–42, 2011.

- [12] B. Gates. Bill Gates: Microsoft's Security Vision and Strategy. *RSA 2006*.
- [13] H. Jo, H. Jin Lee, K. Chun, and H. Park. Interoperability and anonymity for id management systems. *ICACT 2009*, 02:1257–1260.
- [14] R. L. Morgan, S. Cantor, S. Carmody, W. Hoehn, and K. Klingenstein. Federated Security: The Shibboleth Approach. *EDUCAUSE Quarterly*, 27(4):12–17, 2004.
- [15] B. C. Neuman and T. Ts'o. Kerberos: An authentication service for computer networks. *Communications Magazine, IEEE*, 32(9):33–38, 1994.
- [16] W. Sherchan, S. Nepal, and C. Paris. A Survey of Trust in Social Networks. *ACM Comp.Surv.*, 45(4):47:1–47:33, 2013.
- [17] G. Costagliola, V. Fuccella, and F. A. Pascuccio. Towards a Trust, Reputation and Recommendation Meta Model. *JVLC*, 25(6):850–857, 2014.
- [18] G. Costagliola, R. Esposito, V. Fuccella, and F. Gioviale. An architecture for user-centric identity, profiling and reputation services. *DMS 2009*, pages 170–173.

# Joint Fingerprinting and Encryption for JPEG Images Sharing in Mobile Social Network

Conghuan Ye, Zenggang Xiong, Yaoming Ding, Guangwei Wang, Xuemin Zhang, Fang Xu  
College of Computer and Information Science  
Hubei Engineering University  
Xiaogan, China  
ychzzw@163.com

**Abstract:** The advent of mobile social network and smartphone has made social multimedia sharing in social network easier and more efficient. However, it can also cause serious security and privacy problems, secure social multimedia sharing and traitor tracing issues have become critical and urgent. In this paper, we propose a joint fingerprinting and encryption (JFE) scheme based on Game of Life (GL) and singular value decomposition (SVD) with the purpose of protecting JPEG images sharing in mobile social networks. Firstly, the fingerprint code is produced using social network analysis. After that, a fast inter-transformation from block DCTs to DWT is employed. Then, fingerprints are embedded into the LL, HL and LH subbands. At last, GL and SVD are used to for confusion and diffusion respectively. The proposed method, to the best of our knowledge, is the first JFE method using GL and SVD in the JPEG compressed domain for security and privacy in JPEG images sharing. The use of fingerprinting along with encryption can provide a double-layer of protection to JPEG images sharing in social network. Theory analysis and experimental results show the effectiveness of the proposed JFE scheme. Most importantly, the performance of inter-transformation between block DCT and one-level DWT has a profound effect on lowering computational cost in our proposed JFE scheme. In the end, our JFE method can secure JPEG images sharing in social networks and meet the real-time requirement.

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

## I. INTRODUCTION

The advent of mobile social network, cloud computing, and smartphone makes social multimedia sharing become pervasive in our daily lives. A group of users, geographically distributed, share the same social multimedia content-images, video, and audio with their mobile devices in a social networking community. The growth of social multimedia, user-generated, transmitted, consumed or shared in social network[1], underscores potential risks for the unethical use with the emergence of mobile devices such as smartphones. Social multimedia distribution within social network raises distinctive challenges such as privacy and security issues. Preserving privacy in publishing social multimedia becomes an important concern to prevent illegal use of social multimedia.

However, secure social multimedia sharing is still in its infancy and high dependent on both confidentiality and redistribution tracing, therefore, techniques, such as

fingerprinting and encryption [2], need to be carried out. For the purpose of confidentiality of social multimedia sharing, cryptography techniques transform multimedia content into an enciphered, unintelligible form, that keep the encrypted content from illegal access difficult during the distribution processes without the decryption key. In order to achieve such type of security, employing chaotic systems with the properties of initial-value sensitivity and parameter sensitivity in generating the encryption keys has become one of the important topics in secure multimedia communication. Its main advantage lies in the observation that a chaotic signal looks like noise for non-authorized users ignoring the mechanism for generating it. Only the authorized customer who has the correct key can recover the data successfully [3].

A variety of chaos-based image encryption schemes have been proposed [4-6]. These proposed schemes reduce the risk of sensitive content being revealed by one other than the intended recipient. However, these schemes only focus on encrypting. Once users receive and decrypt the data, the content could be copied and delivered to an unauthorized user at their option. There are not ways to continue the work of protecting the multimedia content, therefore the privacy may still be leaked. In this case, extra protection schemes should be adopted to deter content redistribution. Digital fingerprinting, in which a user specific identification mark is embedded into a copy of original content, is a useful tool to trace redistributed content. Although encryption and fingerprinting are used to protect multimedia separately, there are some works[7, 8] used both techniques for the secure sharing of multimedia content. The need to apply both fingerprinting and encryption to digital images keeps rising in recent years [9-11].

However, these existing approaches seem not desirable to be applied to mobile multimedia content encryption because their encryption and decryption procedure in non-compressed domains makes the distribution of content will be very slow for large volume of multimedia data and do not meet the real-time constraint of social multimedia sharing in resources-constrained mobile social network environment. Apart from billions of camera exists, there are billions of smartphones which are being used in the world and majority of them are equipped with camera. If every smartphone clicks only one image per day then what is the volume of images and what amount of network bandwidth required to deliver these images within mobile social network environment is not tough to

imagine. To save computation overhead and transmission bandwidth, compression is a must for image processing and distribution, especially in resource limited environments [12].

Digital watermarking is not new for the multimedia protection but fingerprinting for the compressed encrypted images is relatively new and appealing. To save space, a large volume of data is being kept in the compressed format. For the various reasons, there is a great issue of security protection for these compressed images. Growth of compressed image has magnified the need for more advanced encryption and fingerprinting techniques. [13].

In fact, there are some problems about the existing joint fingerprinting and encryption schemes. First, most existing schemes are not focus on the compressed content. As a result, there are huge volume of multimedia content has to be distributed. Second, there is no scalability in some existing schemes when they are used in secure content sharing in mobile social network environment. Then, the processing speed about encryption and fingerprinting may be a bottleneck when dealing with social multimedia data in social network for both central servers and resources-constrained devices, such as smartphone. At last, the distribution method will cost a lot of resources, especially network bandwidth.

In this article, we mainly focuses on the problem of privacy leakage, the illegal duplication and redistribution of social multimedia content in resources-constrained mobile social network environment. We are trying to deal with the issues of security and privacy in JPEG images sharing in mobile social network environment. We also are trying to address new future challenges of JFE ( joint fingerprinting and encryption ) in the area of JPEG images sharing. we present a novel JFE method using SNA (social network analysis) to deal with the issues of JPEG images sharing. Firstly, we describe a method for fingerprint code produced by dendrogram of hierarchical and overlapping structure of social network. Then, we propose a JFE method based on GL ( Game of Life ) and SVD in the DWT domain directly from the JPEG content. By using our technique, one is well able to design a privacy-preserving and secure social multimedia sharing in resources-constrained mobile social network environment.

According to our best knowledge, there has been no report yet on the implantation of JFE scheme based on GL and SVD using social network analysis for secure social multimedia sharing in resources-constrained mobile social network environment. The remainder of this paper is organized as follows. Related works are introduced in Section 2. In Section 3, techniques used in this paper will be introduced. Section 4 details the proposed JFE scheme based on GL and SVD. Then, the experimental results will be given in Section 5. Finally, conclusions are drawn in Section 6.

## II. RELATED WORKS

There have been some related works on access control to content security [14-16]. Commutative encryption and watermarking (CEW) could be used for providing more comprehensive security protection for multimedia content. D. Bouslimi et al. proposed a joint encryption and watermarking algorithm in [17]. The convergence of the two technologies is

now facilitating privacy and security studies [18]. Two robust watermarking algorithms were proposed to watermark compressed JPEG images in encrypted domain [19] and JPEG2000 compressed and encrypted images [20] respectively. Kundur and Karthik [21] proposed a novel architecture for joint fingerprinting and decryption (JFD) that holds promise for a better compromise between practicality and security. Another joint fingerprinting and decryption (JFD) scheme based on vector quantization is proposed with the purpose of protecting media distribution [22].

However, all the above schemes did not be applied to JPEG images sharing in resources-constrained mobile social network environment. In view of the increasingly important role played by digital imaging in mobile multimedia social network with the emergence of smartphone, it is necessary for large amount of image data to be economically stored and/or transmitted. In these cases, the fingerprinting and encryption should be implemented in the compressed content to avoid the process of fully decoding and encoding. In social networks, practical multimedia contents are stored and transmitted in compressed JPEG format. Understanding the inherent characteristics of JPEG may play a useful role in digital image forensics[23]. As the perceptual information concentrates at low-frequency DCT coefficients, this leads to the research on selective encryption for JPEG images[24]. Lian proposed to encrypt the DC coefficient and the sign bit of all AC coefficients using a spatiotemporal chaotic system [25]. However, Wu and Kuo [26] stated that selective encryption is not suitable for DCT-based compression algorithms because some perceptual attacks were able to restore a perceptual image. The encrypted data is not secure in visual perception since the encryption of signs of DCT coefficients cannot fully scramble the original data.

In addition, the traditional fingerprinting methods do not consider the relationship between users in social network; then they cannot be applied to secure sharing in social network. How to use SNA to embed fingerprint information in encrypted contents and how to make the content sharing system robust against attacks is not deeply considered. In order to address social multimedia sharing, the authors proposed a secure content sharing method in the TSH transform domain in [27] through mapping the community structure of social network into the tree structure wavelet transform. With the different wavelet bases and decomposition levels, the DWT can extract different kinds of information from the media, and is therefore very likely to map community structure of social network into the tree structure of DWT, which can be used to joint fingerprinting and encryption. To encrypt the important content only, DWT domain algorithm can improve the encryption speed.

## III. BASIC THEORY

### A. SVD

SVD is a very useful tool in linear algebra, which is a factorization and approximation technique. From the perspective of image processing, an image can be viewed as a matrix with non negative scalar entries. Mathematically, SVD of a rectangular matrix  $A$  is expressed as



$$A = USV^T \quad (1)$$

where  $S$  is also known as singular value matrix in SVD domain,  $U$  and  $V$  are the unitary matrices. Both of  $U$  and  $V$  components are composed of eigenvectors of matrix  $A$ , and  $T$  represents the conjugate transpose operation.  $U$  and  $V$  are also orthogonal matrices. Therefore, the following conditions are always satisfied

$$I_N = U^T U = U U^T \quad (2)$$

$$I_M = V^T V = V V^T \quad (3)$$

where  $I_N$  and  $I_M$  are identity matrices with size  $N \times N$  and  $M \times M$ , respectively.

### B. Chaotic maps

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

$$x_{n+1} = ux_n(1 - x_n) \quad (4)$$

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. (5):

$$y_{n+1} = F(y_n, \eta) = \begin{cases} y_n / \eta, & 0 \leq y_n < \eta \\ (y_n - \eta) / (0.5 - \eta), & \eta \leq y_n < 0.5 \\ 0, & y_n = 0.5 \\ F(1 - y_n, \eta), & 0.5 \leq y_n < 1 \end{cases} \quad (5)$$

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. MD5

A cryptographic hash function is designed to ensure message integrity. MD5 takes a sequence of data as input and output a 128-bit "fingerprint" or "message digest" of the input. A MD5 hash is typically expressed as a 32 digit hexadecimal number. The MD5 hash function is one way, even if there is only a tiny bit change between two input sequences, the returned MD5 hash value will be totally different. It takes a message with a variable-length and returns a fixed-length output of 128 bits. As MD5 is fast and sensitive to the input

message, it is employed in our JFE scheme to partially determine the control parameter and the initial condition of a chaotic tent map. A small change in the given fingerprints affects the control parameter and the initial condition of the chaotic map. As a result, the change will effectively spread to the whole cipher-image.

### D. CA ( Cellular Automata )

CA [28] is a dynamical complex space and time discrete system. 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. In (2-D) CA, Moore neighborhoods method is used [29]. The Moore neighborhood of range  $L$  is defined by

$$NH(x_0, y_0, L) = \{(x, y) : |x - x_0| \leq L, |y - y_0| \leq L\} \quad (6)$$

At every time step, all the cells update their states synchronously by applying rules (transition function). Each cell computes its new state by applying the following transition rules.

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

For binary cells  $c_1, c_2, \dots, c_9$ , we say that the transition function, at any time  $t$ , for GL rule [30] is of the form:

$$\phi \left( \begin{matrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \\ c_7 & c_8 & c_9 \end{matrix} \right) = \begin{cases} 1, & \text{if } \sum_{i=1}^9 s(c_i, t) = 3 \\ 1, & \text{if } \sum_{i=1}^9 s(c_i, t) = 3, i \neq 5 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

## IV. THE PROPOSED JFE ALGORITHM

### 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
$X_*$	the half of original fingerprint vector

$Sum_*$  the sum of half fingerprint vector  
 $d_k$  the dither sequence  
 $Y_k$  the fingerprinted coefficients vector  
 $G^0$  the initial two-dimensional grids of cells  
 $Pl_r$  the input image patch;  
 $Cp_r$  the output scrambled matrix patches  
 $A_i$  the GL matrix  
 $I^{JFE}$  the encrypted and fingerprinted image

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

We try to use method in [31] to get the overlapping and hierarchical structure of social network for fingerprint encoding. As shown in Fig.1, users are placed into  $c$  communities. These communities are encoded by outer code that is constructed by BS code [32], and the users in each community are encoded by the inner code produced with Tardos scheme [33]. Therefore, Fingerprint code for  $N_u$  users can be concatenated by a multilevel outer code for communities and an inner Tardo code for users in the communities, which is detailed in our previous work [34].

#### B. DWT from DCT directley

DCT is adopted in the JPEG compression standard [9]. The image is first divided in  $8 \times 8$  blocks and each of these is transformed with the DCT. DCTs convert data from the spatial domain into the frequency domain. The transformed blocks are quantized with a uniform scalar quantizer, zig-zag scanned and entropy coded with variable length coding (VLC), this mode is simply regarded as Joint Photographic Experts Group (JPEG). The block-based segmentation of the source image is fundamental limitation of the DCT-based compression system. The JPEG compression methods actually gained widespread acceptance as image compression methods. Most compressed multimedia contents in social network are stored as block DCT coefficients and motion vectors. The compressed JPEG images are partially decoded to obtain block DCT coefficients which are subsequently used to construct one-level DWT.

To lower the computational complexity, we use a fast inter transformation between one-level DWT and block DCTs. Compared with DCT, the wavelet transform is closer to the human visual system (HVS) because it splits the input image into several statistically frequency bands that can be processed independently. DWT also causes fewer visual artifacts than DCT because the wavelet transform does not decompose the image into blocks for processing. In the DWT transform, an image is split into one approximation (also called LL subband) and three details in horizontal, vertical, and diagonal directions which are named (or coefficients in LH subband), (coefficients in HL subband), and (coefficients in HH subband). The LL

subband is then itself split into a second-level approximation and details, and the process is repeated.

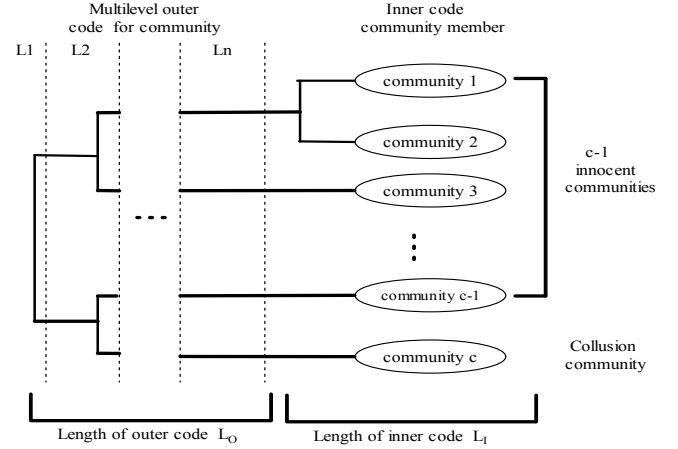


Figure 1. Fingerprint coding using social network analysis

Obtaining one-level DWT of JPEG compressed images is time consuming using existing methods, which first decompress the block DCTs of images into pixel data and then perform DWT on the data. In fact, both DCT and DWT are linear and invertible transforms, and a linear relationship exists between block DCT coefficients and its DWT coefficients [35]. Our JFE method can directly obtain the JPEG image's one-level DWT coefficients for fingerprinting and encryption in the DWT domain from the block DCT coefficients without involving inverse DCT (IDCT) to maintain low computational cost. Because the consumed time of IDCT and that of DCT occupy prodigious proportions of that of full decoding and encoding, respectively [36]. These approaches are inefficient because a large amount of time is spent on the inter-conversion between the DCT coefficients and spatial pixel data. Hence, the fast direct conversion between the block DCTs and one-level DWT coefficients, which will prevent full decoding and encoding, is indispensable for resource-constrained mobile social network environment.

Let's consider an image (or a frame)  $I$  with a size of  $(L \times S) \times (K \times S)$ . We can divided this image into  $L \times K$  blocks, which are denoted as  $BL_{ij}$  with a size of  $S \times S$ .  $C_{ij}(u, v)$  representing the DCT coefficients of blocks can be expressed as

$$C_{ij}(u, v) = \sqrt{\frac{2}{S}} \alpha(u) \sum_{q=0}^{S-1} \sqrt{\frac{2}{S}} \alpha(v) \sum_{p=0}^{S-1} I(p, q) \cos\left(\frac{(2p+1)u\pi}{2S}\right) \cos\left(\frac{(2p+1)v\pi}{2S}\right) \quad (8)$$

where  $u, v = 1, 2, \dots, S$ ,  $\alpha(u), \alpha(v) = \begin{cases} 1/\sqrt{2}, & u = 0 \text{ or } v = 0 \\ 1, & \text{else} \end{cases}$ .

According to Eq.(8), the 2D DCT transform of  $Sb_{ij}$  can be rewritten and represented in matrix formats  $C_{ij}(u, v) = B_1 \times BL_{ij} \times B_1^T$ . The inverse transform of it can be expressed as  $BL_{ij} = B_1^{-1} \times C_{ij} \times (B_1^T)^{-1}$ , where  $B_1$  and  $B_1^T$  are

orthogonal matrix of the block DCT. So the whole image can be expressed as

$$I = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_1 \end{bmatrix}_{LS \times LS}^{-1} \times \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1K} \\ C_{21} & C_{22} & \cdots & C_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ C_{L1} & C_{L2} & \cdots & C_{LK} \end{bmatrix} \times \begin{bmatrix} B_1^T & 0 & \cdots & 0 \\ 0 & B_1^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_1^T \end{bmatrix}_{KS \times KS}^{-1} \quad (9)$$

The three matrices on the right of Eq.(9) are denoted as  $B_4$ ,  $C_{part}$  and  $B_5$ , respectively. We can also compute the one-level DWT coefficients of image  $I$ . Here, the DWT will be taken using the haar wavelet, which is the simplest possible wavelet. It is both separable and symmetric and can be expressed in matrix form

$$KR = H \times I \times Q^T \quad (10)$$

For the Haar wavelet transform,  $H$  contains the Haar basis functions,  $h_z(k)$ . They are defined over the continuous, closed interval  $z \in [0,1]$ . Then the Haar basis functions are

$$h_0(z) = h_{00}(z) = \frac{1}{\sqrt{LS}}, z \in [0,1].$$

$$h_k(z) = h_{pq}(z) = \frac{1}{\sqrt{LS}} \begin{cases} 2^{p/2}, (q-1)/2^p \leq z < (q-0.5)/2^p \\ -2^{p/2}, (q-0.5)/2^p \leq z < q/2^p \\ 0, \text{ otherwise, } z \in [0,1] \end{cases} \quad (11)$$

The inverse Haar wavelet transform can be expressed as  $I = H^T \times KR \times Q$ . Then, we obtain coefficient matrix  $KR$  in DWT domain using this expression:

$$KR = A_1 \times C_{part} \times A_2 \quad (12)$$

where  $A_1 = H \times B_4$ ,  $A_2 = B_5 \times Q^T$ . This may contribute to reduce computational cost. Through the inverse transformation in Eq.(13), we can directly obtain the set of block DCT coefficients from the DWT coefficient matrix using:

$$C_{part} = A_1^T \times KR \times A_2^T \quad (13)$$

In the DWT transform [37], an image is split into  $LL$ ,  $LH$ ,  $HL$ , and  $HH$  subband. In this paper, we transform middle-frequency subbands repeatedly. 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.

### C. The JFE process

The architecture of JFE scheme based on DWT and chaotic CA is designed and shown in Fig.2. The JFE process is composed of two processes: fingerprinting and encryption.

#### 1) Fingerprint embedding

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 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 outer code, and choose another robust coefficients sequence in  $LL$  subband to create a vector,  $X^I = (x_1, x_2, \dots, x_{L^I})$ . The outer code hiding scheme is described in Eq. (14), to hide fingerprint codeword. The hiding scheme is as follow:

$$y_j^{(i)} = q_{x_j} = \text{round}\left(\frac{x_j + d_i^{(j)}}{\Delta}\right) \times \Delta \quad (14)$$

Where  $x_j$  is a vector which represents the host signal with length  $L^O$ ,  $i, j = 1, \dots, L^O$ , round is an operation of *Floor and Ceiling*, and  $\Delta$  is a constant.

In this case, to identify the embedded fingerprint, the multimedia producer needs to obtain the fingerprinted coefficients, which compose a vector  $z$ . By deducting, the difference is as follow:

$$T_k = \|z - y_k\|^2, k=1, \dots, L \quad (15)$$

Here, the least  $T_k$ , which is related to user  $k$ , determines who the traitor is.

#### 2) Encryption algorithm

Digital media contents like image and video are tightly related to visual quality. Generally, social multimedia encryption algorithms should be not only secure against cryptographic attacks but also secure in human perception. The more degraded their visual quality is, the higher the security is.

In this paper, we focus on JPEG images in mobile social network, so the encrypted output must ideally “appear” random to make estimation of the original image from the encrypted one computationally difficult without access to the decryption key; traditional multimedia content encryption algorithms are considered computationally infeasible for high volumes of multimedia content in resources-constrained devices or for near real-time or massively parallel distribution of multimedia content flows [38].

The traditional solution applies a encryption algorithm on the compressed image in JPEG format, with which the total processing time will be longer. In order to overcome some limitations, we propose the notion of partial encryption, in which a smaller subset of the important content in the DWT domain is encrypted to lower computation and delay while integrating the fingerprinting with encryption. CA (Cellular automata) 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 [39]. CA capable of exhibiting chaos is attractive in cryptography because it is possible to have a great number of possible keys in the keyspace. Fast computation helps in achieving this capability. We are

interested to use chaotic CAs in order to take advantage for fast cryptography. SVD performs an optimal matrix decomposition in a least-square domain for matrices in real number domain.

Permutation-only type image cipher is superior in the aspect of efficiency due to its lowest computational complexity. To overcome the drawbacks of conventional permutation-only type image cipher, a novel JPEG image fingerprinting and encryption method based on CA and SVD in the DWT domain is proposed in Fig.2. The encryption process is composed of substitution with GL and diffusion based on SVD of random matrix in the DWT domain. The proposed encryption algorithm can be divided into the following steps:

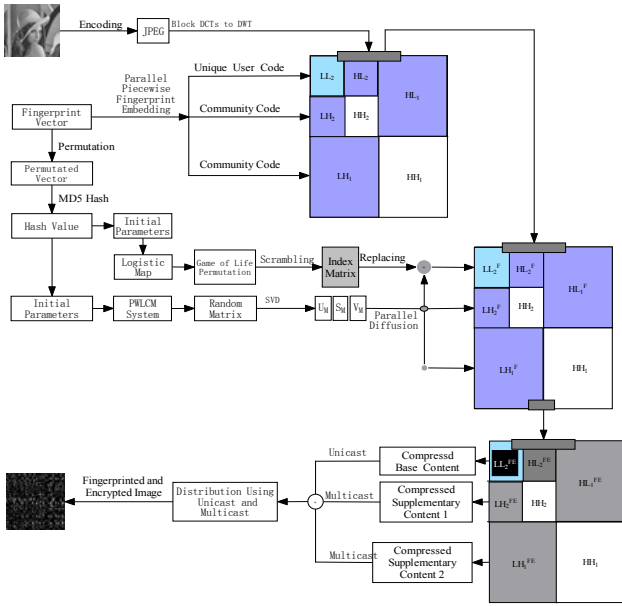


Figure 2. The architecture of JFE scheme

Step 1: Given a user fingerprint vector, permute the vector randomly, and then divide the permuted fingerprint vector into two parts:  $X = X_1 + X_2$ , where  $X_*$  denotes half of the vector, calculate the sum of both parts denoted by  $Sum_{X_1}$  and  $Sum_{X_2}$  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, MD5 is a widely used cryptographic hash function with a 128-bit hash value [40]. The MD5 hash value of  $Th$  is  $V^{Th}$ . According to the order of bits, we segment  $V^{Th}$  into eight 16-bit parts  $V_1^{Th}$ ,  $V_2^{Th}$ , ...,  $V_8^{Th}$ , 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) Initial values  $x_0$  (Logistic

map) and  $y_0$  (PWLCM system); (2) 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: Chaotic cellular automata for scrambling matrix generation. Adjacent coefficients in an image have a strong correlation. To scramble the image, this correlation needs to be reduced. We propose performing coefficient scrambling with the help of a number of generations of the GL. The universe of the GL is an infinite two-dimensional orthogonal grid of square cells, each of which is in one of two possible states, alive or dead. GL will add the diffusion property to the scrambling technique. At each step in time, the proposed scrambling matrix generation algorithm can be described as follows:

(1) Use logistic map to generate sequences  $(x_1, x_2, \dots, x_{M \times N})$  respectively, where  $x_0$  and  $u$  are given in advance as keys. Then we create a two-dimensional grids of cells  $G^0$ , as the seeds of GL by the sequences, 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 dead. Where  $G^0$  is used to permute the DWT transformed coefficient matrixes; An  $M \times N$  GL automaton is set up with an initial random configuration  $A_0$ , and is set to run for  $k$  generations, thus obtaining  $\{A_1, A_2, \dots, A_k\}$  matrices.

(2) Let  $I_G$  denote the original matrix, then getting the patches set of  $\{Pl_1, Pl_2, \dots, Pl_k\}$  in the original matrix.

(3) For every  $Pl_r$  in the patches set (for  $r = 1, \dots, m$ ). Let  $Pl_r$  denote the input image patch;  $Cp_r$  denotes the output scrambled matrix patches and  $A_1$  is the first generation produced by the GL. Set row=1, col=1.

(4) For all  $(i, j)$  such that  $A_1(i, j) = 1$ , take the value of element  $P_e$  (row, col), put it in  $Cp_r(i, j)$ , and increment (row, col) with row-first order to point to the next objective in the input matrix.

(5) For  $p = 2, \dots, k$ , for all  $i, j$  such that  $A_p(i, j) = 1$  and  $A_n(i, j) = 0$  (for  $n = 1, \dots, p-1$ ), take the value of element  $P_e$  (row, col), put it in  $Cp_r(i, j)$ , and increment (row, col) to point to the next objective.

(6) Take the gray value of the remaining objectives in  $Pl_r$ , and put them in row-first order in those  $Cp_r(i, j)$  where, for all  $p = 1, 2, \dots, k$ ,  $A_p(i, j) = 0$ .

(7) Assume every  $Cp_r$  as a independent patch, steps 3, 4, and 5 are used to scramble these patches in the original grid.

Fig.3 (left) displays the first generation of the GL; living cells (cells in state 1) are shaded. Fig.3 (middle) shows the initial step of the algorithm to an matrix with  $6 \times 6$  elements. Fig.3 (right) displays the scrambled matrix after applying step (3) of the algorithm. To recover the original matrix, the inverse of the scrambling algorithm must be executed, using as keys the initial random configuration  $A_0$  and the number of iterations  $ki$ .

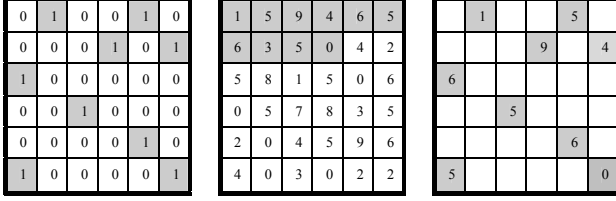


Figure 3. Encryption using CA. (left, first generation GL, middle, original patch, right, encrypted patch)

Step 3: For a compressed JPEG image, we calculate the one-level DWT coefficient matrix of the image from the block DCTs. Then we can get four sub-bands: the approximation coefficients LL, and the detailed coefficients HL, LH, HH. The low-frequency LL subband of the one-level DWT is a down-sampled image of the origin image. This process can reduce computational cost and lower complexity because most of the modified differences of all DWT coefficients are zero. Then, perform two-level DWT decomposition.

Step 4: The corresponding plain coefficients of the LL subband are put to the scrambling coefficient matrix according to the index matrix one by one.

Step 5: To protect content further, diffusion processes with the PWLCM map and SVD can enhance the resistance to attack. Using the PWLCM map to generate chaotic sequences

$RP_{M \times N}' = \{rp_1', rp_2', \dots, rp_{M \times N}'\}$ , then we can get the sequences

$CP_{M \times N}' = \{cp_1', cp_2', \dots, cp_{M \times N}'\}$ ,  $cp_i = \text{ceiling}(fp_i)$ , which is one-to-one correspondent with the coefficient sequence in DWT domain. The obtained chaotic sequence is arranged in the form of a matrix of dimension  $M \times N$ , which is denoted by  $CP'$ , as a random matrix. Perform SVD on  $CP'$ , we get  $CP' = U_{CP} V_{CP} V_{CP}^T$

Step 6: Deform all coefficients of each subband using orthonormal matrices  $U_{CPK}$  and  $V_{CPK}^T$ , as

$$I^{FE} = \begin{cases} U_{CP} I V_{CP}^T, M \leq N \\ V_{CP} I U_{CP}^T, M > N \end{cases}$$

Then, We can get the scrambled and fingerprinted image  $I^{JFE}$ .

## V. EXPERIMENT RESULTS AND SECURITY ANALYSIS

The performance of the proposed JFE technique is demonstrated using MATLAB platform on a computer having a Pentium(R) Dual-Core E5700 CPU and 2-GB RAM. We used eight types of test images with different spatial and frequency characteristics: Elena, Peppers, Airplane, Fishingboat, Baboon, and Watch. We set parameters  $x_0 = 0.98968389485321$ ,  $u = 3.9978859364826$ ,

$y_0 = 0.4576412939342$ ,  $\eta = 0.45967789391392$ . Fig. 4(a) shows the encrypted image. Fig. 4 (b) show the decrypted image with fingerprints under the correct key. From the results of our experiment, we can see it is difficult to recognize the original image from the encrypted one.

### A. Perceptual Security

A good multimedia content encryption algorithm should be sensitive to the cipher keys, and the key space should be large enough to make brute force attack infeasible. Generally, the encrypted image should be unintelligible for confidentiality. In the proposed scheme, the LL coefficients in DWT domain are encrypted by permutation via GL firstly. Then the scrambled values of coefficients are changed using SVD. The visual impact of the proposed encryption scheme is demonstrated in Fig.4(a). 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.

### B. Imperceptibility of the Fingerprint

The fingerprint is embedded in the image during the decryption process. In order to preserve visual quality, the fingerprint in the fingerprinted copy should be imperceptible and perceptually undetectable. Fig. 4(b) shows some experimental results of decrypted fingerprinted images. It can be observed that the quality of the fingerprinted image does not have any change observably.

### C. Ability of resisting exhaustive attack

The total key space includes two processes of 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  $10^{-16}$  [41], The total key space is about  $10^{16 \times 4} = 10^{64}$ . This key space is large enough to resist the brute-force attack.

### D. Resistance to statistical attack

#### 1) The grey histogram analysis

Since, the proposed scheme is a rapid selective encryption in the compressed domain. Hence, the basic idea is to compare the histograms of the original and encrypted media. If the histograms of the encrypted images are fairly uniform and is

significantly different from the histogram of the original one, then the encryption is said to be perfect. Fig. 4(c), (d) shows the grey-scale histograms. Comparing the two histograms we find that the pixel grey values of the original images are concentrated on some values, but the histograms of the encrypted images are significantly different from the histograms, and the histograms of the encrypted ones are very similar, which makes statistical attacks difficult.

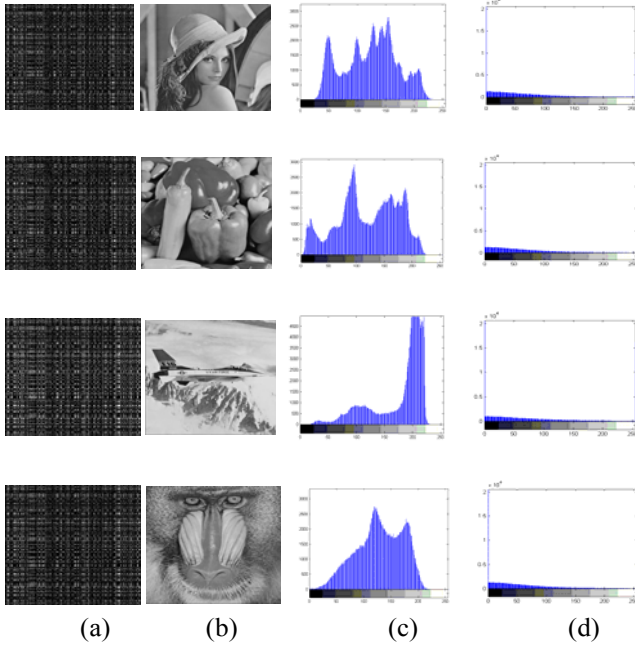


Figure 4. The experimental results : (a) the encrypted images, (b) the decrypted images with fingerprints, (c) the grey histogram of the original images, (d) the grey histogram of the encrypted images.

## 2) Correlation coefficient analysis

The correlation analysis says that a good encryption technique must break the correlation among the adjacent image pixels. We randomly select 2000 pairs (horizontal, vertical and diagonal) of adjacent pixels from the original image and the encrypted image. Fig.5(a), (b) show the correlation of two adjacent pixels in the original Lena image and its encrypted image. Fig.5(b) shows that the correlations of adjacent pixels in the encrypted image are greatly reduced.

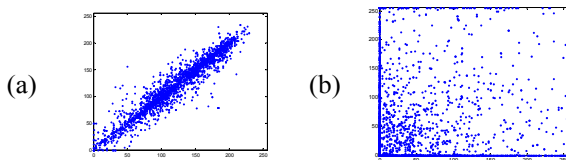


Figure 5. Correlation of two adjacent pixels

## E. Discussion of the encryption process

We knew that the permutation process only enhances the unintelligibility of the encrypted image. Although single coefficient in 2-level LL subband permutation via GL can

achieve better effect than  $4 \times 4$  blocks in the 2-level LL subband permutation, however, the first method took 16 times as much as time that the latter took. In fact, the latter can get almost the same encryption effect that  $4 \times 4$  blocks permutation in all subbands of 1-level DWT via GL could achieve, in this comparison, permutation performed only took 1/16 time that permutation in all subbands took. Therefore,  $4 \times 4$  blocks permutation in the 2-level LL subband can get better performance than the others.

On the other hand, even if the chaotic map used in GL is cracked, the hacker still cannot decrypt the image since the random matrix key of diffusion in SVD encryption process 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 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.

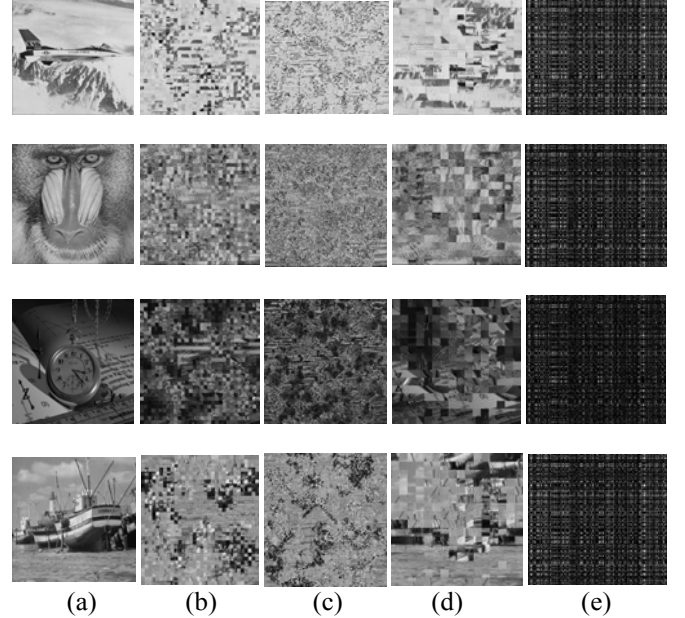


Figure 6. Evaluation of the encryption process. (a) Original images, (b)  $4 \times 4$  blocks in the 2-level LL subband permutation via GL, (c) Single coefficient in 2-level LL subband permutation via GL, (d) Permuted  $4 \times 4$  blocks in all subbands of 1-level DWT via GL respectively, (e) image encryption with permutation on 2-level LL subband via GL and diffusion using SVD distortion.

## F. Encryption Efficiency

### 1) Comparative Analysis

This subsection presents a comparative analysis of the proposed technique with the existing state of art. The considered technique is a joint encryption/watermarking algorithm for verifying medical image presented by D. Bouslimi et al.[42]. 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. On the other hand, watermarking and encryption are conducted in the spatial domain. The approach is inefficient because a large amount of time is spent on the inter-conversion between the DCT coefficients and spatial pixel data. The proposed algorithm is able to overcome the aforementioned weaknesses by incorporating chaotic map with CA and SVD in the compressed domain. It is evident that the chaotic maps are very sensitive to their initial seed. Therefore, a slight change in the initial seed will cause the significant change in their final value. Furthermore, the proposed technique is perceptually efficient. The use of orthogonal matrices obtained by SVD produces a nonlinear process for the encryption of LL subbands. Another benefit of using the SVD is that it ensures the matrices used in encryption are invertible so that the inverse process will perfectly reconstruct the subbands. This proves an improvement by the proposed technique over the existing watermarking and encryption technique.

## 2) Time efficiency

In the case of multimedia distribution in social networks, if a technique requires a huge amount of time to encrypt/decrypt a 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 0.9 s 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 1.** TIME EFFICIENCY

Images	Lena	Peppers	Airplane	Baboon	watch	Fishingboat
Time(s)	1.02	0.92	0.90	0.92	0.92	0.90

## VI. CONCLUSION

The traditional JFE methods don't consider the relationship between users and resource-constrained mobile devices, therefore then cannot be applied to secure multimedia sharing for resource-constrained mobile social network because of the tremendous scale of social network and the limited resources that mobile devices have. In this paper, the first JFE method based on CA and SVD in the DWT transform domain for mobile social network to deal with the issues of JPEG images 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, and statistical attacks. Our methods does not require a great deal of computation time in comparison with full decoding because the proposed algorithm can transform JPEG images into DWT domain directly. Therefore, the efficiency is desirable, our algorithm is meant to be a good candidate to ensure the security of JPEG images distribution. Above all, the proposed scheme can continually protect decrypted content from being illegally

distributed by an authorized member. Therefore, by the proposed JFE mechanisms the risk of the illegal distribution of authorized users can be reduced. The fundamental goal of our research has been to provide a useful synthesis of social network analysis for the field of secure JPEG images sharing for mobile social network.

## ACKNOWLEDGMENT

This work is supported by NSF of China Grants ( 61370092, 61370223 ), Natural Science Foundation of Hubei Province of China (No.2015CFB236, No.2014CFB188), and Youth innovation team project in Hubei Provincial Department of Education (No.T201410).

## REFERENCES

- [1] P. Belimpasakis and A. Saaranen, "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] S. Lian and X. Chen, "Traceable content protection based on chaos and neural networks," *Applied Soft Computing*, vol. 11, pp. 4293-4301, 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] J. Guo, P. Zheng, and J. Huang, "Secure watermarking scheme against watermark attacks in the encrypted domain," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 125-135, 2015.
- [8] A. Qureshi, D. Megias, and H. Rifà-Pous, "Framework for preserving security and privacy in peer-to-peer content distribution systems," *Expert Systems with Applications*, vol. 42, pp. 1391-1408, 2015.
- [9] C. Ye, Z. Xiong, Y. Ding, G. Wang, J. Li, and K. Zhang, "Joint fingerprinting and encryption in hybrid domains for multimedia sharing in social networks," *Journal of Visual Languages & Computing*, vol. 25, pp. 658-666, 2014.
- [10] 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.
- [11] B. Czapski and R. Rykaczewski, "Matrix-based robust joint fingerprinting and decryption method for multicast distribution of multimedia," *Signal Processing*, 2014.
- [12] T. Xiang, C. Yu, and F. Chen, "Secure MQ coder: An efficient way to protect JPEG 2000 images in wireless multimedia sensor networks," *Signal Processing: Image Communication*, vol. 29, pp. 1015-1027, 2014.
- [13] R. Gupta and S. Jain, "A review on watermarking techniques for compressed encrypted images," in *Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014 International Conference on*, 2014, pp. 10-13.
- [14] W. Wang, "Team-and-role-based organizational context and access control for cooperative hypermedia environments," in *Proceedings of the tenth ACM Conference on Hypertext and hypermedia: returning to our diverse roots: returning to our diverse roots*, 1999, pp. 37-46.

- [15] S. K. Chang, G. Polese, M. Cibelli, and R. Thomas, "Visual authorization modeling in e-commerce applications," *IEEE Multimedia*, vol. 10, pp. 44-54, 2003.
- [16] M. Giordano and G. Polese, "Visual Computer-Managed Security: A Framework for Developing Access Control in Enterprise Applications," *IEEE Software*, vol. 30, pp. 62-69, 2013.
- [17] 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.
- [18] 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.
- [19] A. Subramanyam and S. Emmanuel, "Robust watermarking of compressed JPEG images in encrypted domain," 2011, pp. 37-57.
- [20] 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.
- [21] D. Kundur and K. Karthik, "Video fingerprinting and encryption principles for digital rights management," *Proc. IEEE*, vol. 92, pp. 918-932, 2004.
- [22] 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 Processing*, 2012.
- [23] W. Luo, J. Huang, and G. Qiu, "JPEG error analysis and its applications to digital image forensics," *IEEE Transactions on Information Forensics and Security*, vol. 5, pp. 480-491, 2010.
- [24] C. H. Yuen and K. W. Wong, "A chaos-based joint image compression and encryption scheme using DCT and SHA-1," *Applied Soft Computing*, vol. 11, pp. 5092-5098, 2011.
- [25] S. Lian, "Efficient image or video encryption based on spatiotemporal chaos system," *Chaos, Solitons & Fractals*, vol. 40, pp. 2509-2519, 2009.
- [26] C. P. Wu and C. C. J. Kuo, "Design of integrated multimedia compression and encryption systems," *IEEE Transactions on Multimedia*, vol. 7, pp. 828-839, 2005.
- [27] C. Ye, H. Ling, F. Zou, and C. Liu, "Secure content sharing for social network using fingerprinting and encryption in the TSH transform domain," in *Proceedings of the 20th ACM international conference on Multimedia*, 2012, pp. 1117-1120.
- [28] S. Wolfram, "A new kind of science," *Wolfram Media*, 2002.
- [29] S. Wolfram, "Theory and applications of cellular automata," 1986.
- [30] A. Adamatzky, *Game of life cellular automata*: Springer, 2010.
- [31] 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.
- [32] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inf. Theory* vol. 44, pp. 1897-1905, 1998.
- [33] G. Tardos, "Optimal probabilistic fingerprint codes," *J. ACM* vol. 55, p. 10, 2008.
- [34] 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.
- [35] B. J. Davis and S. H. Nawab, "The relationship of transform coefficients for differing transforms and/or differing subblock sizes," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1458-1461, 2004.
- [36] L. Wang, H. Ling, F. Zou, and Z. Lu, "Real-Time Compressed-Domain Video Watermarking Resistance to Geometric Distortions," *IEEE Multimedia*, vol. 19, pp. 70-79, 2012.
- [37] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674-693, 1989.
- [38] D. Kundur and K. Karthik, "Video fingerprinting and encryption principles for digital rights management," *Proceedings of the IEEE*, vol. 92, pp. 918-932, 2004.
- [39] S. Wolfram and M. Gad-el-Hak, "A new kind of science," *Applied Mechanics Reviews*, vol. 56, p. B18, 2003.
- [40] H. Liu and X. Wang, "Color image encryption based on one-time keys and robust chaotic maps," *Computers & Mathematics with Applications*, vol. 59, pp. 3320-3327, 2010.
- [41] 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.
- [42] D. Bouslimi, G. Coatrieux, and C. Roux, "A joint encryption/watermarking algorithm for verifying the reliability of medical images: Application to echographic images," *Computer Methods and Programs in Biomedicine*, vol. 106, pp. 47-54, 2012.

# An Interaction Mining Approach for Classifying User Intent on the Web

Loredana Caruccio, Vincenzo Deufemia, Giuseppe Polese  
Department of Management and Information Technology  
University of Salerno  
84084 Fisciano (SA), ITALY  
{lcaruccio, deufemia, gpolese}@unisa.it

## Abstract

*Predicting the goals of internet users can be extremely useful in e-commerce, online entertainment, and many other internet-based applications. One of the crucial steps to achieve this is to classify internet queries based on available features, such as contextual information, keywords and their semantic relationships. Beyond these methods, in this paper we propose to mine user interaction activities in order to predict the intent of the user during a navigation session. However, since in practice it is necessary to use a suitable mix of all such methods, it is important to exploit all the mentioned features in order to properly classify users based on their common intents. To this end, we have performed several experiments aiming to empirically derive a suitable classifier based on the mentioned features.*

## 1 Introduction

During an Internet navigation session the user performs several actions that can provide hints on his/her future activities. Being able to capture and interpret the hidden goals behind such actions can provide organizations with a competitive advantage. For instance, e-commerce organizations might predict user needs, and advertise the products that users will most likely buy. Thus, multimedia catalogues, web and information retrieval systems need to embed search engines capable of capturing user intent, which is the focus of user intention understanding (UIU) research area [26].

Many approaches for user intent understanding are based on the analysis of search behaviors [4, 6, 7, 8, 10, 17], such as clicked URLs [31] and submitted queries. Most of them aim to capture semantic correlations among search behaviors of the same user, in order to let search engines produce customized results for each individual user.

Other studies analyzed user interactions with *Search Engine Result Pages* (SERPs) to infer their intent [2, 3, 14, 18, 22, 30]. However, by limiting the analysis to results

contained in a SERP, such methods ignore many important interactions and contents visited from such results. For this reason, some approaches to user behavior analysis focus on user interactions with web pages to infer clues on their interest and satisfaction with respect to the visited contents [1, 16]. Following this trend, in this paper we define a new model for UIU analyzing both interactions with SERP results and those on the visited web pages. The interaction features considered in the proposed model are local page level statistics, that is, they are fine-grained and refer to portions rather than the whole web pages. This provides the basis for a more promising prediction of the user intent, since several experiments with eye-trackers revealed that users analyze web page contents by sections, overlooking those of low interest [27].

Other than interaction features, the proposed model considers additional features, such as query keywords and contextual information, all feeding a classification algorithm to understand user intent. The classification process uses a two-level taxonomy in which the first level defines *navigational*, *informational*, and *transational* types of queries[5], where the last two are further decomposed in the second level [28].

We also provide experimental results highlighting the efficiency of the proposed model for query classification. The proposed set of features has been evaluated with several classification algorithms. To this end, in order to more precisely compare the achieved results, and detect the most promising features, we have introduced a metric to evaluate the performances of the different classifiers.

The rest of this paper is organized as follows. In Section 2, we provide a review of related work. Then, we present the model exploiting interaction features for UIU in Section 3. Section 4 describes experimental results. Finally, conclusions and future work are given in Section 5.

## 2 Related Work

As said above, many approaches for user intent understanding analyze search behaviors of users while they navigate and submit queries through the web [4, 6, 7, 8, 10, 17].

In the early 90s, a pioneer study on search behaviors focused highlighted three browsing strategies [9]: *scan browsing*, in which new information is scanned based on its relevance to changing tasks, representing transient browse goals; *review browsing*, in which, with respect to scan browsing, the scanned information is also reviewed and integrated; finally, *search-oriented browsing*, in which the new information is scanned, reviewed, and integrated based on its relevance to a fixed task.

Morrison *et al.* proposed three taxonomic classification schemes based on user responses to web activities that significantly impacted on their decisions and actions [24]. In particular, they formalized the main questions users ask themselves before starting a search session: *why*, *how*, and *what*, which represent the primary purpose of the search, the method used to find the information, and the content of the searched-for information, respectively, yielding three different taxonomies.

Sellen *et al.* extended previously defined taxonomies by extensively monitoring user search activities [29]. They ended up with a classification dividing web activities into six categories, in which two new types were introduced: *transacting* and *housekeeping*. The first concerns using the web to execute secure transactions targeted at products and services, such as ordering a product or filling out a questionnaire. The second concerns using the web to check or maintain the accuracy and functionality of web resources.

A taxonomy focusing on search queries has been defined by Broder [5], who identified the following three classes of queries based on user's intent: *navigational*, aiming to reach a particular web site, *informational*, aiming to collect information from one or more web pages, and *transactional*, aiming to perform some web-mediated activities, that is, to reach a web site where some service is offered, and from which further interactions are expected.

Kang *et al.* focused on analyzing two types of search activities [18]: *topic relevance*, that is, searching documents guided by a given topic, of informational type, and *homepage finding*, aiming to search main pages of several types of navigational web sites. Starting from common information used by Information Retrieval (IR) systems, such as web page content, hyperlinks, and URLs, the model proposes methods to classify queries based on the two categories mentioned above.

Agichtein *et al.* proposed a predictive model derived from real case studies, which is based on the analysis and the comprehension of user interactions during web navigation [2]. The model tries to elicit and understand user navigation

behaviors by analyzing several activities, such as clicks, scrolls, and dwell times, aiming to predict user intention during web page navigation. Moreover, the study proposes to analyze features that are used to characterize the complex interactions following a click executed on a result page. Such interactions have been exploited also by Guo *et al.*, since they considered them useful to accurately infer two particular tightly correlated intents: search and purchase of products [14].

Lee *et al.* proposed a feature based model for the automatic identification of search goals, focusing on navigational and informational queries [22]. The model has been developed starting from experimental studies on real user navigation strategies, which have primarily revealed the possibility of effectively associating most queries to one of two categories defined within the taxonomy. They observed that queries not effectively associable to a category are usually related to few topics, such as proper nouns or names of software systems. More specifically, the model proposes two features: *past user-click behavior* to infer users intent from their past interactions with results, and *anchor-link distribution*, which uses possible targets of links sharing the same text with the query.

While the strategies described so far aim to classify search queries exclusively using features modeled to characterize search queries, Tamine *et al.* propose to analyze search activities previously performed in the same context [30]. To this end, the set of past queries represents the *query profile*, which helps deriving data useful for inferring the type of the current query.

## 3 A Model for User Intent Understanding

In this section we describe the model and the features used for the classification process. The model of this work is based on the model proposed in [12].

### 3.1 A two-level taxonomy for web queries

During a web search the user has a specific goal, generally described by a textual query, and classifiable in a taxonomy. In what follows, we introduce the two-level taxonomy that will be used in the proposed approach for classifying user queries, which is shown in Figure 1. It synthesizes concepts defined in the taxonomies proposed in [5, 28], which have been refined based on the analysis of the query set used in our experiments.

A brief description of the categories on both levels of the taxonomy follows:

- *Informational*: The aim of this kind of query is to learn something by reading or viewing web pages;

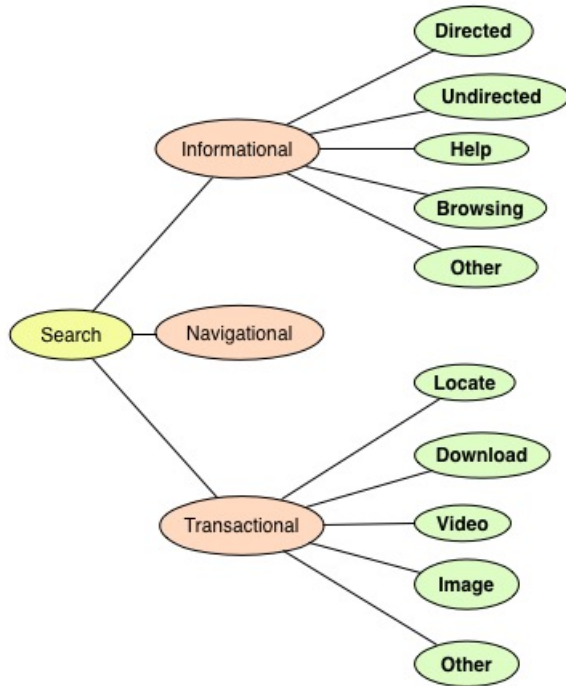


Figure 1: Two-level taxonomy.

- **Directed:** when searching something about a topic;
  - **Undirected:** when the user wants to learn anything/everything about a topic;
  - **Help:** when the user searches for advices, ideas, suggestions, or instructions;
  - **Browsing:** when the user searches something like news, forums, or manuals;
  - **Other:** when the informational query does not fall in any of the categories above.
- **Navigational:** The aim of this kind of query is to reach a known website. The only reason of this kind of search is that it is more convenient than typing the URL, or perhaps if its URL is not precisely known.
  - **Transactional:** The aim of this kind of query is to retrieve a resource available on some web page.
    - **Download:** when the user aims to download a resource;
    - **Video:** when the user aims to watch a video;
    - **Image:** when the user aims to get an image;
    - **Locate:** when the user aims to verify whether or where some real world service or product is offered;

- **Other:** when the query is transactional, but it does not fall in any of the categories above.

### 3.2 Search model: session, search, interaction

Several studies have proven the usefulness of user interactions to assess the relevance of web pages [1, 15, 16, 19], and to determine the intent of search sessions [14, 13]. However, there are additional interactions originating from SERP's contents, such as browsing, reading, and multimedia content fruition, which can potentially provide additional useful clues to UIU.

The proposed approach extends existing predictive models, by mining interactions between users and web pages during a search session. We believe that the actions performed on the visited pages, contrasted to the page format, provide a valuable source of knowledge to predict user intent. As an example, scrolling a web page containing flat text might imply a given user intent, which is different from the scrolling actions performed on framed web pages including both textual and multimedia contents.

In general, a web *session* can be seen as a sequence of search activities aimed at achieving a given goal. When the submitted query does not provide the desired results, the user tries to gradually approach the target, by refining or changing search terms and keywords. A *search* activity can be seen as the combination of the following user actions: submission of a query to a search engine, analysis of search results, and navigation through one or more hyperlinks inside them. The last two types of activities are accomplished by means of several types of *interactions*, which include mouse clicks, page scrolling, pointer movements, and text selection. If combined with features such as dwell time, reading rate, and scrolling rate, such interactions allow us to derive an implicit feedback of user experience with the web pages [12].

The proposed approach prescribes a fine-grained analysis of the traced interactions between users and web pages. Indeed, user interaction analysis is restricted to portions of web pages, e.g., blocks of text, images, multimedia content, which can have a variable length. The use of *subpage*-level analysis provides additional information in the assessment of the user interactions with respect to a global analysis of the entire page.

The data concerning user interactions during web navigation have been encoded into features, which are used by predictive models to characterize user behaviours. We organize the set of features into the following categories: *query*, *search*, *interaction*, and *context*.

**Query.** These features are derived from characteristics of a search query such as keywords, the number of keywords, the semantic relations between them, and other characteristics of a search or an interaction.

**Search.** These features act on the data from search activities such as: results, time spent on SERP, and number of results considered by the user. The *DwellTime* is measured from the start of the search session until the end of the last interaction originated by the same search session. The reaction time, *TimeToFirstInteraction*, is the time elapsed from the start of the search session and the complete loading of the first selected page. Other features dedicated to interactions with the results are *ClicksCount*, which is the number of visited results, and *FirstResultClickedRank*, determining the position of the first clicked result.

**Interaction.** These features act on the data collected from interactions with web pages and subpages, taking into account the absolute dwell time, the effective dwell time, all the scrolling activities, search and reading activities. The *DwellRate* measures the effectiveness of the permanence of a user on a web page, while the reading rate *ReadingRate*, measures the amount of reading of a web page [12]. Additional interactional features are: *ViewedWords*, the number of words considered during the browsing, *UrlContainsTransactionalTerms*, which verifies if the URL of the page contains transactional terms (download, software, video, watch, pics, images, audio, etc.), *AjaxRequestsCount*, which represents the number of AJAX requests originated during browsing.

**Context.** These features act on the relationship between the search activities performed in a session, such as the position of a query in the sequence of search requests for a session.

### 3.3 Logging Web Interaction Data

In the following we describe the module YAR we implemented for logging the user interaction actions, from which we derive the set of features contributing to the mining of user intent.

The architecture of the YAR system is depicted in Figure 2. It is based on a client/server model, where data concerning user interactions are collected on the client side by the *Logger*, and evaluated on the server side through the *Log Analyzer*. The *Logger* is responsible for “being aware” of the user’s behavior while s/he browses web pages, and for sending information related to the captured events to the server-side module. The latter is responsible for analyzing the collected data and for applying metrics to derive the candidate taxonomy categories.

The *Logger* is based on the *AJAX* technology [25] to capture and log user’s interactions with a web system through a pluggable mechanism, which can be installed on any web browser. Thus, it does not require modifications to the web sites, or any other legacy browser extensions.

## 4 Experiments

In this section we describe the dataset constructed for evaluating the proposed approach and the results achieved with different classification algorithms. In the following, we first provide an overview on the used evaluation metrics and the considered subsets of features, then experimental results are presented.

### 4.1 Experiment Setup

In order to build the dataset for evaluating the proposed model we recruited 31 participants, whose profiles are described in Table 1. For each participant the table shows the gender (18 males vs. 13 females), the age (ranging from 20 to 65 ages), and their experience in using the Web (ranging from 1 to 23 years). Since age, education, and Web experience might significantly influence the approach to Web search, we have tried to involve a balanced mix of profiles, in order to gain unbiased conclusions. Thus, we involved people with heterogeneous ages and web experience; similar considerations apply for education, even though the majority of them have a computer science or technical background (18 out of 31).

Gender	Age	Education	Web exp. (yrs)
M	36	Tech. High School Diploma	13
M	65	Tech. Professional Qualification	1
F	32	MSc in Graphics	11
F	59	Accountant Qualification	10
M	31	MSc in Computer Science	20
M	24	BSc in Computer Science	11
M	23	Undergrad. student in Biology	15
F	23	BSc in Biology	16
F	25	BSc in Computer Science	12
M	60	Tech. High School Diploma	23
M	25	BSc in Computer Science	13
M	27	BSc in Computer Science	10
M	24	Undergrad. student in Political Science	10
M	25	Undergrad. student in Computer Science	14
M	25	Undergrad. student in Computer Science	15
M	25	BSc in Computer Science	10
M	25	BSc in Computer Science	10
M	25	Grad. student in Computer Science	7
F	20	Undergrad. student in Linguistics	5
F	22	Undergrad. student in Civil Eng.	10
M	29	Grad. student in Computer Science	14
M	24	Grad. student in Computer Science	15
F	25	BSc in Computer Science	15
F	27	BSc in Education	11
M	25	BSc in Computer Science	9
F	33	MD specializing in Pediatrics	10
M	34	Grad. student in Microelectronics Eng.	20
F	25	Grad. student in Linguistics	8
F	25	Undergrad. student in Sociology	10
F	24	High School Diploma in Arts	8
F	27	BSc in Education	11

Table 1: Profiles of participants to the evaluation.

All participants were requested to perform ten search sessions organized as follows:



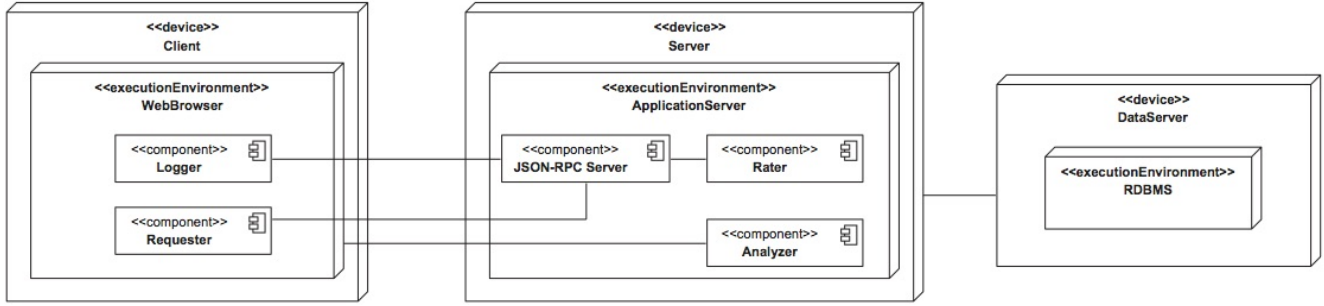


Figure 2: The YAR System Architecture.

- four guided search sessions;
- three search sessions in which the participants know the possible destination web sites;
- three free search sessions in which the participants do not know the destination web sites.

In the following, we list the goals of the guided search sessions:

- the London Metro map image;
- the official video of U2 song Vertigo;
- the e-mail address of an administrative office at the University of Salerno;
- the size of *Mona Lisa*, the famous painting of Leonardo.

This led to 129 sessions and 353 web searches, which were recorded and successively analyzed in order to manually classify the intent of the user according to the two-level taxonomy in Figure 1. Starting from web searches, 490 web pages and 2136 sub pages were visited. The interaction features were logged by the YAR plug-in for Google Chrome/Chromium [12].

#### 4.1.1 Feature subsets

In order to analyze the effectiveness of the considered features, we have grouped them into several subsets:

- **All**: subset of all the proposed features: *query*, *search*, *interaction*, and *context*;
- **Query**: subset of all the features related to *queries*;
- **Search**: subset of all the features related to *search* and *context*;
- **Interaction**: subset of all the features related to *interactions*;

- **Query+Search**: subset of the features derived as union from *Query* and *Search*. The goal is to evaluate the effectiveness of query classification by using the features considered in other studies [16, 2, 14];
- **Transactional**: subset of all the features related to interactions over transactional queries *ViewWords*, *AjaxRequestsCount*, *ScrollingDistance*, *ScrollingCount*, and *UrlContainsTransactionalTerms*. The goal here is to evaluate the classification of transactional queries by adopting more specific features;
- **Interaction–Transactional**: subset derived by the exclusion of the transactional features from the set *Interaction*. The goal here is to evaluate the effectiveness of the classification of transactional queries by comparing results achieved with interaction features to those achieved by excluding transactional features.
- **All–Transactional**: subset derived by the exclusion of the transactional features from the set *All*. The goal here is to evaluate the effectiveness of the classification of transactional queries by comparing results achieved with all features to those achieved by excluding transactional features.

The set of features captured during the search sessions are available for download<sup>1</sup>.

#### 4.1.2 Classifiers

We considered three classifiers to evaluate the proposed feature model: SVM [11], CRF [20], and LDCRF [23].

In the context of query classification, SVM assumes that the queries in a user session are independent, Conditional Random Field (CRF) considers the sequential information between queries, whereas Latent Dynamic Conditional Random Fields (LDCRF) models the sub-structure of user sessions by assigning a disjoint set of hidden state

<sup>1</sup><https://goo.gl/ypH2ij>

variables to each class label. They have been configured as follows:

1. **SVM.** We used MSVMpack [21] as the SVM toolbox for model training and testing. The SVM model is trained using a linear kernel and the parameter C has been determined by cross-validation.
2. **CRF.** We used the HCRF library<sup>2</sup> as the tool to train and test the CRF model. For the experiments we used a single chain structured model and the regularization term for the CRF model was validated with values  $10^k$  with  $k = -1 \dots 3$ .
3. **LDCRF.** We used the HCRF library for training and testing LDCRF model. In particular, the model was trained with 3 hidden states per label, and the regularization term was determined by cross-validation to achieve best performances.

#### 4.1.3 Evaluation Metrics

In order to evaluate the effectiveness of the proposed model, we adopted the classical evaluation metrics of Information Retrieval: *accuracy*, *precision*, *recall*, and *F1-measure*, whose definition is given below:

$$\text{Accuracy} = \frac{\#TP + \#TN}{\#TP + \#TN + \#FP + \#FN}$$

$$\text{Precision} = \sum_{\text{Category}(i)} \left( \frac{\# \text{correctly classified queries}}{\# \text{classified queries}} \times \frac{\# \text{category queries}}{\# \text{total queries}} \right)$$

$$\text{Recall} = \frac{\# \text{correctly classified queries}}{\# \text{total queries}}$$

$$\text{F1 - measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

In addition, in order to simplify the comparison of performances for the different classifiers, in what follows, we apply more a suitable metrics. In fact, in order to evaluate the effectiveness of a classifier, the features need be grouped into several subsets, and executing each classifier by considering each subset of features once. Then, to contrast performances of classifiers we need to compare the results achieved on different pairs (*classifier, feature subset*). Thus, in our case, we need to compare 336 values since we have 3 classifiers, each executing on 8 feature subsets, for each of which we need to calculate 14 parameters.

<sup>2</sup><http://sourceforge.net/projects/hcrf/>

The proposed metrics is based on the mean squared error (*MSE*), which is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2 \quad (1)$$

where  $\hat{x}_i$  is the *i-th* predicted value, while  $x_i$  is the *i-th* correct value. For our purposes, we used MSE calculated on the *accuracy* measure. Thus, given the vector of *accuracy* values  $\hat{a}$ , the definition of the *Accuracy Mean Squared Error* (*AMSE* or  $MSE(\hat{a})$ ) is

$$AMSE = \frac{1}{n} \sum_{i=1}^n (\hat{a}_i - a_i)^2 \quad (2)$$

where  $a_i$  is equal to 1. We computed a relative AMSE value for each pair *Classifier-SubsetFeatures*.

AMSE is able to gain knowledge about the performance of the classifiers and the subsets of features, and how they influence each other.

Let

$$I = \{All, Query, Search, Transactional, Interaction, Query+Search, All-Transactional, Interaction-Transactional\}$$

$$J = \{CRF, LDCRF, MSVM\}$$

be the set of *SubsetFeatures* and the set of compared *Classifiers*, respectively. We designed four AMSE-based values for gaining knowledge about the classifier performances:

- **Global:** it returns the pair *Classifier-SubsetFeatures* with the minimum AMSE

$$\min \left( AMSE_{i,j} \right) \quad \forall i \in I, j \in J \quad (3)$$

It is useful to catch the best performance;

- **Subsets:** it predicts the classifier better performing on each subset of features.

$$\min_{j \in J} \left( x_{i,j} \right) \quad \forall i \in I \quad (4)$$

so that we can easily derive the best performing pairs *Classifiers-SubsetFeatures*;

- **FeaturesBehavior:** it computes the average behavior for each subset of features

$$\frac{1}{|J|} \sum_{j \in J} AMSE_{i,j} \quad \forall i \in I \quad (5)$$

allowing us to gain knowledge about the subsets of features on which a classifier performs better;

- **ClassifiersBehavior:** it computes the average behavior for each *Classifier*

$$\frac{1}{|I|} \sum_{i \in I} AMSE_{i,j} \quad \forall j \in J \quad (6)$$

allowing us to detect the best performing classifiers.

## 4.2 Results

In order to simulate an operating environment, 60% of user queries were used for training the classifiers, whereas the remaining 40% were used for testing them.

Figures 3-6 report the statistics based on the CRF classifier, which give an idea of how complex is the evaluation with conventional measures. On the other hand, Figures 7 and 8 provide a synthetic overview with all the used classifiers, which appears to be more effective. In particular, Fig. 7 highlights that MSVM achieves the best average performance, followed by CRF, which has almost the same MSVM value. Notice that, the lesser the AMSE value the better are the performances, since AMSE is an error measure.

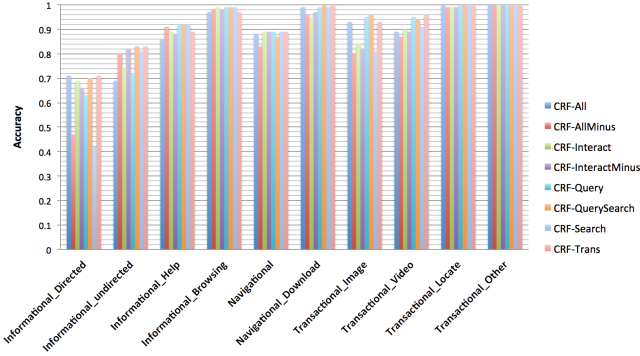


Figure 3: Accuracy obtained with the CRF model.

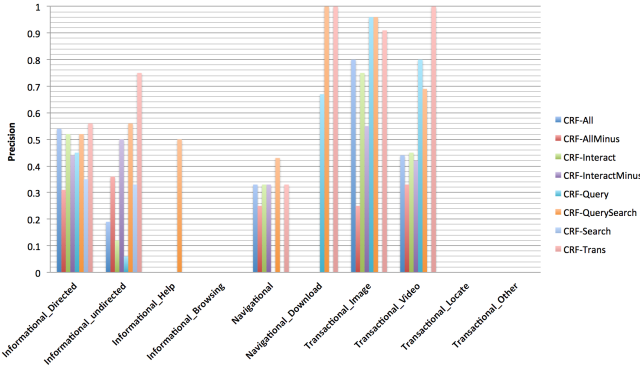


Figure 4: Precision obtained with the CRF model.

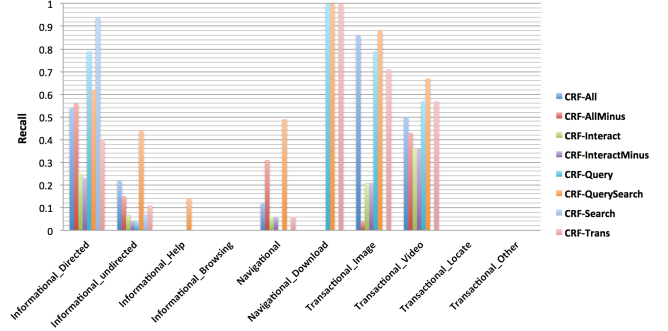


Figure 5: Recall obtained with the CRF model.

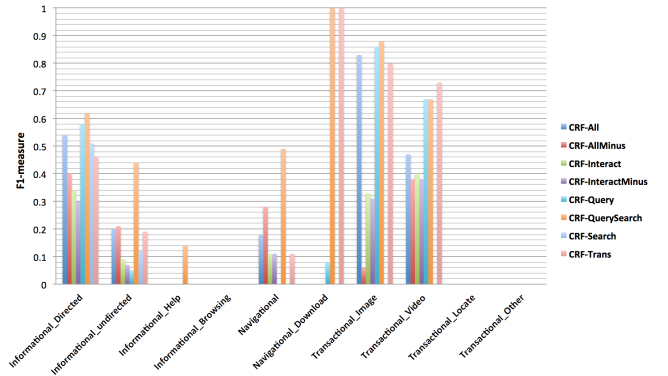


Figure 6: F1-measure obtained with the CRF model.

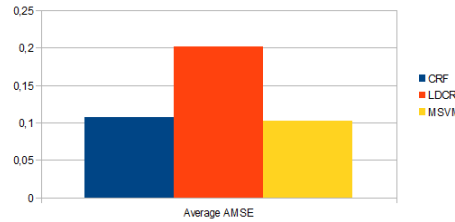


Figure 7: AMSE *ClassifiersBehavior* values of classifiers.

Fig. 8 shows the AMSE values obtained for the different subsets of features. In particular, *Transactional* achieves the best performances, followed by *Query*, and *Query+Search*. Instead, the worst performances are given by the subset *All*, since it yields the maximum value for AMSE.

The *AMSE global* values in Fig. 9 highlight that the best pair (classifier, features) has been *CRF-Transactional*. The *Transactional* features have shown a good discriminative power, since there are two classifiers achieving the best AMSE based on them. Conversely, the *AMSE Subsets* values shown in Fig. 10 highlight that the *CRF* classifier is the one showing best performances for most subsets of features, since it outperforms the other classifiers on 4 out of 8 subsets of features).

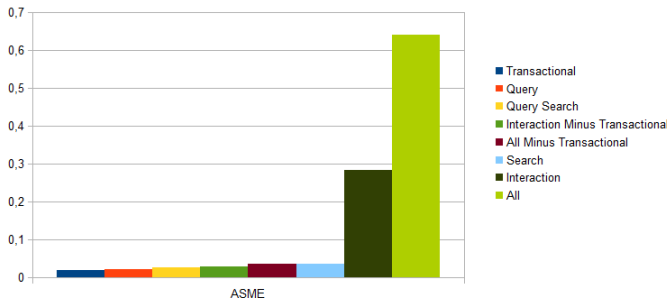


Figure 8: *AMSE Subset* values for each subset of features.

Global		
Classifiers	SubsetFeatures	AMSE
CRF	Transactional	0,01446
LDCRF	Transactional	0,02086
MSVM	Query	0,02105

Figure 9: *AMSE Global* values for each classifier.

Subsets		
SubsetFeatures	Classifiers	AMSE
All	MSVM	0,58953
All Minus Transactional	LDCRF	0,02903
Interaction	CRF	0,02262
Interaction Minus Transactional	CRF	0,02204
Query	LDCRF	0,02102
Query Search	CRF	0,01475
Search	MSVM	0,02391
Transactional	CRF	0,01446

Figure 10: Lower *AMSE FeaturesBehavior* values for the considered classifiers.

### 4.3 Discussion

From the experimental results we can conclude that the use of interaction features to mine the intent of the user during search sessions is a promising approach. In fact, we have observed best classification performances when using the transactional features, which embed a considerable amount of interaction actions. However, we have observed best performances when interaction features are analyzed in a specific context (transactional), rather than in generic contexts. This is due to the fact that is easier to mine the intent when the interaction is performed on a specific type of web page. Vice versa, when no specific assumption can be made on the structure of the web page, each interaction action can convey many different meanings. For instance, a scrolling action yields different interpretations if it is performed on a plain text web page with respect to framed pages like those of online magazines.

## 5 Conclusions and Future Work

We have proposed a model for UIU focusing on both interactions with SERP results and on the visited web pages. The model predicts user intent by exploiting local page level statistics, and additional features, such as query keywords and contextual information, all feeding a classification algorithm. The latter uses a two-level taxonomy, defining *navigational*, *informational*, and *transational* query types at first level [5], furtherly decomposing the last two types at the second level [28].

We have also empirically compared the performances of main classifiers, and have devised a suitable metrics to detect the best classifier and the best subset of features. In particular, the experiments highlighted that the MSVM classifier achieves the best average performances, followed by the CRF classifier, whereas the *Transactional* features outperformed the others, followed by *Query* and *Query + Search* feature.

In the future, other than investigating the possibility of monitoring additional features, we would like to investigate machine learning approaches for inferring a suitable predictive model from a larger set of training data. Moreover, we need to perform a precise classification of web site types, in order to customize the interpretations of user interactions on the specific type of web page. We also need to perform similar investigations in order to tailor the interpretation of the user interaction actions to the type of client device. In fact, smartphones and tablets use different interaction paradigms for which we need further experiments to understand how to mine user interactions for intent understanding purposes.

## References

- [1] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *Proc. of the International Conference on Research and Development in Information Retrieval, SIGIR'06*, pages 19–26. ACM, 2006.
- [2] E. Agichtein, E. Brill, S. Dumais, and R. Ragno. Learning user interaction models for predicting web search result preferences. In *Proc. of the International Conference on Research and Development in Information Retrieval, SIGIR'06*, pages 3–10. ACM, 2006.
- [3] A. S. B. J. Jansen, D. L. Booth. Determining the user intent of web search engine queries. In *Proceedings of the International Conference on World Wide Web, WWW'07*, pages 1149–1150. ACM, 2007.
- [4] S. Beitzel, E. Jensen, O. Frieder, D. Lewis, A. Chowdhury, and A. Kolcz. Improving Automatic Query Classification via Semi-Supervised Learning. In *Proceedings of the fifth IEEE International Conference on Data Mining*, pages 42–49, 2005.
- [5] A. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.
- [6] H. Cao, D. H. Hu, D. Shen, D. Jiang, J.-T. Sun, E. Chen, and Q. Yang. Context-aware query classification. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '09*, pages 3–10, New York, NY, USA, 2009. ACM.
- [7] H. Cao, D. Jiang, J. Pei, E. Chen, and H. Li. Towards context-aware search by learning a very large variable length hidden markov model from search logs. In *Proceedings of the 18th International Conference on World Wide Web, WWW '09*, pages 191–200, New York, NY, USA, 2009. ACM.
- [8] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. Context-aware query suggestion by mining click-through and session data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08*, pages 875–883, New York, NY, USA, 2008. ACM.
- [9] E. Carmel, S. Crawford, and H. Chen. Browsing in hypertext: a cognitive study. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 22:865–884, 1992.
- [10] Z. Cheng, B. Gao, and T.-Y. Liu. Actively predicting diverse search intent from user browsing behaviors. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pages 221–230, New York, NY, USA, 2010. ACM.
- [11] C. Cortes and V. Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, Sept. 1995.
- [12] V. Deufemia, M. Giordano, G. Polese, and G. Tortora. Inferring web page relevance from human-computer interaction logging. In *Proc. of the International Conference on Web Information Systems and Technologies, WEBIST'12*, pages 653–662, 2012.
- [13] Q. Guo and E. Agichtein. Exploring mouse movements for inferring query intent. In *Proceedings of the International Conference on Research and Development in Information Retrieval*, pages 707–708. ACM, 2008.
- [14] Q. Guo and E. Agichtein. Ready to buy or just browsing? detecting web searcher goals from interaction data. In *Proceedings of the International Conference on Research and Development in Information Retrieval, SIGIR'10*, pages 130–137. ACM, 2010.
- [15] Q. Guo and E. Agichtein. Towards predicting web searcher gaze position from mouse movements. In *Proceedings of the International Conference on Human Factors in Computing Systems, CHI EA'10*, pages 3601–3606. ACM, 2010.
- [16] Q. Guo and E. Agichtein. Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior. In *Proceedings of the International Conference on World Wide Web, WWW'12*, pages 569–578. ACM, 2012.
- [17] Q. He, D. Jiang, Z. Liao, S. C. H. Hoi, K. Chang, E.-P. Lim, and H. Li. Web query recommendation via sequential query prediction. In *Proceedings of the 2009 IEEE International Conference on Data Engineering, ICDE '09*, pages 1443–1454, Washington, DC, USA, 2009. IEEE Computer Society.
- [18] I. Kang and G. Kim. Query type classification for web document retrieval. In *Proceedings of the Conference on Research and Development in Informaion Retrieval, SIGIR'03*, pages 64–71. ACM, 2003.
- [19] D. Kelly and J. Teevan. Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum*, 37(2):18–28, 2003.
- [20] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the International Conference on Machine Learning, ICML'01*, pages 282–289, 2001.
- [21] F. Lauer and Y. Guermeur. MSVMpack: a multi-class support vector machine package. *Journal of Machine Learning Research*, 12:2269–2272, 2011.
- [22] U. Lee, Z. Liu, and J. Cho. Automatic identification of user goals in web search. In *Proceedings of the International Conference on World Wide Web, WWW'05*, pages 391–400. ACM, 2005.
- [23] L.-P. Morency, A. Quattoni, and T. Darrell. Latent-dynamic discriminative models for continuous gesture recognition. In *Proceedings of IEEE Conference Computer Vision and Pattern Recognition, CVPR'07*, pages 1–8, 2007.
- [24] J. Morrison, P. Piroli, and S. C. SK. A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. In *Extended Abstracts on Human Factors in Computing Systems, CHI EA'01*, pages 163–164. ACM, 2001.
- [25] G. Murray. Asynchronous javascript technology and XML (ajax) with the java platform. <http://java.sun.com/developer/technicalArticles/J2EE/AJAX/>, 2006.
- [26] H. Nguyen. Capturing user intent for information retrieval. In D. L. McGuinness and G. Ferguson, editors, *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications*

*of Artificial Intelligence, July 25-29, 2004, San Jose, California, USA*, pages 997–998. AAAI Press / The MIT Press, 2004.

- [27] J. Nielsen. F-shaped pattern for reading web content, April 2006. <http://www.useit.com/articles/f-shaped-pattern-reading-web-content/>.
- [28] D. Rose and D. Levinson. Understanding user goals in web search. In *Proceedings of the International Conference on World Wide Web*, WWW'04, pages 13–19. ACM, 2004.
- [29] A. Sellen, R. Murphy, and K. Shaw. How knowledge workers use the web. In *Extended Abstracts on Human Factors in Computing Systems*, CHI'02, pages 227–234. ACM, 2002.
- [30] L. Tamine, M. Daoud, B. Dinh, and M. Boughanem. Contextual query classification in web search. In *Proceedings of International Workshop on Information Retrieval Learning, Knowledge and Adaptability*, LWA'08, pages pp. 65–68, 2008.
- [31] J. Wang, A. P. de Vries, and M. J. T. Reinders. A user-item relevance model for log-based collaborative filtering. In M. Lalmas, A. MacFarlane, S. M. Rger, A. Tombros, T. Tsikrika, and A. Yavlinsky, editors, *ECIR*, volume 3936 of *Lecture Notes in Computer Science*, pages 37–48. Springer, 2006.



# MOSAIC+: tools to assist virtual restoration

Daniel Riccio  
University of Naples Federico II  
80121 Campi Flegrei  
Naples, Italy  
daniel.riccio@unina.it

Sonia Caggiano  
Master of Architecture and PhD  
in Digital Painting Restoration  
Salerno, Italy  
soniacaggiano@yahoo.it

Maria De Marsico  
Sapienza University of Rome  
Via Salaria, 113  
Rome, Italy  
demarsico@di.uniroma1.it

Riccardo Distasi, Michele Nappi  
University of Salerno  
84084 Fisciano  
Salerno, Italy  
{distasi, mnappi} @unisa.it

**Abstract**—In many cases, virtual restoration is the only way to have an idea of the original appearance of an artwork. In particular, in the archeological field, it is useful to both assist and guide the operator in physical reconstruction, and to provide to the final visitor a complete vision of the artwork even though the original is damaged or lacks some parts. In this work, we propose a set of tools useful in two different restoration steps. The first one helps the expert to carry out reconstruction of fragmented artifacts in an easier and more effective way. The second provides a view of the artifact after virtually eliminating a craquelure.

**Keywords**—*virtual restoration; mosaic fragmentation; painting craquelure; color indexing; shape indexing*

## I. INTRODUCTION

We start by the statement of the practical problem that we want to contribute to solve. It happens, either fortunately during archeological excavation campaign or unfortunately after destructive events, like earthquakes, to recover ruins originally covered by frescos. In the former case, the scene depicted is not known, in the latter pictures may exist, but the situation is not that better if fragmentation is dense. In both cases the automatic re-composition is very hard to carry out and requires patience and a lot of careful, slow work. Even worse, the inherent difficulty of the task is increased by its possibly extreme delicacy and by the required caution. The expert often faces a collection of painted fragments that crumble to dust if not handled with the utmost gentleness. On the contrary, reconstructing at least part of the original design would require much manipulation: correctly putting the pieces side by side requires an infinite number of repeated rotations, tentative alignments, and more operations on single pieces as well as batched of already dovetailed ones. In this scenario, the most time consuming action is to repeatedly explore the pool of available fragments to locate a possible candidate piece to join the portion of surface at hand. Each time a piece is touched or

moved it might break, or at least its edge might be further ground away, depending on the materials used to build and coat the old wall. At the end, when one luckily arrives at the end, it is the case that the recovered surface appears as cracked by missing pieces as well as by patchy edges. For the final visitor of the artwork, the artistic experience may be significantly improved by attenuating the visual effect of such irregularities.

This paper presents the advancements in the implementation of a set of tools formerly proposed as Multi-Object Segmentation for Assisted Image reConstruction (MOSAIC) in [1]. MOSAIC+ also includes procedures to attenuate the visual effect of craquelure. Such tools are devised together with field experts to support the work of archaeologists and cultural heritage operators, when reconstructing fragmented (plain) artifacts. No information about the original appearance of the whole artwork is assumed to be available. To this aim, fragment images are inserted in a repository suitably indexed. The system provides the operator with complete workflow from photo-acquisition onwards. In the repository population phase, the fragments are photographed and their captured images are suitably processed and organized. The repository is indexed according to features such as color distribution, shape and texture. Images can be retrieved through query-by-example, using any fragment image as the key. If more results are available, as it is almost always the case, they are displayed to the user from the most to the least similar to the key. The operator can pick returned fragment images, rotate and translate them, and try to dovetail them to reconstruct the original picture, as when solving a puzzle. In most cases, holes will be present and the result will appear as highly fragmented, even in the virtual reconstruction. To further support the operator, as well as to provide an idea to the artwork visitors of its original appearance, once the reconstruction is completed a technique to attenuate craquelure is applied. Actually, this technique is also useful as a preprocessing step during population, before extracting shape and color information, when the fragments present inner craquelure. We applied our techniques on a number of

simulations, and on the real use case of the reconstruction of a fresco from fragments found in the St. Trophimena church in Salerno (Italy).

## II. RELATED WORK

MOSAIC+ includes an automated system for computer-aided reconstruction of jigsaw puzzles. In algorithm literature, puzzles are grouped into two broad categories, that given the different characteristics need quite different approaches: in apictorial puzzles gather pieces of uniform surfaces that do not show figures, so that the only kind of information to be used to guide reconstruction is the fragment shape; in pictorial puzzles, texture and color information is available but, unlike most commercial puzzles, this does not necessarily imply that the solution image is known a priori. This the category of puzzles that restores usually handle.

The available literature offers several approaches to solve both types of jigsaw puzzles. Being the puzzle pictorial or apictorial does not significantly affect the computational complexity of automatic solution methods. In the first case, the classic paper by Freeman and Gardner, which was one of the earliest to tackle the problem of apictorial jigsaw puzzles, demonstrates that an exact algorithmic solution is NP-complete: the computing time is super-polynomial with respect to the problem size [2]. The paper also suggests five fundamental puzzle properties: orientation (unknown a priori), connectivity (presence or absence of internal "holes"), perimeter shape (known/unknown a priori), uniqueness (does the problem admit only one solution?), radiality (topology of fragment juncture). The contours of the fragments are represented as chain codes, and code length is used as a heuristic for search space dimensionality reduction.

Some techniques can provide an approximate solution in a shorter time [3]. Applications are popular mostly in the fields of cultural heritage and ancient document reconstruction. A survey of literature is out of our scope, and in any case we do not tackle the problem of automatic solution, but rather that of relieving the expert by the burden of an extremely long trial-and-error process and from the anxiety of manipulating critically fragile material. Nevertheless, it is worth mentioning some example system devoted to our same application field.

It is to notice that most techniques used in the cultural heritage field address the problem of pictorial puzzles, rather than apictorial. On the other hand, if the artifact to restore is a fresco, it is possible that it does not represent a natural scene but rather a set of repeated geometrical patterns. This latter circumstance may cause the lack of uniqueness of the combination of fragments, or of their automatically computed features at least.

The paper by Papaodysseus et al. facing the problem of reconstructing wall paintings [4] is paper particularly interesting for the present discussion. The focus is on the real-world issues that arise when dealing with a fresco: lack of information about the original content of the painting, possible non-uniqueness when geometric shapes are involved, and especially non-connectedness arising from the presence of very small fragments that are not available to the problem solver.

The technique for finding the correct correspondences deals with missing information using local curve matching.

Brown is among the authors of a semi-automatic system, used for reconstructing frescoes at the Akrotiri site in Thera, present Santorini in Greece [5]. The system carries out 3D data acquisition, but most results are obtained via 2D image feature extraction.

Brown is also among the experts dealing with the restoration of the Roman site in Tongeren (*Atuatuca Tungrorum*), the oldest town in Belgium. The site contained a number of artifacts that have been at least partially reconstructed, e.g., the Vrijthof Wall Decoration 1 [6]. The fragments were preliminarily acquired by an ad hoc 3D scanner. The shapes extracted by fragment image processing were matched by an ad-hoc software. Notwithstanding the efforts to refine image processing procedures to allow less expensive equipment, the total cost remained quite high. The obtained increased number of true matches is not that satisfactory, considering the actual numbers: 3 true matches were found manually, and in the same situation the system proposed 6103 tentative matches that became 17 true matches after human screening.

From results reported in literature, it is clear that better solutions can be obtained by fully exploiting all the possible available information. For instance, Chung et al. use color [7], while Sagiroglu and Ercil use texture [8]. It is to say that, in most cases, actual testing has been limited to problems involving a relatively small number of fragments. A completely different approach is presented by Nielsen et al. [9]. Fragment shape is neglected altogether. Rather, the method relies on features of the whole represented pictorial scene. The reported results for this technique show low error margins: the solution to a 320-fragment problem only had 23 pieces out of place—an error margin of 7.2% obtained by using only color and texture information. However, this approach implies to have at least a partial knowledge of the represented scene.

We can conclude by recognizing that the virtual reconstruction of pictorial fragments is an intrinsically hard problem. Approximate solutions as well as ones were the human-in-the-loop strategy is assumed are often all we can get. Several advanced image processing techniques are being incorporated in most recent systems. The most promising are based on local texture analysis, chrominance analysis and contour analysis on single fragments. When the original appearance is known or can be at least partially inferred, methods based on properties of the whole scene are expected to be quite powerful, and can provide further features to consider. Such techniques can produce representations that allow users to refine the solution progressively.

## III. CRAQUELURE

In the artistic lexicon, craquelure is the pattern of fine and dense "cracks" on the surface of artworks. It can affect different materials, and it can be either due to an intentional characteristic of the production process, or be caused by defects in such process or more frequently by ageing. Though being a more general term, the most popular use refers to paintings, in particular those produced by tempera or oil, where it is causes

mainly by ageing. In particular, in paintings on wood, it first appears following the direction of wood fibers. When referring to ceramics, it usually denotes a special manufacturing, and the term is usually modified as "crackle". Furthermore, it can appear on old ivory carvings, and on painted miniatures on ivory. Being a typical sign of ageing of the pictorial layer, the microscopic analysis of the craquelure (either natural or artificial, either deep or surface) is also used to determine the age and therefore the authenticity of paintings. We will slightly extend the term also to denote irregular and occasional cracking, especially on frescos (Fig. 1(b), cracking deriving from reconstruction with missing (small) parts, and virtual simulation of cracking (Fig. 1(e)). In particular, we will apply techniques to reduce the visual effect of craquelure, both on single fragments and after a virtual reconstruction of a fragmented fresco. Examples of craquelure are shown in Fig. 1.

In order to hide the inter-fragment cracks in a reconstructed image, the craquelure pattern must first be detected, then corrected. The first methods for detection were only half-automated and required human intervention [15], since the operator had to indicate a pixel in each connected component of the craquelure pattern, so that the system could locate a maximal connected region. Subsequent research produced methods that required less human intervention. The approach used by Giakoumis et al. [16] is based on mathematical morphology, as is the work by Spagnolo and Somma [17]. Both approaches require the operator to locate at least a small subset of the crack pattern, in order to train the system to detect cracks with either a light or a dark background. The solution proposed in this paper works with both types of craquelure without the help of a human operator, i.e. the technique adopted works notwithstanding the kind of craquelure.

Correction of the craquelure is performed by inpainting – which is also used, e.g., to remove logos and similar superimposed marks from images, or to visually repair rips in the acquired version of a damaged image. There are several inpainting techniques. The more sophisticated ones aim at preserving the isophotes – curves of equal luminance in the underlying original image – or the luminance gradient in the neighborhood [18, 19, 20]. From the preserved isophote, a Laplacian smoothing or other interpolation can be used to extend the restoration to adjacent pixels. A simpler option is that of applying a spatial convolution mask to diffuse the surrounding pixels over the crack [21]. The tradeoff is between more accurate results in the case of isophotes preservation - or gradient-preserving techniques, vs. reduced computing time in the case of simple geometric processing.

#### IV. MOSAIC+ REPOSITORY

Mosaic+ can be classified among the toolsets for Jigsaw pictorial puzzle solving, where texture and color information is available but not the information about the original picture. Color and texture information is used together with shape information. As mentioned above, our proposal relies on a human-in-the-loop approach, i.e. it was expressly designed to support archaeologists and restorers facing fresco recomposition from fragments, and it does not implement an automatic approach. The aim is not to perform a completely automatic reconstruction, but rather to relieve the expert from

most of the burden and anxiety implied by reordering fragile fragments and grouping them in similar clusters.

MOSAIC+ is composed of a number of modules, which implement different procedures. The first set of modules implements a procedure for image acquisition and processing. The result is a catalogued repository of the single fragments, which are clustered according to their texture/color and shape features. This preliminary grouping allows a quite quick answering to user queries, so that reconstruction is made easier, quicker and more effective. The interface module of the application provides a virtual workbench. Among the other actions, the user can virtually perform the actions that would have been performed on real fragments in a real reconstruction attempt, i.e., rotate, translate and search for similar fragments. In particular, a query engine allows searching the archive for relevant fragments, while the manipulation interface allows the user to manipulate them virtually to attempt recovery of the broken picture.

##### A. Fragment image acquisition

In the image acquisition phase, the physical fragments are laid in a white tray, whose bottom is covered by a dark foam to reduce reflexions. The distance among fragments must be sufficient to allow the following image segmentation to insulate the single pieces. The tray is placed inside a box for photographic acquisition, which is closed by a white curtain and bears two lateral spotlights. Close to the tray, a colorimeter is used by the operator to check for the need for automatic color corrections. The tray is then photographed. For this work, we used an 8-Mpixel Canon camera), orthogonally pointing it from a height of 90 cm.

##### B. Fragment image segmentation

This operation aims at correctly separating each fragment appearing in a same image, so that individual features can be extracted from each of them. Segmentation entails two steps. The image is first binarized and turned into B/W with no shades of gray. As can be expected, in our case no single threshold value is effective across all trays, unless some pre-processing occurs to enhance the image color separability. Too low values fail in separating pieces, while too high ones produce "holes" inside pieces. In extreme cases the piece may even come out as two separate fragments—an error that is quite hard to correct later. Therefore, the process of binarization that we carry out first entails pre-processing the raw image in order to amplify the difference between the brighter pixels (fragments) and the darker ones (the background foam is a dark shade of gray—almost black, but not quite). This pre-processing is described in detail in [1]. At the end of the segmentation process, specific information about the fragments found is computed, namely its area, its perimeter, and its orientation. The obtained binary connected component will be used as a mask to retrieve the fragment from the original image by a pixel-wise logical AND operation, in order to separately extract features from each fragment. Notice that, when putting fragments in the tray to acquire the tray image, fragments from very different groups can lay together, while fragments from the same group may lay in different trays.

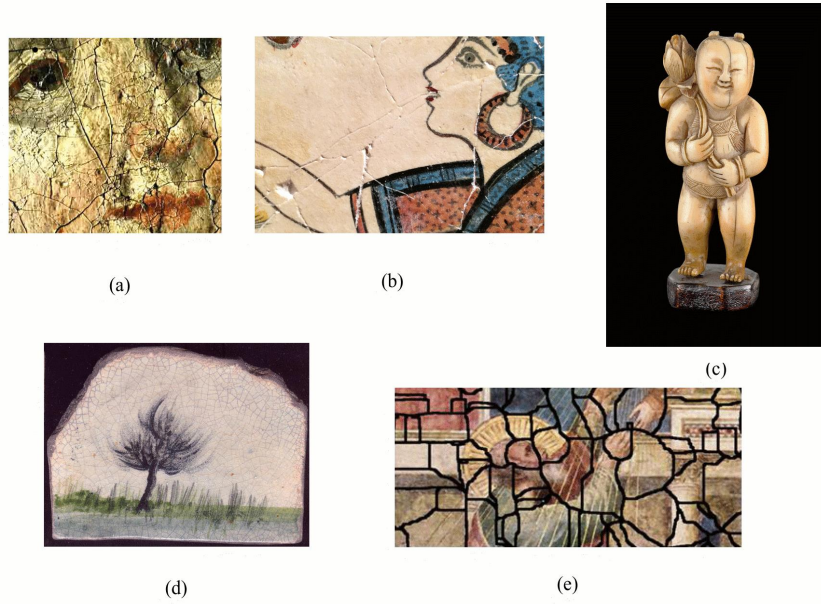


Figure 1. Examples of craquelure: (a) painting; (b) fresco; (c) ivory; (d) raku artwork; (e) virtual craquelure

### C. Fragment feature extraction

Using the masks computed during segmentation, the module for feature extraction insulates the corresponding fragments, so that these can be individually indexed to allow a convenient successive retrieval. Indexing/retrieval are carried out according to the (basic) shape(s) depicted on the fragment, and to a spatiogram, which describes the spatial distribution of colors on the fragment surface [10]. The user can search the fragment catalogue by color, shape, or spatial color distribution, in order to retrieve fragments similar to a given “key” one. Details on feature extraction can be found in [1]

### D. Fragment search

Extracted color information is stored by the fragment spatiogram; comparison is performed by related techniques. Shape information is represented in a more articulated way. Each shape on a fragment is represented as a triple

$$S = \langle v, \omega, c \rangle \quad (1)$$

where  $v = (v_1, \dots, v_7)$  is the vector of the first 7 central moments of the shape (see [11, 12]), and  $\omega$  and  $c$ , are the shape smoothness and mean color value, respectively. A fragment  $F_h$  containing  $s_h$  shapes is therefore characterized by  $s_h$  such triples. We compare two shapes  $S_1 = \langle v_1, \omega_1, c_1 \rangle$  and  $S_2 = \langle v_2, \omega_2, c_2 \rangle$ , by computing their similarity as the normalized dot product of their moment vectors (i.e., the cosine of the angle between them), weighted by the product of their smoothness values:

$$\text{sim}(S_1, S_2) = \omega_1 \omega_2 \frac{v_1 \cdot v_2}{|v_1| |v_2|} \quad (2)$$

Since each fragment contains more shapes, the similarity between two fragments  $F_1$  and  $F_2$  is given by the maximum shape-to-shape similarity.

In the most common case, the query key is represented by a single shape  $S$  and the search goes through each fragment indexed in the database, looking for shapes with high values of similarity to  $S$ . The similarity score assigned to a fragment is the maximum similarity score achieved by a shape it contains. Smaller shapes are discarded as not relevant.

## V. CRAQUELURE ATTENUATION

### A. Craquelure detection

Detecting and isolating craquelure traditionally required slightly different processing depending on the luminance of the cracks – either dark or light background. We adopt here a solution based on Mathematical Morphology (MM). Though MM is most often applied to digital images, it can be also employed on graphs, 3D surface meshes, and other spatial structures. Morphological operations on images entail the use of a suited structuring element, whose shape depends on the problem at hand, and which is used as a probe to draw conclusions on how this shape fits or misses the shapes in the image. Basic operations are *erosion* and *dilation*. Given an image  $I$  and a structuring element  $r$ , the erosion of  $I$  by  $r$  is the locus of points reached by the center of  $r$  when it moves inside  $I$ . The dilation of  $I$  by  $r$  is the locus of points covered by  $r$  when its center moves inside  $I$ . Two further fundamental operations are derived, namely *opening* and *closing*. The opening (closing) of  $I$  by  $r$  is obtained by the erosion (dilation) of  $I$  by  $r$ , followed by dilation (erosion) of the resulting image by  $r$ . Slightly different definitions hold when applying MM to either grayscale or color images. Details on morphological operators are out of the scope of this paper, but the reader can refer to the still extremely valid books [22] and [23]. In the case of MM-based processing, the most common solutions used the *bottom-hat* operator for dark cracks and the *top-hat* operator for light cracks. Bottom-hat of an image  $I$  and a structuring element  $r$  is

the difference between the closure of  $I$  by  $r$  and  $I$ , while top-hat is the difference between  $I$  and its opening by  $r$ .

$$Bhat(I, r) = (I \bullet r) - I \quad (3)$$

$$That(I, r) = I - (I \circ r) \quad (4)$$

The solution proposed in [24] to detect the crack pattern first transforms the image in grayscale, then uses the sum of bottom-hat and top-hat – namely, the difference between closure and opening. This operator is able to detect and isolate the cracks whether they have a dark or light background.

$$Bhat(I, r) + That(I, r) = (I \bullet r) - (I \circ r) \quad \dots \quad (5)$$

The difference between closing and opening returns a grayscale image where the points of maximum and minimum luminance are made more evident than the rest of the image. However, lighter or darker brush traits provide false positives, therefore the image undergoes a thresholding operation to eliminate such false positives and create a map of areas with cracks. From the analysis of the histograms and the values of mean, standard deviation and modal value it comes out that pixels corresponding to these false positives have the gray values less than those that identify the true cracks. The threshold value is the sum of the mean value of the image pixels and the standard deviation. After thresholding, pixels that were discarded but have very similar values to the returned ones are added again, to obtain a more complete map. With the same goal, a dilation operation is finally performed. Differently from other approaches in literature, the solution we propose works indifferently with either light and dark craquelure, or with a mixture of them, without needing a human operator identifying the (possibly local) kind of craquelure in advance.

### B. Craquelure correction

The resulting craquelure map is fed as input to the correction phase. Correction is performed by an inpainting method partly based on the Fast Marching Method illustrated in Telea's work [18]. The method fills the gaps with textures; however, in actual use cases the size of realistically correctable cracks seldom justifies texture creation, so a slightly different approach has been chosen: rather than creating a texture to fill the gap, the surrounding pixels are stretched by interpolation. This is similar to what an actual restorer does when repairing physical paintings – extending the remaining paint over the cracks by patient stretching.

## VI. EXPERIMENTS AND RESULTS

For our experiments we used both true fresco fragments found in the St. Trophimena church in Salerno (Italy), and a number of virtually cracked images.

A first consideration that is worth mentioning is that the measure of the structuring element used in MM depends on the thickness of the cracks, which in the real fragments are usually thinner than in the artificial images that we used. Therefore the structuring element is  $3 \times 3$  for fresco fragments, and  $9 \times 9$  for virtually cracked images. We denote by  $I_{MM}$  the image resulting from applying the above mentioned morphological operations

to the original image  $I$ .  $I_{MM}(x, y)$  is the value of pixel in position  $(x, y)$ . The formula used to detect cracks and create map  $M$  is:

$$M(x, y) = \begin{cases} 1 & \text{if } I_{MM}(x, y) > (\text{mean}(I_{MM}) + \text{std}(I_{MM})) \dots \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

By performing the experiments we also noticed that in large images (the artificial ones) the algorithm has difficulty in detecting cracks on the whole surface. We hypothesized that this depended on the fact that, being the images very large and variegated, the mean value of the entire image aggregates too much information and certain areas become unrepresentative. Therefore we decided to apply the algorithm to image patches. In other words, we divide the image into square patches of size  $N \times N$  and apply the algorithm to each of them, so rearranging the image of cracks as a collage of all the processed patches. In this way, each patch has its own average value, which is more representative and makes the detection of the cracks more accurate. The result, in fact, is far much acceptable.

We report the results of the following experiments: 1) craquelure detection with different dimensions of the patch and inpainting of the virtually cracked “Assunzione di san Giovanni Evangelista” by Giotto, Cappella Peruzzi cathedral of Santa Croce in Florence, about 1318-1322; 2) true fresco fragments before and after craquelure detection and inpainting; 3) a portion of the true reconstructed fresco before and after inpainting. The quality of results results can be appreciated visually. Fig. 2 shows the original image and the virtually cracked one of “Assunzione di San Giovanni Evangelista”. Fig. 3-6 show the results of our procedure with patches of increasing size. It can be readily seen that as the patch size increases the inpainting procedure repairs the craquelure with less residual holes, and the result visually improves approaching the original appearance. This experiment is quite stressing due to the thickness of the craquelure pattern.

Fig. 7 and Fig. 8 show the images of two real fresco fragments (original image, craquelure map, inpainting result). As mentioned above, the best size of the patch to use depends on the resolution of the input. For fragments, all of which are relatively small, we used only  $64 \times 64$  patches.

Finally, Fig. 9 shows the image of a true part of the fresco rebuilt. Since the reconstructed part is bigger, we tested three possible choices:  $64 \times 64$ ,  $128 \times 128$  and  $256 \times 256$ . Fig. 10 and 11 show the results for  $64 \times 64$  and  $256 \times 256$  patches.

## VII. CONCLUSIONS

We presented MOSAIC+ (Multi-Object Segmentation for Assisted Image reConstruction), the evolving version of a system providing a set of tools to support the real as well as virtual reconstruction of fragmented pictorial artworks. We aim both at supporting the delicate work of experts, by facilitating the reconstruction of fragments, and at enhancing the experience of a visitor. The extraction of relevant features related to color and shape allows cataloging and indexing of the fragments, which support queries for similar pieces.





Figure 2. Examples of virtual craquelure: original fresco (left); virtual craquelure (right)



Figure 3. Results with 32×32 patches: (left) craquelure map; (right) inpainting



Figure 4. Results with 64×64 patches: (left) craquelure map; (right) inpainting



Figure 5. Results with 128×128 patches: (left) craquelure map; (right) inpainting





Figure 6. Results with 256×256 patches: (left) craquelure map; (right) inpainting

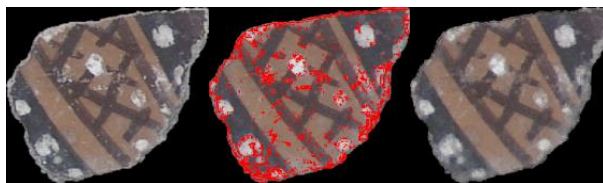


Figure 7. Fragment n. 00054-15 original (left), craquelure map (centre), inpainting (right)

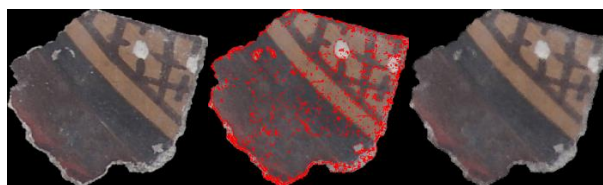


Figure 8. Fragment n. 00054-17 original (left), craquelure map (centre), inpainting (right)



Figure 9. A portion of reconstructed fresco



Figure 10. Virtual restoration of the part in Fig. 9 using 64x64 patches: (left) craquelure map and (right) inpainting



Figure 11. Virtual restoration of the part in Fig. 9 using 256x256 patches: (left) craquelure map and (right) inpainting

The results of the comparison with the stored virtual fragments are sorted by similarity to the query key, and this can speed up the manual reconstruction process significantly. We also experimented the effect obtained by virtually restoring craquelure, i.e., the presence of crack patterns due to age as well as to other negative factors. The system has been tested both via computer simulations and on a real case. The examples reported visually underline the quality that is possible to achieve in virtual restoration of the artwork appearance. This is deemed to improve the artistic experience of both experts and occasional visitors. Our future work will entail the implementation of better inpainting procedures, strategies for automatic setting of the right parameters according to the kind of artwork (size of the structuring pattern as well as of the image patches) and the addition of further tools.

## REFERENCES

- [1] S. Caggiano, M. De Marsico, R. Distasi, and D. Riccio. Multi-Object Segmentation for Assisted Image reconstruction. In Proceedings of 4-th International Conference on Pattern Recognition Applications and Methods - ICPRAM 2015, Vol. 2, pp. 100-107 (2015)
- [2] H. Freeman, L. Garder, "Apictorial jigsaw puzzles: The computer solution of a problem in pattern recognition." IEEE Transactions on Electronic Computers 2(EC-13), 118-127 (1964)
- [3] T.S. Cho, S. Avidan, W.T. Freeman, "A probabilistic image jigsaw puzzle solver." In: CVPR, pp. 183-190. IEEE (2010), <http://dblp.uni-trier.de/db/conf/cvpr/cvpr2010.html#ChoAF10>
- [4] C. Papaodysseus, T. Panagopoulos, M. Exarhos, "Contour-shape based reconstruction of fragmented, 1600 bc wall paintings." IEEE Transactions on Signal Processing 6(50), 1277-1288 (2002)
- [5] B. Brown, C. Toler-Franklin, D. Nehab, M. Burns, D. Dobkin, A. Vlachopoulos, C. Dumas, S. Rusinkiewicz, T. Weyrich, "A system for high-volume acquisition and matching of fresco fragments: Reassembling Thera wall paintings." ACM Transactions on Graphics (Proc. SIGGRAPH) 27(3), 1-10 (2008)
- [6] B. Brown, L. Laken, P. Dutrè, L.V. Gool, S. Rusinkiewicz, T. Weyrich, "Tools for virtual reassembly of fresco fragments." In: Proceedings of the 7th International Conference on Science and Technology in Archaeology and Conservations. pp. 1-10. SCITEPRESS (2010)
- [7] M.G. Chung, M. Fleck, D. Forsyth, "Jigsaw puzzle solver using shape and color." In: Proceedings of the 4th International Conference on Signal Processing (ICSP '98). vol. 2, pp. 877-880 (1998)
- [8] M. Sagioglu, A. Ercil, "A texture based matching approach for automated assembly of puzzles." In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06). pp. 1036-1041 (2006)
- [9] T.R. Nielsen, P. Drewsen, K. Hansen, "Solving jigsaw puzzles using image features". Pattern Recognition Letters 14(29), 1924-1933 (2008)
- [10] S.T. Birchfield, S. Rangarajan, "Spatiograms versus histograms for region-based tracking." In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1158-1163 (Jun 2005)
- [11] M. Hu, "Visual pattern recognition by moment invariants." IRE Trans. Inf. Theor. IT-8, 179-187 (1962)
- [12] M. Mercimek, K.G.T.V. Mumcu, "Real object recognition using moment invariants." Sadhana, Academy Proceedings in Engineering Science 30(6), 765-775 (2005)
- [13] D. Comaniciu, P. Meyer, "Mean shift: A robust approach toward feature space analysis." IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 24(5), 603-619 (2002)
- [14] E.D. Demaine, M.L. Demaine, "Jigsaw puzzles, edge matching, and polyomino packing: Connections and complexity." Graphs and Combinatorics 23 (Supplement), 195-208 (2007), special issue on Computational Geometry and Graph Theory: The Akiyama-ChvatalFestschrift.
- [15] M. Barni, F. Bartolini, V. Cappellini, "Image processing for virtual restoration of artworks", Tools for Building Virtual Heritage, IEEE Multimedia, 34-37 (2000)
- [16] I. Giakoumis, N. Nikolaidis, F. Pitas, "Digital image processing techniques for detection and removal of cracks in digitized paintings, IEEE Transactions on Image Processing 15 (1) 178-188, (2006)
- [17] G. Spagnolo, F. Somma, "Virtual restoration of cracks in digitized image of paintings", Journal of physics. Conference series (vol. 249) (2010)
- [18] A. Telea, "An image inpainting technique based on the Fast Marching Method", Journal of graphics tool (vol. 9), 25-36 (2003)
- [19] M. Bertalmio, A. L. Bertozzi, G. Sapiro, "Navier-Stokes, fluid dynamics and Image and Video inpainting", Proc. ICCV 2001, 1335-1362 (2001).
- [20] M. Oliveira, B. Bowen, R. McKenna, Y. Chang, "Fast digital image inpainting", Proc. VIIP 2001, 261-266 (2001)
- [21] A. Gupta, V. Khandelwal, A. Gupta, M. C. Srivastava, "Image processing methods for the restoration of digitized paintings", Thammasat Int. J. Sc. Tech., Vol. 13, No.3, July-September 2008, 66-71 (2008)
- [22] J. Serra. "Image Analysis and Mathematical Morphology", Volume 1. Academic Press (1984)
- [23] J. Serra. "Image Analysis and Mathematical Morphology: Theoretical advances", Volume 2. Academic Press (1988)
- [24] G. Mercadante, "Una proposta di metodologia per il restauro virtuale delle scopolature dei dipinti" (A proposal of a methodology to virtual restoration of cracks in paintings), MSc thesis, Sapienza University of Rome, Information Engineering, Informatics, and Statistics (2012)

# On the Experience of Using Git-Hub in the Context of an Academic Course for the Development of Apps for Smart Devices

Rita Francese,<sup>1</sup> Carmine Gravino,<sup>1</sup> Michele Risi,<sup>1</sup> Giuseppe Scanniello<sup>2</sup> and Genoveffa Tortora,<sup>1</sup>

**Abstract**—In this paper, we present the experience we gained in a Mobile Application Development course for Computer Science students at the University of Salerno. The course foresaw a project work conducted by students organized in teams. The goal of the project work was to design and develop Android-based applications for smart devices. The learning approach was based on collaboration (intra-team) and competition (extra-team). Students cooperated using GitHub as Computer-Supported-Collaborative-Learning tool for the implicit and explicit communication among team members and distributed revision control and management of software artifacts (e.g., source code and requirements models). All the developed applications underwent a final public competition prized by IT managers of national and international companies. IT managers expressed a positive judgment both on the students' competition and on the developed applications for smart devices. Also, the students provided very good feedback on the competition and on the GitHub support.

## I. INTRODUCTION

Internet and mobile applications are converging and from their union a new society is going to appear. According to Gartner [1], the interest of business users and customers in mobile devices and applications is increasingly growing. The digital enterprise becomes a mobile enterprise. In addition, mobile devices offer a rich set of embedded sensors, such as accelerometer, digital compass, gyroscope, GPS, microphone and camera. These sensors enable the production of new applications addressed to a wide variety of domains. In this scenario, the mobile application developer is one of the most demanded and fastest growing IT career.<sup>1</sup> Nevertheless, the development of mobile applications is not an easy task: the developer is required to master a wide range of technologies and capabilities, including programming languages (e.g., Objective C, C++, C# or Java) and operating systems (e.g., Android and iOS) [2], [3], [4], [5].

In this paper, we present the learning experience related to the second edition of the mobile application development course that fosters teamwork and encourages students to explore new ideas. Indeed, students were required to design their applications for smart devices (also simple apps, from here on) by considering the market needs, usefulness, audience, and viability. The course was organized in blended learning modality: the lectures on the Android operating system were given in presence, while students' projects where asynchronously managed by using GitHub, a largely

adopted tool in technology areas that require collaboration and, more recently, also in education [6]. During the analysis and development activities the students exploited the distributed revision control, source code management (SCM), and asynchronous communication functionalities offered by GitHub. The communication among students took place also implicitly<sup>2</sup> through the developed software artifacts (e.g., software models). The GitHub use also allowed us to support a learning approach based on collaboration (intra-team) and competition (extra-team). The lecturer and two tutors supervised the projects by fixing strict deadline and monitoring the project status on GitHub. A distinguished panel of corporate IT managers were asked to judge and give a prize to the three best apps produced during this course. The selection was based on the team live presentations conducted during a public event organized at the University of Salerno. Indeed, IT managers judged for each app its originality, the estimated business value, the pleasantness of the User Interface, the estimated technical quality, and the team presentation. To complete our study, we also performed a qualitative evaluation on the student's opinion concerning their learning experience and the used technologies.

The paper is structured as follows. In Section II, we discuss background. In particular, we provide the concept of Project-Based Learning and the technological solutions adopted for it, and successively we describe the main issues behind the mobile development in Android. In Section III, we detail our experience, while we describe the evaluation performed by the industrial partners and the collected student perceptions in Section IV. Final remarks and future work conclude the paper in Section VI.

## II. BACKGROUND

### A. Project-Based Learning

Project-based learning (PBL) is a model that organizes learning around projects [7]. It is based on both the constructive learning theory [8], where learners become active constructors of their knowledge, and on cooperative/collaborative learning [9], [10]. PBL enables students to cooperate in solving real problems, performing activities typical of the job world, which results in higher student involvement. The production of an artifact that is of interest since others can use or view it is a very motivating factor. Motivation can make the difference between success and failure of a learning experience more than any other factor.

<sup>1</sup>University of Salerno, Italy {francese, gravino, mrisi, tortora}@unisa.it

<sup>2</sup>University of Basilicata {giuseppe.scanniello}@unibas.it

<sup>1</sup><http://www.itcareerfinder.com/it-careers/mobile-application-developer.html>

<sup>2</sup>It is a knowledge transfer process based on communication through a shared mental or abstract model.

The instructor has a less central role, and students are responsible for their own learning (learner-centered education [11]), while learning is generated by their interaction (learner-learner interaction) [12]. When this methodology is supported by technology it is empowered by the capability of engaging learners by providing rapid, compelling interaction and feedback.

The adoption of PBL in Computer Science courses is growing [13], [14]. The main reason is that it enables to train students in principles, methods and procedures under conditions similar to developing real software products [15]. The development of a software product is the result of the team effort which requires both technical skills and soft skills, including the ability to communicate, to work as a team, to partition, assign and monitor task progress, and to assume responsibility for making choices. In the various project phases there is also the need of producing documentation which follows determined standards and not only to concentrate on the coding activity [14].

### B. Technology Support for PBL

Several technological solutions have been proposed to support PBL in computer science courses [14], [16], [17]. As an example, Macias [17] adopted a Moodle-based e-portfolio to support PBL activities involving a lot of deliverables and organizational resources.

Ardaiz-Villanueva *et al.* validate the effectiveness of Wikideas and Creativity Connector tools to stimulate the generation of ideas and originality by university students involved in PBL activities [18]. On the other hand, Zagalsky *et al.* [6] examined how GitHub is adopted as a collaborative platform for education. GitHub initially supported code and project management for software development; recently it is used also in other domains that involve collaborative work, including education, mainly for managing students and their work. It is exploited as submission platform, for hosting course material.

Kizaki *et al.* use GitHub as supporting tool for an Agile course consisting in scrum-based PBL [14]. The paper is mainly focused on the scrum methodology.

In this paper we conducted an in-depth case study of how GitHub plays a role in a specific course related to the learning of an emerging technology, Android-based apps development.

### C. Mobile application development in Android

One of the mobile app challenges is to deal with multiple platforms during mobile development [13]. Developers can create mobile apps by using either native development tools for each of the major mobile platforms, such as iOS, Android, Microsoft Windows Mobile, Symbian, BlackBerry, or cross-platform environments, including PhoneGap and Titanium [2], [4]. At the present, developers separately create the app for each platform. Indeed, the features of a specific operating system may not be available in another. Alternatively, developers can develop a cross-platform app that runs on any environment, but has more limited functionalities.

For example, to create an app that exploits in the better way the features of an Android device, developers have to master development skills related to the Android operating system and the associated development environment and resources. Android is an operating system whose demand is immensely expanding day by day.

As for the available resources, smartphones are equipped by sensors, such as accelerometer, gyroscope, GPS, brightness and temperature, offer communication features, including phone calls, SMS, email, and camera functionalities. The main Android components are: the activity and the service.

An activity is an app component that provides a screen with which users can interact in order to do something, such as dial the phone, take a photo, send an email, or view a map. Each activity is given a window in which to draw its user interface. For simplicity reason, we will refer to an activity as a GUI.

The app execution flow is continuously interrupted by the verification of asynchronous events. For this reason, the developer has to implement the activity logic taking into account its life-cycle. For example, when an activity is suspended (e.g., for the arrival of a phone call) the app has to perform specific work that is appropriate to that state change, or, when the device is rotated, an appropriate GUI has to be shown.

A service is a background component that performs either long-running operations or works for remote processes. A service does not provide a user interface. For example, a service might fetch data over the network without blocking the interaction the user has with an activity.

The market of Android hardware devices is very fragmented in terms of different screen sizes, processor types, custom APIs, etc. The main challenge is to maintain similar execution performances and user experiences in all these variations. Also testing is difficult, since it is practically impossible to test the app on all the available devices and OS versions. Android manages different configurations by exploiting non-code app resources (images, strings, layout files, etc.), which should include alternatives for each considered configuration.

Developers can create the GUI of an Android activity directly in Java or by using an XML-based layout file. The latter approach has two main advantages. It allows to 1) separate logic from presentation; 2) to maintain different parallel layouts for difference screen sizes.

The adoption of XML for specifying GUI requires, rather than setting the content view to be a view created in Java code, setting it to a reference to the XML layout.

## III. OUR TEACHING EXPERIENCE

*Goal.* The main objective of the Mobile Application Development course was to increase student interest, knowledge and practical experience in mobile development through an engaging and empowering PBL experience.

*The software platform.* The course was focused on the Android operating system because the barriers to entry in Android remain much lower. Indeed, with respect to iOS,



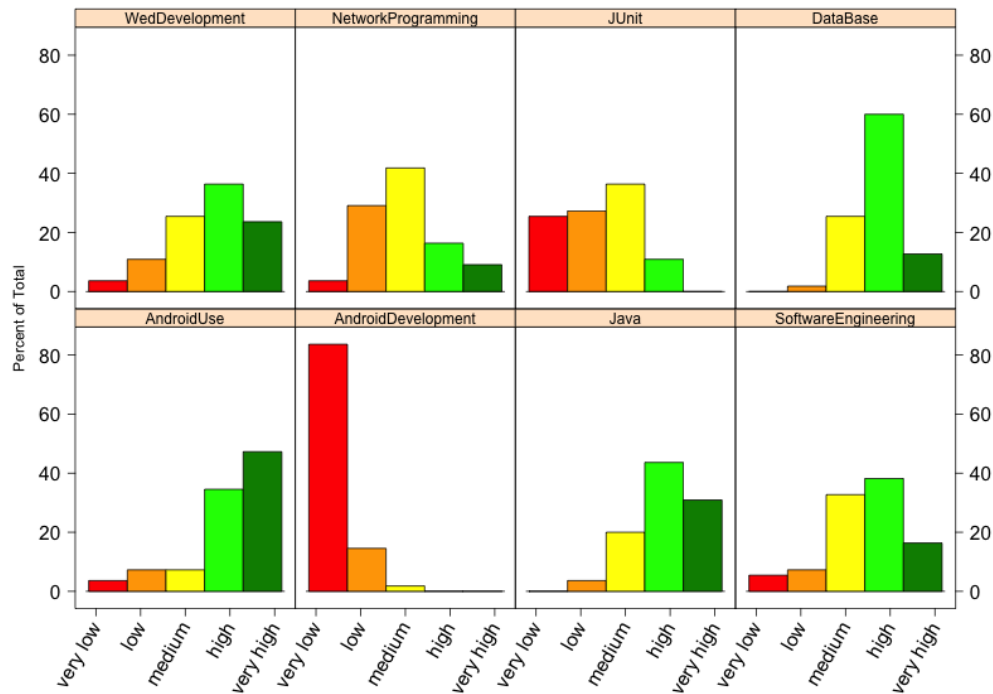


Fig. 1. Students' development competencies.

developers can iterate and test their designs quicker on Android, and marketing costs are significantly lower. At the present, Android users grow and are the largest overall smartphone market [1].

*Context.* The students were 55 Computer Science students at the University of Salerno. They were enrolled to the third year of the Bachelor program. Before the course, the students were asked to fill in a pre-questionnaire, which included questions (i.e., statements) regarding the following points: experience on the Android device use (as smart-users) and development, knowledge on Software Engineering and Web Development principles, experience on Java and Network Programming, and JUnit and Database knowledge. The questions admitted answers according to a 5-point Likert scale. Possible answers ranged from "Very Low" (1) to "Very High" (5). As shown in the histograms of Figure 1, students are generally smart-users, they did not know Android Programming, most of them thought to be good Java developer, 30 students affirmed to have a good competence in Software Engineering, 33 to be good web developer. Less (14 students) declared to have good network programming abilities. This was due to the contemporaneity of the network programming and the mobile development course. Very few known JUnit, while most of them were experienced in database development.

*Course organization.* The course consisted in two main modules: *i)* lectures on the Android operating system; *ii)* project work concerning the development of an app for smart devices accessing services available on the web. The course lasted 12 weeks. The course was performed in blended

modality. In particular, the first part of this course was in presence, with the didactic material available on-line on our learning platform. The course topics were the following: Android Activity lifecycle, modern interfaces, accessing to the web, threads, Android services, access to RESTful web service by JSON, accessing to native functionalities (i.e., GPS, sensors, camera, SMS, call), monetization. The second part of the course was conducted in distance modality and consisted in a project work, as better detailed in the following of this paper.

#### A. Project organization

*Teams.* Students were divided in 27 teams of two people, except one of three. The pedagogy of project-based learning suggest that to obtain good results groups should be composed of students with similar ability and interest in the topics being learned [19]. However, the debate on the effectiveness of homogeneous and heterogeneous groups is still open and needs further investigation [20], [21]. Thus, teams were composed according the students' preferences. We did not decide to randomly assign team members because the students had previous experience of project work in several courses and, at the last term, they know which is the classmate more appropriate to work with.

*Software Projects.* Each team directly proposed the app to be designed and developed in the project. The rationale behind this choice was related to stimulate student's entrepreneurship actions and creativity. Each proposal was accurately motivated, by performing a detailed market analysis. The projects had to respect the following nonfunctional

requirements: the app had to interact with a remote server, through JSON. It had to exploit native device functionalities, including maps, GPS, sensors, call, SMS. It had to handle device rotation and to use SQLite. Games were admitted if they exposed backend functionalities, such as account management, multiuser, bonus management, and app upgrade. The project started after the approval of the lecturer.

*Lifecycle model.* We addressed the students to follow an incremental prototyping development lifecycle. We did not discourage the adoption of pair programming, namely a software development practice where two programmers work as a pair on the same computer. Pair programming is an effective practice largely adopted in industrial settings [22]. As for students, it has been observed that their performances improve [23], namely they produce higher quality source code, are more confident in their work, and enjoy this more.

*Deadlines.* Students were informed of the project deadlines related to the presentation of the deliverables. The first deliverable was the project proposal, the second was the Requirement Analysis Document (RAD). Successive deliverables are referred to the GUI prototype, the mobile app prototype, the complete app, including the external server.

*Documentation.* The documentation required had to follow the templates proposed by the lecturer. In particular, a Project Proposal has to present the idea underlying the project, the motivation, a summary study of the market, also considering the apps available on the main app stores, and has to convince about the novelty of the proposed app. The RAD had to better detail requirements and also provide system models (actors, use case diagrams, class diagrams, sequence diagrams, navigational diagrams and user interface Mocks up). Black box test cases had also to be produced in order to test the final version of the app.

### B. On the Use of GitHub

Student projects were managed through GitHub, a distributed revision control and source code management (SCM) system [24]. It currently hosts over one million code repositories, and has 340,000 registered contributors. Each repository on GitHub has a dedicated project page that hosts the source code files, commit history, open issues, and other data associated with the project [25]. As investigated by Zagalsky *et al.* [6], GitHub can be a powerful learning management tool, differently used by various educators even in similar environments (e.g., technical background) and with similar requirements (e.g., class size, course type).

The lecturer and the tutors monitored the quality and the times of the projects, supervising that teams correctly performed their work to be able to participate to the public App Challenge. The lecturer creates a GitHub account for each team, downloading on it the documentation templates to be provided. In particular, he uses the GitHub mechanism for milestones, typical of many project systems. A new milestone simply has a title, description, and a date. GitHub also provides a graph view that summarizes project activities. In this way the lecturer had a high-level view of the students' activities during the app development.

The team uses the GitHub communication feature. In this way, the communication is handled in one centralized place rather than across emails and visible to all the team. Using labels team members (also the lecturer) can create issues for discussion. Team members can set up email notifications when people comment or tag them in an issue. The communication features offered by GitHub favor awareness, which has a very relevant value of activity information for small teams [26]. Indeed, notifying members of actions on shared artifacts helps them maintain mental models of others activities [27] and avoid potential coordination conflicts [28]. In particular, when a deliverable was completed, the team notified the lecturer that it was ready. The lecturer could accept it or notify the change to perform by adding a checklist-based revision to the teams' GitHub account. The communication between the lecturer and the team members was easier with respect to other learning management systems.

It is also important to point out that the transparency on GitHub supported learning from the actions of other students. Indeed, they are able to look at the documentation of students of the same team and of students of other teams, how the other students coded, what they paid attention to, and how they solved problems. The availability of this information enables them to learn better ways to code and access to superior knowledge [25]. Also competition is favored, since a team can monitor the state of the others and is stimulated to perform better.

## IV. EVALUATING TEAM WORK AND DEVELOPED APPS

The evaluation of the team work and the developed apps went through the following evaluation levels:

- *Lecturer.* The artifacts each team produced were constantly monitored by the lecturer. In particular, she used GitHub to monitor the progress of the projects and to assess whether teams respected deadlines for the delivery of software artifacts. GitHub was also used to enable the communication among the lecturer and the students. The communication among the lecturer and the students took place in presence when needed. For example, the students had to show three versions of their app and in this case revision meetings were planned and conducted in presence. A wrap-up meeting was also conducted before the App Challenge<sup>3</sup>, a public competition where students were asked to participate. The goal was to prepare students to the competition.
- *App Challenge.* The main goal of App Challenge was to stimulate students in engaging in the project, as well as to have excitement throughout the course. The participation to the App Challenge was on voluntary base, namely students participated only if interested. In our case, all the students participated in the App Challenge. During the competition, students gave a demonstration (the imposed time limit was eight-minute) to a panel formed by external IT managers of national and

<sup>3</sup><http://www.zerottonove.it/unisa-grande-successo-per-la-prima-edizione-di-app-challenge/>



TABLE I  
THE PERCEPTION QUESTIONNAIRE

ID	Question
P1	Managing my project with GitHub was easy.
P2	Using the Software Configuration Management features offered by GitHub (e.g., commit, check-out) was easy.
P3	Using the communication features offered by GitHub (e.g., notification, tagging) was easy.
P4	I think that the app I developed is complex.
P5	Basing on your experience, the development of mobile app is easy with respect to traditional desktop applications.
P6	Basing on your experience, the development of mobile app is easy with respect to traditional web applications.
P7	My experience in the development of mobile app during this course was involving.
P8	The final competition was a stimulus for improving the quality of my work with respect to a traditional exam.
P9	After this course, my Android development competences are: .....

international companies, whose business included the development of apps for smart devices. Each team of students had to show that their app meets the market needs, explain which technologies they selected and why, discuss their choice on the User Interface, and present a live demonstration of the developed app. The first three projects received a prize from the jury composed of 10 IT managers. The first prize consisted in two iPads, the remaining were external hard disks. We asked the students to fill a grid scored from 1 to 10 concerning the following aspects: originality, business value, User Interface, technical quality, presentation. The prizes were assigned considering the results of the IT managers' evaluation. One of the main goal related to the organization of our App Challenge was to assess the students' apps from a professional perspective.

- *Students.* We were also interested in collecting some feedback from students about their perception in: using GitHub, developing apps for smart devices equipped with an Android operating system, and participating in App Challenge. To this end, we asked the students to fill in the questionnaire reported in Table I.
- *Software and project metrics.* We collected both software and project metrics [29]. A metric is a quantitative measure of a degree to which a software system and/or process possesses some property. We collected software and project metrics for two main reasons: (i) to assess team productivity and work and (ii) to study the value of these metrics to estimate the effort needed to develop mobile apps. For space reason, we will focus here only on the first point. We collected the following metrics:

- *Requirements*, the number of functional require-

ments individuated in the RAD.

- *Checkouts*, the number of local working copy taken from the repository by the team members. It provides an indicative idea of how much the team members are active and how their work is distributed;
- *Time*, the time the students were active on the development phase of their project. It has been obtained by analyzing the activity log of GitHub;
- *User Interfaces*, the number of graphical components composing the user interface of an app. In particular, we considered the number of XML file describing the Android activity user interfaces;
- *LOC*, the number of lines of code, source code comment excluded.
- *Cyclomatic Complexity*, a measure of the control complexity of a program. It measures the amount of decision logic in a source code function. It is a measure of how is structured a program. A high Cyclomatic Complexity denotes a bad structure and high risk of errors.
- *Depth Inheritance Tree (DIT)*, which measures the software complexity of an inheritance hierarchy. It is the length of the longest path from a given class to the root class in the inheritance hierarchy. Some studies have shown that higher DIT rate corresponds with larger error density and lower quality [29]. The smaller the DIT, the more abstract and simpler the class would become, but decreases the class reusability. While the more a class inherits, the more difficult to understood the design is.

Our choice in selecting these metrics was mostly based on their simplicity in collecting and because they are well known and widely adopted (e.g., [29]).

Together with OO and traditional size code metrics, we also measured method calls in mobile apps. Method calls

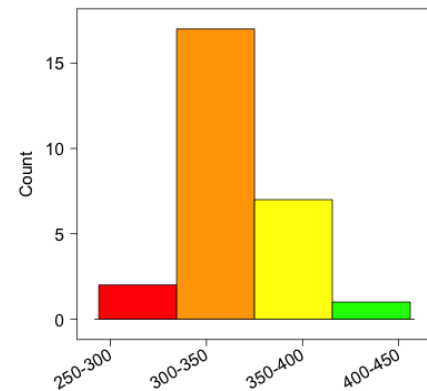


Fig. 2. The scores attributed by the jury.

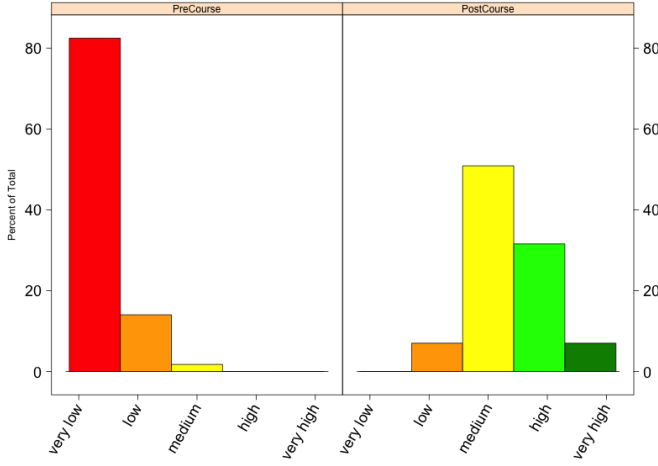


Fig. 3. The student perceptions on their Android development competencies Pre and Post course.

classified as both internal method and API (Application Programming Interface) calls. Internal method calls are invocations to methods the original developer implemented in the app, while API calls represent invocations to methods that Android provides. These metrics quantifies which use the app makes of native functionalities.

## V. RESULTS

All the teams completed their app and took part to the final competition. 11 apps were games, such as graphics retro-based games, or based on word guessing or math ability; 13 apps supported productivity (e.g. apps providing information on a City Hall, supporting people management or personal training), the remaining were social apps, e.g. for sharing their own travel diary or meeting people of interest.

### A. App Challenge

The scores of the jury are graphically summarized in Figure 2. Descriptive statistics are also reported in Table II. The possible scores could range from 1 to 500. Thus, a mean score 331 with only two apps that scored less than 300 revealed a good opinion of the IT managers on the students' apps. The app considered the best obtained 405 as the score. The app was a very captivating game. The developers were very able to present and motivate their app, also performing a particularly suggestive spot. The User Interface was very simple and fascinating. The technical complexity was lower because the game mainly worked on the mobile device, except for the server-side score management.

### B. Student perception

Concerning the opinions the students had on their Android development competencies, they perceived a notable improvement before and after the course, as shown in Figure 3. In particular, this figure depicts the histogram related the perceptions before and after the course, collected by the pre-course and the perception questionnaires (question

TABLE II  
JURY EVALUATION STATISTICS

Min	Max	Median	Mean	St. Dev.
282	405	328	331	27.17

P9), respectively. It is of practical interest to estimate the magnitude of performance difference perceived first and after the course. To this aim, we adopted the *Cohen d* effect size. The effect size is considered negligible  $d < 0.2$ , small for  $0.2 \leq d < 0.5$ , medium for  $0.5 \leq d < 0.8$ , and large for  $d \geq 0.8$ . In our case (paired analyses), it is defined as the difference between the means ( $M_{POST}$  and  $M_{PRE}$ ) divided by the standard deviation of the (paired) differences between samples  $\sigma_D$ .

$$d = \frac{M_{POST} - M_{PRE}}{\sigma_D}$$

Since the effect size is  $d = 2.28$ , we can consider that the students perceived that the course has had a considerable positive effect on their Android development competencies.

The answers to the perception questionnaire are graphically summarized in Figure 4. In particular, the greater part of the students asserted that GitHub eases the management of projects, 31 expressed a positive judgment (question P1). 36 students positively judged the CSM support offered by GitHub (P2), while 33 expressed a positive judgment on its communication feature (P3). 30 students judged complex the app they developed (P4). Most students (36) considered easier to develop mobile apps with respect to desktop ones (P5), while most of them considered easier develop web apps (P6). A high number of students (50) perceived the course involving (P7) and the final competition was very appreciate by 55 students (P8).

### C. Project metrics

The values of the considered metrics are summarized in Figure 5. In general, the produced apps do not have a large number of functional requirements: the projects were characterized by median 8. The number of checkouts is not elevated (54 on average), probably because often students worked in pair programming modality, on the same PC. The time took to develop the app on average was 59 working days. The time to accomplish the analysis phase should also be added (about one month). The User Interfaces produced for each app were 33 on average. The number of Line Code (LOC) was on average 3609. Cyclomatic Complexity was on average 1.83, which denotes a good modular structure of the code (low risk of errors for values less than 10). DIT was 4 on average. This means that the classes are not much reused, i.e., teams develop for each functional requirements.

Table III reports the descriptive statistics for API (Application Programming Interface), internal method calls and the total number of calls. Internal method calls are invocations to methods the original developer implemented in the app, while API calls represent invocations to methods that Android provides. Half of the apps made more than 726 API

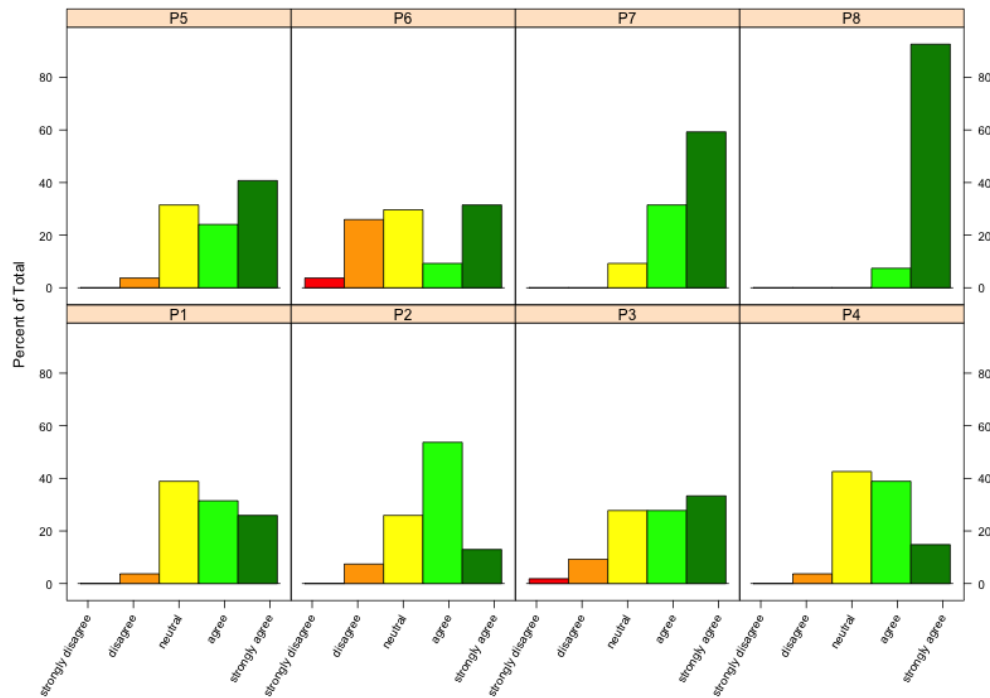


Fig. 4. The Perception Questionnaire results.

calls, that is they made a large use of the functionalities offered by the Android operating system.

## VI. CONCLUSION AND DISCUSSION

In this paper, we have presented a teaching experience gained in the context of an academic course at the University of Salerno for the development of applications for smart devices. In such a course, students arranged in teams implemented apps and cooperated using GitHub. A competition to establish the best developed app was also conducted and the jury in charge of judging the apps was composed by IT professional managers.

The findings gained from our teaching experience can be considered positive: all the students delivered the projects on time, with a good level of quality and completeness with respect to the established requirements. The possible motivations could be related to the following aspects: first, all the students were enthusiastic in developing apps for smart devices; secondly, their activity was monitored thanks to the use of a GitHub which enabled continuous monitoring of the team work in all the phases of the development process, starting from the project proposal. Last but not least, let the students present their work to IT managers belonging to top

IT companies. Indeed, by examining the project activity of the teams, when they knew of the company involvement their production notably increased. The lecturer and the tutors continuously motivated the students, also providing suggestions on the way they had to communicate. The App Challenge was successfully also because allowed the best students to be placed or to increase their familiarity with the work market. For example, a TLC company involved in this competition hosted the winning students for a stage because they demonstrated to be young talent with a strong ability to innovate. Many other students were required by the other companies involved in the App Challenge. Overall, all the companies manifested a positive judgment on the competition and on the work the students did. In fact, many of these companies asked to be informed and involved in future similar initiatives.

As future work, we plan to fully involve the IT professional managers in the next edition of the Mobile Application Development course. In particular, we would involve them as the role of coach. Future work will be also devoted to introduce in the next year course cloud platforms for implementing the back-end of the apps for smart devices.

## ACKNOWLEDGMENT

We thank all the students to this course and the IT managers, who judged the apps these students developed.

## REFERENCES

- [1] Gartner, *Top 10 Mobile Technologies and Capabilities for 2015 and 2016*. <https://www.gartner.com/doc/2665315>, 2014.

TABLE III  
DESCRIPTIVE STATISTICS ON THE API AND INTERNAL METHOD CALLS

Method call	Min	Max	Median	Mean	St. Dev.
API	36	11449	726	1398	2236
INTERNAL	9	1325	180	331	400

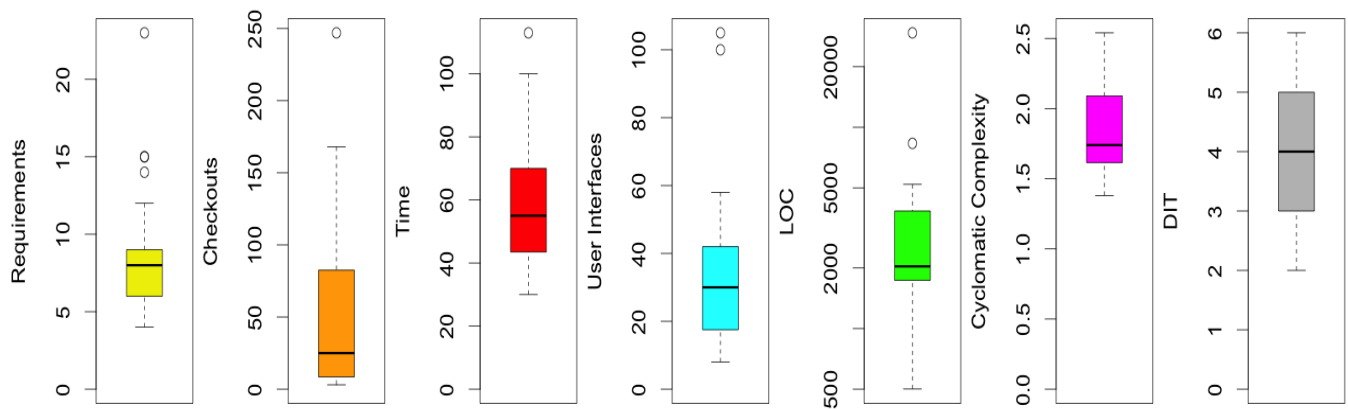


Fig. 5. The boxplots of the project metrics.

- [2] R. Francese, M. Risi, G. Tortora, and G. Scanniello, "Supporting the Development of Multi-Platform Mobile Applications," in *Proceedings of the 15th IEEE International Symposium on Web Systems Evolution (WSE)*, pp. 87–90, 2013.
- [3] R. Francese, M. Risi, G. Tortora, and M. Tucci, "Visual Mobile Computing for Mobile End-Users," *IEEE Transactions on Mobile Computing*, 2015.
- [4] M. Cimitile, M. Risi, and G. Tortora, "Automatic Generation of Multi Platform Web Map Mobile Applications," in *Proceedings of the 17th International Conference on Distributed Multimedia Systems (DMS)*, pp. 84–89, 2011.
- [5] R. Francese, M. Risi, and G. Tortora, "Management, Sharing and Reuse of Service-Based Mobile Applications," in *Proceedings of the 2nd ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, pp. 105–108, 2015.
- [6] A. Zagalsky, J. Feliciano, M.-A. Storey, Y. Zhao, and W. Wang, "The Emergence of GitHub As a Collaborative Platform for Education," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*, pp. 1906–1917, ACM, 2015.
- [7] P. C. Blumenfeld, E. Soloway, R. W. Marx, J. S. Krajcik, M. Guzdial, and A. Palincsar, "Motivating Project-Based Learning: Sustaining the Doing, Supporting the Learning," *Educational psychologist*, vol. 26, no. 3–4, pp. 369–398, 1991.
- [8] J. R. Savery and T. M. Duffy, "Problem Based Learning: An Instructional Model and Its Constructivist Framework," *Educational technology*, vol. 35, no. 5, pp. 31–38, 1995.
- [9] P. Dillenbourg, "What do You Mean by Collaborative Learning?," *Collaborative-learning: Cognitive and Computational Approaches*, pp. 1–19, 1999.
- [10] G. Scanniello and U. Erra, "Distributed Modeling of Use Case Diagrams with a Method Based on Think-Pair-Square: Results from Two Controlled Experiments," *Journal on Visual Languages and Computing*, vol. 25, no. 4, pp. 494–517, 2014.
- [11] D. A. Norman and J. C. Spohrer, "Learner-Centered Education," *Communications of the ACM*, vol. 39, no. 4, pp. 24–27, 1996.
- [12] J. Laffey, T. Tupper, D. Musser, and J. Wedman, "A Computer-mediated Support System for Project-Based Learning," *Educational Technology Research and Development*, vol. 46, no. 1, pp. 73–86, 1998.
- [13] M. E. Joorabchi, A. Mesbah, and P. Kruchten, "Real Challenges in Mobile App Development," in *Proceedings ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pp. 15–24, ACM Press, 2013.
- [14] S. Kizaki, Y. Tahara, and A. Ohsuga, "Software Development PBL Focusing on Communication Using Scrum," in *Proceedings of the IIAI 3rd International Conference on Advanced Applied Informatics (IIAIAI)*, pp. 662–669, 2014.
- [15] S. Yadav and J. Xiahou, "Integrated Project Based Learning in Software Engineering Education," in *Proceedings of the International Conference on Educational and Network Technology (ICENT)*, pp. 34–36, 2010.
- [16] Q. C. Lang, "Developing an Asynchronous Computer Mediated Communication Tool for Project-Based Learning," in *Proceedings of the 4th International Conference on Distance Learning and Education (ICDLE)*, pp. 222–225, 2010.
- [17] J. Macias, "Enhancing Project-Based Learning in Software Engineering Lab Teaching Through an E-Portfolio Approach," *IEEE Transactions on Education*, vol. 55, no. 4, pp. 502–507, 2012.
- [18] O. Ardaiz-Villanueva, X. Nicuesa-Chacón, O. Brene-Artazcoz, M. L. S. de Acedo Lizarraga, and M. T. S. de Acedo Baquedano, "Evaluation of Computer Tools for Idea Generation and Team Formation in Project-Based Learning," *Computers & Education*, vol. 56, no. 3, pp. 700–711, 2011.
- [19] Y.-T. Lin, Y.-M. Huang, and S.-C. Cheng, "An Automatic Group Composition System for Composing Collaborative Learning Groups Using Enhanced Particle Swarm Optimization," *Computer & Education*, vol. 55, no. 4, pp. 1483–1493, 2010.
- [20] C. A. Bowers, J. A. Pharmer, and E. Salas, "When Member Homogeneity is Needed in Work Teams: A Meta-Analysis," *Small Group Research*, vol. 31, no. 3, pp. 305–327, 2000.
- [21] S. Hooper and M. J. Hannafin, "The Effects of Group Composition on Achievement, Interaction, and Learning Efficiency during Computer-Based Cooperative Instruction," *Educational Technology Research and Development*, vol. 39, no. 3, pp. 27–40, 1991.
- [22] L. Williams and R. Kessler, *Pair Programming Illuminated*. Addison-Wesley Longman Publishing Co., Inc., 2002.
- [23] C. McDowell, L. Werner, H. E. Bullock, and J. Fernald, "Pair Programming Improves Student Retention, Confidence, and Program Quality," *Communications of the ACM*, vol. 49, no. 8, pp. 90–95, 2006.
- [24] GitHub. <https://github.com>.
- [25] L. Dabbish, C. Stuart, J. Tsay, and J. Herbsleb, "Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository," in *Proceedings of the ACM International Conference on Computer Supported Cooperative Work (CSCW)*, pp. 1277–1286, ACM, 2012.
- [26] P. Dourish and V. Bellotti, "Awareness and Coordination in Shared Workspaces," in *Proceedings of the ACM International Conference on Computer-supported Cooperative Work (CSCW)*, pp. 107–114, ACM, 1992.
- [27] T. Gross, C. Stry, and A. Totter, "User-Centered Awareness in Computer-Supported Cooperative Work-Systems: Structured Embedding of Findings from Social Sciences," *International Journal of Human-Computer Interaction*, vol. 18, no. 3, pp. 323–360, 2005.
- [28] A. Sarma, Z. Noroozi, and A. van der Hoek, "Palantir: Raising Awareness among Configuration Management Workspaces," in *Proceedings of the 25th International Conference on Software Engineering (ICSE)*, pp. 444–454, 2003.
- [29] V. R. Basili, L. C. Briand, and W. L. Melo, "A Validation of Object-Oriented Design Metrics As Quality Indicators," *IEEE Transactions on Software Engineering*, vol. 22, no. 10, pp. 751–761, 1996.

# Teaching Computer Programming in a Platform as a Service Environment

Mauro Coccoli

DIBRIS

University of Genoa

Genoa, Italy

mauro.coccoli@unige.it

Paolo Maresca

DIETI

Federico II University

Naples, Italy

paolo.maresca@unina.it

Lidia Stanganelli

Università Telematica

eCampus

Novedrate (CO), Italy

lidia.stanganelli@ecampus.it

Angela Guercio

Dept. of Computer Science

Kent State University

Kent OH, USA

aguercio@kent.edu

**Abstract**—In this paper we recall a previous research on the future development of the current model of higher education, which highlighted that the labor market is looking for people with competences and skills reflecting a T-shape model. As a consequence, universities should include a wider mix of disciplines in the curricula of their courses. Hence, to overcome existing criticisms and to provide some suggestions on how to enhance universities' performances, we thought of education as a process with inputs, outputs, and relevant dependencies. We based our research on a smart-city-like model due to the fact that next generation networks and relevant services are going to be more and more integrated with existing infrastructure and information management systems. Thus, it is mandatory that smart solutions are the most prominent assets of modern university environments, to improve the effectiveness of higher education. We called such a university a “smarter university” in which knowledge is a common heritage of teachers and students. In this paper, we report experimental results from a specific case study of collaboration between industry and university, which could be used as a reference for the definition of patterns to be applied in the redesign of the current education systems, even though it is referred to a technological application scenario.

**Keywords**—cloud computing; smart applications; collaborative systems; technology enhanced learning.

## I. INTRODUCTION

Owing to modern technologies, ever-growing computing power, miniaturization, innovation in network infrastructures and networking solutions, people and things are connected each other like never before. In this context, the most advanced solutions, such as the use of applications based on the Internet of Things (IoT) and the exploitation of the semantic web capabilities, can give a new impulse to e-learning [1] and push the interaction of people with both learning objects and learning environments, in a real semantic web of things [2]. Moreover, information and data are created and spread at high speed, in settings with less and less boundaries. After analyzing this general scenario in a previous work, we observed the actual teaching model at universities. We found some weaknesses and strengths, and we identified the main building blocks for the definition of a reference model of what we called a “Smarter University” [3]. In our study, we refer to smarter university, instead of smart university, due to the fact that, generally speaking, today's universities widely adopt cutting-edge technologies and systems, thus we can consider they already have the desired smartness characteristics. However,

that may be not enough and they should become “smarter” to improve their effectiveness, to enhance their performances, to be more flexible, and, last but not least, to be able to cope with novel and emerging needs of both society and labor market. In fact, according to the current situation, university teachers perform on-the-edge research activities, which make them the exclusive holders of knowledge, and they act consequently. If regarded from a theoretical point of view, it looks like a good model but, in the actual conditions, the fact of having very specific and very deep knowledge in a very narrow research sector does not match with the labor market needs that are focused on flexibility and require more and more interdisciplinary competences. At present, it is not so easy to find university courses that provide people with such skills. In fact, people with such skills should come from both technical faculties, such as, e.g., engineering, and social sciences ones; but at the same time they should hold knowledge and abilities in management, social behavior and human interaction, communication, teamwork, problem posing and solving, creativity, lateral thinking, and resilience.

This is the hard work that universities are challenged to do. And “How should they achieve it?” is the big question. Beside the disciplinary matters and the strategies leading to the choice of a suited mix of classes in specific courses, we are convinced that one of the most powerful enabling factors to find a solution is the tight collaboration between academia and industry, on common projects, with common objectives, to drive students to learn how to apply theoretical concepts for the solution of some real world problems.

In such a vision, universities, organizations and companies should cooperate to develop together an ecosystem in which they could learn from each other. In this way, universities will be able to achieve the new “smarter” level and be ready to teach novel design and methodologies and new reasoning paradigms, while industry and organizations could find new market shares to conquer. Finally people could find new jobs.

According to this vision, and making reference to the pillars of smarter universities listed in the previous work, we focus our attention on a specific technological issue. In particular, we consider a novel research trend that is gaining consensus in the scientific community and that we expect to have a prominent role in the very next future, that is, the exploitation of the Platform-as-a-Service (PaaS) paradigm in software production [4]. This will allow using remote virtual machines in place of



local hardware and software, thus avoiding time-consuming and expensive installation procedures as well as annoying maintenance tasks. More in details, in this paper, we are going to showcase the outcome of an educational activity that we carried on in a distributed environment based on cloud computing and services. Specifically, we make reference to the most recent stage of a long-term project aimed to empowering collaboration skills in software engineering students, the ETC (Enforcing Team Collaboration) project [5]. In this project, students are grouped together to create small teams regardless of they might be from different universities as well as from different countries. Then, within the tasks they are assigned to, they must cooperate to achieve common objectives, which include, among others, the development of working prototypes of some web-based applications. All the students that enrolled to the laboratory could rely on a bunch of professional tools specifically crafted to support software development and for the management of the software lifecycle, which are made freely available at their universities through the IBM academic initiative. In this framework, in the past, we started experimenting throughout the Jazz ecosystem in conjunction with the renowned open source Eclipse IDE (Integrated Development Environment) and, in the last year, we started using also the IBM Bluemix platform and its relevant facilities.

The remainder of the paper is the following. Section II explores related works then, in Section III we present the ETC Project and the framework in use. Section IV describes the Eclipse framework as a learning environment and Section V outlines the ETC-BLUE project. A summary of the results is reported in Section VI and, finally, a glance on future work concludes the paper.

## II. RELATED WORKS

The Computer Supported Collaborative Learning (CSCL) strategy implemented through the PaaS paradigm can be found in the literature in other related works. According to Silverman [6], *“the adoption of a collaborative learning strategy can be useful in many situations (and it can be realized with or without the use of technology).”* Moreover, Dong et al. [7] assert *“the current models of e-learning ecosystems lack the support of underlying infrastructures, which dynamically allocate the required computation and storage capacities for an e-learning ecosystem. Cloud computing is a promising infrastructure which provides computation and storage resources as services.”* In addition, in [8] the authors conclude that *“e-learning systems can use benefits from cloud computing using: Infrastructure (i.e. use an e-learning solution on the provider’s infrastructure), Platform (i.e. use and develop an e-learning solution based on the provider’s development interface), Service (i.e. use the e-learning solution given by the provider)”*. Given these general considerations, we observe that, the use of the IBM Bluemix platform (the same one we used in this work) is also reported in [9] where the authors describe how their students worked on a database in the cloud, in a virtual laboratory environment and in [10] where the author says: *“by hosting the entire development environment, PaaS increases productivity, lets organizations release products faster, and reduces software’s cost.”* Considering these results, this work proposes a smart education model which creates a CSCL and exploits the tools available in the

IBM Bluemix platform to create applications in Java, which run over the Android operating system, and to achieve the following advantages: shorter learning time, better quality of the prototypes/products, and implementation of the T-shaped model [11]. In [12] the authors say that *“seamless and pervasive intelligence is already proving disruptive in education, with traditional campus-based education models changing as new teaching methods evolve, augmented by automated and interactive learning outside the classroom and distance participation”* and also *“we also project that courses will involve less instruction and lecturing and more dialogue with expert professors, resulting from the ability to use technology for interaction outside the classroom”*. The perspective is that education will be seamless and ubiquitous for those who can afford information technologies. To conclude this quick review, we mention [13] where the author says that *“the potential of cloud computing for improving efficiency, cost and convenience for the educational sector is being recognized by a number US educational (and official) establishments.”* and *“there is also an increasing number of educational establishments that are adopting cloud computing for economic reasons.”*

## III. ENFORCING TEAM COLLABORATION

The Enforcing Team Collaboration (ETC) project was created with the need of developing cooperation skills between university students when they are required to work in groups. Teamwork ability is an essential skill for students to acquire; learning and practicing this skill can give a glance to their future team-working experience. In particular, we are talking about students involved in software engineering activities and, consequently, the ETC project creates an effective CSCL system for higher education that targets the area of software engineering, computer programming, and team cooperation for software analysis and software development [14]. In the rest of the paper we will refer to this particular case study. However, the same principles applied by ETC to the above areas, can be applied, with different tools, to other areas beyond computer programming.

We can face the problem of developing such skills from different points of view. On one hand, from the educational perspective, we observe that people must be duly trained to acquire competences in software engineering models and techniques, as well as in project management and human relations. On the other hand, from the technological perspective, we observe that a complex system is needed to enable and support collaboration as well as to ease interactions between the participants. Finally, we observe that distributed architectures and cloud computing can foster new behavioural paradigms in acquiring and disseminating knowledge and sharing experiences, thus they are needed in the learning process as well [15]. In support to this we consider the vision on how recent advancements in grid and cloud computing and mobile communications have significantly changed many concepts at the basis of e-learning as presented in [16]. In particular, we can envisage new learning models, which ease the implementation of hands-on activities and can fully exploit users’ interaction, due to the absence of located machineries, physical devices and structures as well as working environments such as computer rooms with limited number of



seats and time constraints access. We want to demonstrate that this can improve learning outcomes and accelerate the education process, while making more flexible the design of courses, lectures and practical activities to be assigned to students for the assessment and evaluation of their competences. Moreover, sharing resources and collaboratively constructing reusable learning assets, can significantly reduce costs in terms of both time and money [17].

The project was born by noticing that, in many cases, the software production process can become hard due to lack of a full integration among the tools and meta-tools of different teams such as database, interface, processes, knowledge sharing, and so on. This has nothing to do with being able to design and write good code and can cause significant loss of time and demotivation during the learning process. Hence, the proposed solution is teaching both teamwork and computer programming in parallel. Based on novel programming paradigms and tools specifically created for supporting teamwork, a suited software platform, enabling effective team working, was setup in order to coordinate the cooperation in developing code among students that study in different universities, have different working time-frames, and may be from different countries.

Many different systems and tools are available for the coordination of the software development process activities. At the same time recent integrated environments are shifting the focus to remote cooperation, which is considered the best way to cut down time and money. For such environments to be effective, we need something like an “orchestra director” over the development process. Generally speaking, the orchestra director is the one who knows exactly when each instrument must be played, and how to leverage the quality of the overall execution. Specifically, we found all of these features within the Jazz development platform. This complex platform, released by IBM, is usually adopted worldwide by IBM researchers for the development of software in cooperation. Before the ETC project was launched, such a kind of complex platform had never been used in an academia setting for a geographically distributed project. Consequently, in the ETC project, together with the Jazz environment, we adopted the renowned open source Eclipse IDE as a development platform. It is worth noticing that, in addition to writing clean and working lines of code in the preferred programming language, students have to cope with other tasks such as debugging, compiling, and, finally, the deployment of activities on specific hardware platforms and operating systems. This requires the inclusion of resources necessary to run experimental distributed software architectures.

Based on such considerations, we can summarize that the ETC project consists of experimenting with the realisation of collaborative activities based on the Eclipse community and tools in which different teams have to complete a group of tasks that have to be integrated with systems or subsystems developed by other teams. To reach the project main objective the IBM Rational software tools were integrated into software engineering academic projects. The project was sponsored By IBM Italia and the University Federico II of Naples received an IBM Faculty award in 2011 for the project. A variety of Italian universities participated in the project, such as: the University

of Napoli Federico II, the University of Milano Bicocca, the University of Bologna Alma Mater, the University of Bergamo, the University of Genova and its regional campus in Savona, the University of Bari and its regional campus in Taranto. Each university formed teams of students from different courses. Specifically, the course of software engineering from the University of Napoli Federico II, the University of Bologna and the University of Milano Bicocca; the course of web design for the University of Genova and the University of Bari; the course of advanced programming and testing from the University of Bergamo. Heterogeneous teams were composed of students from different universities with one teacher as a tutor for each group. Moreover, for each University, a *champion student* (usually a computer engineering or computer science Ph.D. student) is put in charge of corresponding with a teacher and acts as a responsible for each local group and support. One computer engineering Ph.D. student is put in charge of the technical direction of the entire ETC platform (both software and hardware).

Based on the encouraging results deriving from the ETC experience, we aimed at building wider team cooperation projects from lessons learned in open communities of practice [21] and we have extended the original project by designing new activities for groups of students that included the Kent State University, thus creating a more complex and broad working environment. The project was called ETC-plus [22].

#### IV. THE ECLIPSE FRAMEWORK AS A LEARNING ENVIRONMENT

In this Section we discuss how the Eclipse IDE can be considered as the inner center of a learning framework where students of computers programming write down their code and can easily interface with a number of external tools and services for a wide variety of different purposes. In fact, Eclipse is an open universal platform for tool integration, is an open and extensible IDE, and an open source community as well. The aim of Eclipse creators, and hence of the Eclipse-based tools, is to give to developers the freedom of choice in a multi-language, multi-platform, multi-vendor environment supported by multiple vendors. In addition, Eclipse provides a unique environment for members of the academic community to build new tools for teaching, doing research, and fostering further growth of the Eclipse community [18]. We point out that integration is part of the software development process and it occurs through tools inside and outside the IDE. In order to maximize the collaboration results with the minimal effort, we have joined the Jazz project, which seeks to integrate collaborative capabilities into the Eclipse IDE, thus enabling small teams of software developers to work together in a more productive way [19]. In brief, team cooperation in the context of ETC is enabled by the Jazz platform via the following tools:

- (i) Rational Team Concert (RTC);
- (ii) Rational Quality Manager (RQM);
- (iii) Rational Requirement Composer (RRC).

These three tools assist teams in developing in cooperation software specifications while maintaining quality constraints. An overview of the results obtained with the use of Eclipse on the Jazz platform is presented in [20].

After having successfully used and developed systems on the Eclipse-and-Jazz integrated platform in the ETC-plus project [21, 22], since last October 2014, we have joined the IBM Bluemix program. The use of the IBM Bluemix platform has made easier to cope with issues related with the management of data, infrastructures, connectivity, and servers. In fact, its use allows a paradigm shift so that now we can exploit advanced solutions according to the Software as a Service (SaaS) model, on a Platform as a Service (PaaS) environment. Consequently, we do not have to worry about managing servers, databases, virtual machines, and multiple releases of instances. Furthermore, the extensive use of the cloud relieves from data management issues, including security of both network and software.

## V. THE ETC-BLUE PROJECT

The lessons learned from past experiences have driven us to the definition of a new scenario, which takes into account that requirements change very quickly due the fact that activities are bounded learning tasks. Also in this case we found a solution on the IBM shelf in the recently released IBM Bluemix platform, which allows developers to use a combination of the most prominent open source computer technologies to power their apps, by handling in a seamless way the integration with apps and systems running elsewhere and managing data in the cloud [23]. We believe that the adoption of the IBM Bluemix platform will foster further developments and increase the educational results achievable through collaborative work activities, resulting in students that learn faster and acquire competences and skills in different fields.

To prove the validity and the effectiveness of the presented concepts, we have created the most recent release of the ETC project, called the ETC-Blue project. The idea behind the experiment is that of “grafting” a university course in a formalized company internal training process with the aim of getting T-shaped students. In this experiment we involved a pool of university students of a software engineering degree course with the main objective of:

- (i) strengthening the vertical part of the T, which is made of a deep and narrow knowledge of computer programming and operating systems;

- (ii) completing the horizontal layer of the T by developing skills in project management, collaboration, and leadership.

The project participants are the University of Naples Federico II and IBM Italia. The participating students are the ones enrolled in the “Computer programming - I” course of the Software Engineering degree at the Univ. of Naples. IBM has supplied the students with a crash course on the Bluemix platform to make them aware of the features of the platform and to speed up the learning curve. In this way students could quickly focus on the design phase and start implementing sophisticated functionalities for their apps without having to worry about databases, server connections, security issues, etc., which are ready-to-use available services of the platform and, thus, transparent to the developer. Moreover, the IBM Bluemix platform is accessed through a web browser and no software has to be installed on local machines, allowing students to bring

their own computers without having to rely on the university facilities. In addition, they can continue studying and experimenting at home or anywhere else, at any time, e.g., at nighttime. Another positive side effect is that universities are not any more requested to maintain a huge number of machines with specific configurations in students’ laboratories, according to the model of virtual laboratories, saving money and resources that could be used more fruitfully. This should increase the students’ satisfaction, which can get better services.

In more details, the crash course took two days, for a total of eight hours of lectures, including hands-on lab and it was given two times in November 2014. The students who attended the course were 120 but a total amount of 150 people was involved, including, among these, professors from local and nearby universities (i.e., the University of Cassino), Ph.D. students and professionals as well coming from different cities, such as, i.e., from Milan.

As a follow up of the crash course, many students have started individual projects in small groups immediately after. It is worth noticing that one of the above-cited projects involves the Kent State University at Stark, USA, and this should be regarded as an example of really distributed team collaboration. In fact, working together, in this case, implies significant geographic distances, different time, and different languages spoken as well. The collaborative environment, which was already intensively tested in the previous years with the ETC-plus project, has proved once more to be very effective and the first results achieved within this cooperation are really encouraging.

## VI. ACTIVITIES AND RELEVANT RESULTS

After the above-cited crash course, 92 students were arranged in 26 groups, each of these made by 4/5 people. Then, for each group we selected 2 students and assigned them the roles of team leader and deputy, so that they assumed the responsibility for the management of the whole project and for the external communications (i.e., with the teacher) too. In the following we list some of the running projects to give a flavor of the type of activities carried on, also providing short descriptions.

- 1) **Knowledge Hound.** This work aims to facilitate the community building around specific activities carried on in the university for both education and research. In fact, if some students may need support to the solution of specific problems, it is possible that some other students are working on the same problem. For example, there are students who, while working at a master thesis, have acquired a deep knowledge about that issue. Through the Knowledge Hound, knowledge can easily circulate and students teach themselves exchanging and merging own individual competences. In this respect, the project has the aim of developing a proximity-based app in which every student can state personal abilities and skills and search for the missing ones in other students’ profiles. The development of this project has started over the IBM Bluemix platform and the expected outcome is an Android app running on different devices such as smartphones, tablets, laptops and desktop computers. In Figure 1 we show some screenshots.

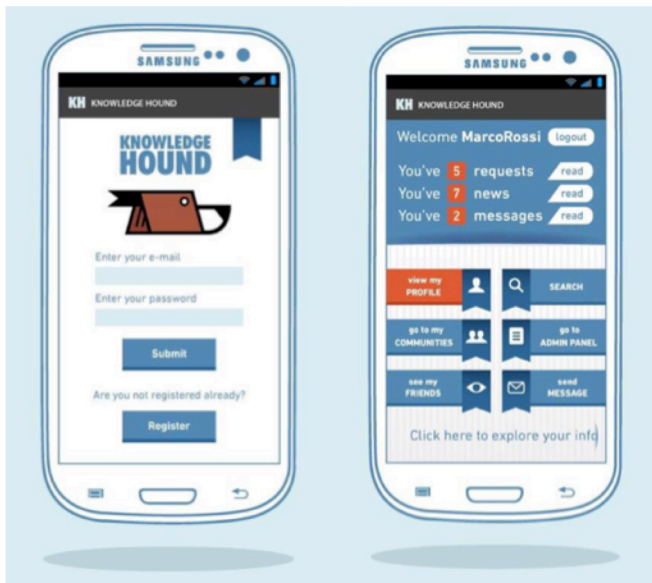


Figure 1: The login/registration page and the home page of the Knowledge Hound app

2) **K12**. This is a project launched in collaboration with the Kent State University at Stark, which fully exploits the features offered by the IBM Bluemix platform. The objective of this activity is defining innovative educational materials supporting both teachers and students in their relevant learning activities. According to its name, the project addresses K12 students by providing students and teachers with open source and reusable learning resources. Learning activities include, among the others, quizzes at different difficulty levels, which can be customized to individual students' profiles. In Figure 2 we show some screenshots from the student app, composed by a main screen, a quiz screen and a chat.

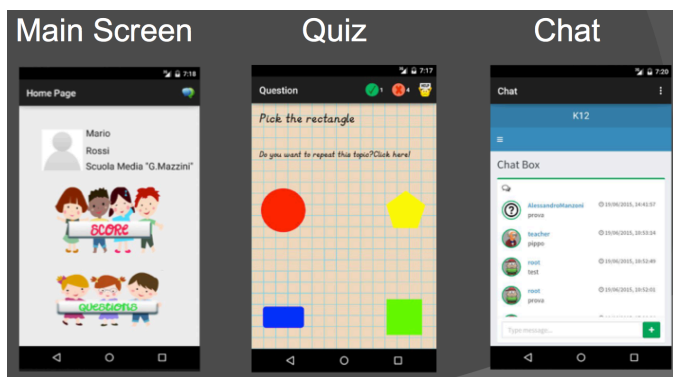


Figure 2: The K12 app for math. Particular attention was given to the graphical interface, to be attractive for kids.

3) **SmartApp**. This project has the objective of developing an app really useful, not only a mere academic exercise, which can be delivered through the application stores to a wide public. The project is in charge to 2 different groups of 5 students each. The target public should be composed of tourists or, generally speaking, traveling people. The main functionality of the app is collecting apps based on location criteria. "Where are you? And, hence, what specific apps could you need now?"

Possible suggested apps could be, e.g., local transportation routes and timetables, local museums, local weather forecast, and others. In addition the app can ask, "Who are you?" and provide the user customized replies listing, e.g., only apps related to music rather than sport events, find contents in your language and suited to your age or other criteria.

4) **ElectionUp**. This project involves 5 groups of 5 students each, and is the design of an app made to follow in real-time ballots in elections. This involves real-time communication with a shared database and the need of suited tools and algorithms for data analysis and visualization. Moreover, data should be accessible through common interfaces and APIs to other systems and a friendly user interface is needed. Figure 3 and Figure 4 show the user interface and the data visualization screen.

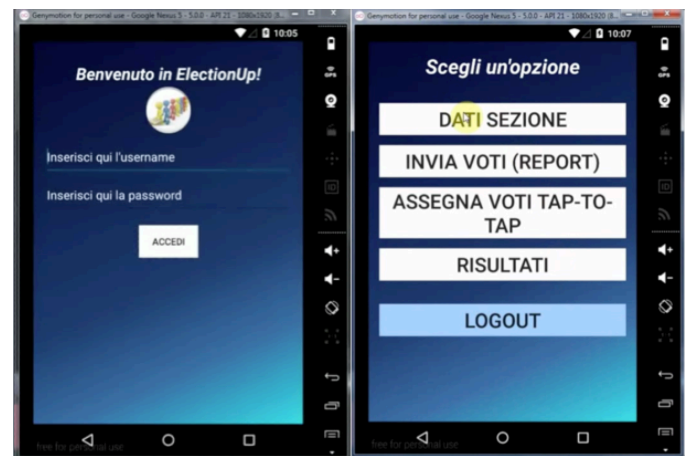


Figure 3: The ElectionUp app for Italian system election. Particular attention was given to the graphical interface, to be user friendly.

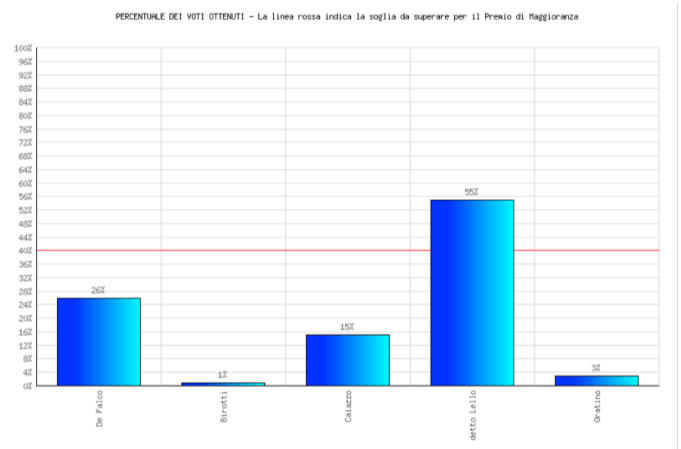


Figure 4: The ElectionUp app for Italian system election: the percentage of votes obtained.

Specifically, 12 groups are on the "Knowledge Hound", 2 groups are on the "K12 math" (plus a team of 4 from the American side), 2 groups are on the "SmartApp" and 5 groups are on the "ElectionUp". Summarizing, a total amount of 92 people have been working on the IBM Bluemix platform and duly finished their assigned tasks. After the first stage, those students who have finished the activity they were assigned to,

assume the role of project managers for the newly entering teams of students and charge other fifty people with new tasks to improve their previous works, fixing bugs, finishing uncompleted tasks and implementing new features, for the refinement of existing prototypes, based on the results of previous evaluation and assessment. It is worthwhile noticing that evaluation of software performances and usability happens between peers, while assessment is in charge to the teacher.

Based on this philosophy, projects outcome can be incremental and every team can start the assigned work, inheriting parts already developed by previous ones, adding or improving functionalities of an app that will become more and more complete, easy to use, and powerful. At the actual stage, two different teams have laid the foundations for a fan of projects. The former developed some basic building blocks and a common knowledge base, which constitute the substratum for the forthcoming groups to operate on. The latter, developed a variety of interfaces for the Android operating system, exploiting the Eclipse ADT (Android Development Tools) and the foundations provided by the IBM Bluemix boilerplates.

From the educational perspective, we highlight that, beside the development of the above-cited components, students involved in this first stage developed specific training materials to enable other future students to use the common workspace. In addition they also setup suited tools for the management and coordination of groups. But this educational activity has interesting points even if regarded from the software engineering perspective. In fact, the groups involved used the IBM Bluemix life cycle management illustrated during the initial crash course, passing through Jazz and DevOps.

As an example, in Figure 5 we show a screenshot taken from the IBM Bluemix dashboard. The picture illustrates the modules used within the above-cited Knowledge Hound project. Each tile gives access to the configuration environment for the relevant service. They include services for:

- (i) Mobile Application Security,
- (ii) Mobile Quality Assurance,
- (iii) Mobile Data,
- (iv) Push, and
- (iv) an SQL database server.

Of course, all of them are provided, as a service, by the Bluemix platform itself and this is a great advantage from the point of view of maintenance but also for the achievement of software engineering skills and design capabilities, since students are free to think of their computer programs in terms of architecture and high level interfaces, regardless specific implementation issues and the software available at their universities. Moreover, we highlight that the adoption of cloud-based solutions for software services and storage could solve a huge number of problems to computer laboratories of universities that should not install and maintain numerous software packages for many different purposes. One sad note about this from the authors is that in Italy having large bandwidth connections is still a serious problem in many geographical areas and this could slow down the deployment of cloud services.

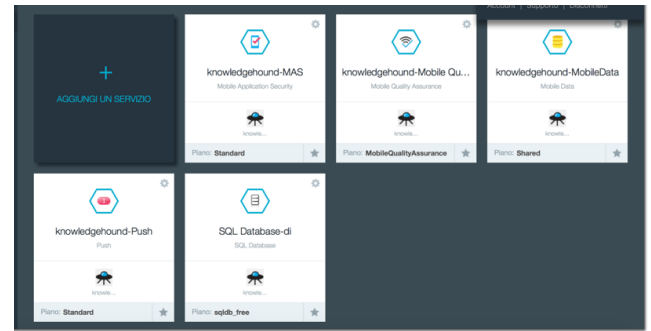


Figure 5: The services used in the Knowledge Hound app. The view from within the IBM Bluemix dashboard

## VII. CONCLUSIONS AND FUTURE WORK

In conclusion, the more important lesson learned is that the joint effort from university and industry together can give outstanding results for a wide range of reasons. Summarizing, the main motivations are: students learn to cooperate in small teams; the collaboration implies a split of tasks and drives everyone to make the best out of his effort and this implies that individuals' competences emerge; among these, leadership is one of the abilities which clearly appears and, consequently, portending e-leaders can be identified. In this newly developed framework the collaboration among university and industry has been give-and-take. University took the training course from industry, yet contributing to the definition of their contents, with the aim of readily exploiting them in specific projects for the dissemination at a students' level.

These preliminary results demonstrate that the use of the IBM Bluemix platform, tacked on the complex eco system based on the Eclipse IDE seamlessly integrated with the Jazz products and solutions that was developed in the past years, can greatly improve performances of the students, which gain core competences faster and in a realistic working environments.

Summarizing, we can say that ETC-Blue has reported several advantages and we observe that the most important part of the architecture is collaboration (smart ETC-Blue). In fact, collaboration has made possible the design and development of resources useful to the all of the teams. Moreover, ETC-Blue drives standards and forces open innovation networks, requires mature organizations and produces high quality products. It is noted that if there are mature organizations to act as a driver of an experience like that, one can get high quality products (i.e., software) but also students trained in an excellent manner as also reflected in the results achieved.

ETC-Blue fosters learning methods that are student-led versus instructor-led, with professors playing a mentor role in the learning process. This is a student-centric paradigm, which constitutes the basis for collaboration between people in teams and among groups. Within the large number of students involved in this experiment, we observed that some groups were *crawling*, other groups were *walking*, others were *running* and some were even *flying*. The teacher, acting as a coach, should identify those groups that fly and motivate them so that they can achieve quickly the best results and so that then they can spread to other groups, helping them to reach a superior level.

In a broader vision, we highlight that ETC-Blue can nurture the creation of smarter campuses, which are interconnected, enriched and fed by on-the-ground knowledge being developed over social networks. ETC-Blue favors the creation of smarter universities and forces teachers to have the most updated and relevant curricula, which then attract the best students who then will have the best formations, creating a virtuous circle of collaboration between universities and companies. ETC-Blue implements team-based projects across geographical, disciplinary and institutional boundaries and sustains a community that enables the formation of “T-shaped” people. Finally ETC-Blue fosters leadership and e-leadership.

Future work will be dedicated to finalize projects that are still open and, besides, we want to reserve a specific space for a students’ session within the forthcoming annual workshop of the Italian Eclipse community (*Eclipse-IT 2015* hosted in Rome, Italy, on October 14<sup>th</sup>, 2015), where they will have the opportunity to show what they did to a wide audience including academics and professionals. Moreover, we want to carry on new experiments involving more companies, and even startups, to prove that innovative working environments that push collaboration can enhance their productivity and that they can profit from the university think tanks through students’ internships (or by other means of collaboration) even before they are graduate and, thus, participating in this way to the education process, which can bring a great added value to consolidated realities. It is worthwhile noticing that the same experience can be replicated in other settings, regardless the chosen PaaS platform.

Another target is to launch an IBM Bluemix ecosystem tracing previous experiences made with IBM Jazz in the various releases of the ETC projects depicted in this paper. In fact, merging these working environments has created a very powerful educational experience that students lived with enthusiasm attaining encouraging results. Moreover, at the end of their activity, we observed a high degree of satisfaction and a growth in personal appreciation as well. Despite this, some students complained about a steep learning curve. However, we have kept into account their feedback and we have taken specific actions to overcome this criticism. To this aim, we put a significant effort on the development of instructional materials, which have been duly realized in the form of video tutorials, user guides, handbooks on essentials, mind maps and more. Such learning assets will be readily available to future people involved in similar activities and they will be the starting point in their learning experience.

#### ACKNOWLEDGMENTS

The authors wish to thank IBM and the Italian Eclipse Community for their friendship and their kind collaboration.

#### REFERENCES

- [1] G. Adorni, M. Coccoli, I. Torre, “Semantic web and internet of things supporting enhanced learning,” *Journal of E-Learning and Knowledge Society*, vol.8, no.2, 2012, pp. 23-32.
- [2] M. Coccoli, I. Torre, “Interacting with annotated objects in a semantic web of things application,” *Journal of Visual Languages and Computing*, vol.25, no.6, 2014, pp. 1012-1020.
- [3] M. Coccoli, A. Guercio, P. Maresca, L. Stanganelli, “Smarter universities. A vision for the fast changing digital era,” *Journal of Visual Languages and Computing* vol.25, no.6, 2014, pp. 1003-1011.
- [4] G. Lawton, “Developing Software Online With Platform-as-a-Service Technology,” *Computer*, vol.41, no.6, 2008, pp.13-15.
- [5] M. Coccoli, P. Maresca, L. Stanganelli, “Enforcing team cooperation: an example of computer supported collaborative learning in software engineering,” *Proceedings of the 16<sup>th</sup> International Conference on Distributed Multimedia Systems, Workshop on Distance Education Technologies*, pp. 189-192, 2010.
- [6] B.G. Silverman, “Computer Supported Collaborative Learning,” *Computers Education*, vol.25, no.3, 1995, pp. 81-91.
- [7] B. Dong, Q. Zheng, J. Yang, H. Li, M. Qiao, “An e-learning ecosystem based on cloud computing infrastructure,” *Proceedings of the 9<sup>th</sup> IEEE International Conference on Advanced Learning Technologies*, 2009.
- [8] N. Radhakrishnan, N. Poorna Chelvan, D. Ramkumar, “Utilization of cloud computing in e-learning systems,” *Proceedings of the 2012 IEEE International Conference on Cloud Computing, Technologies, Applications & Management*, 2012.
- [9] E.P. Holden, J.W. Kang, G.R. Anderson, D.P. Bills, “Databases in the cloud: a status report,” *Proceedings of the ACM International Conference on Information Technology Education*, pp. 171-176, 2011.
- [10] G. Lawton, “Developing software online with Platform-as-a-Service technology,” *Published by the IEEE Computer Society*, June 2008.
- [11] T in <http://tsummit2014.org/t>, last accessed July, 2015.
- [12] H. Alkhatib, P. Faraboschi, E. Frachtenberg, H. Kasahara, D. Lange, P. Laplante, A. Merchant, D. Milojicic, K.Schwan, “What will 2022 look like? The IEEE CS 2022 report,” *IEEE Computer*, March, 2015.
- [13] N. Sultan, “Cloud computing for education: a new dawn?” *Int. J. of Information Management*, vol.30, no.2, 2010, pp.109-116.
- [14] M. Coccoli, P. Maresca, L. Stanganelli, “Computer supported collaborative learning in software engineering,” *Proceedings of Global Engineering Education Conference*, pp. 990-995, 2011.
- [15] L. Caviglione, 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.
- [16] L. Caviglione, M. Coccoli, V. Gianuzzi, “Opportunities, integration and issues of applying new technologies over e-learning platforms,” *Proceedings of International Conference on Next Generation Networks and Services*, pp. 12-17, 2011.
- [17] L. Caviglione, M. Coccoli, E. Punta, “Education and training in grid-enabled laboratories and complex systems,” in *Remote Instrumentation for eScience and Related Aspects*, pp. 145-157, Springer, 2012.
- [18] IBM, “Eclipse,” <http://www.research.ibm.com/eclipse>, 2010, last accessed July, 2015.
- [19] L. Cheng, S. Hupfer, S. Ross, and J. Patterson, “Jazzing up Eclipse with collaborative tools,” in *Proceedings of the 2003 OOPSLA Workshop on Eclipse Technology Exchange*, pp. 45-49, 2003.
- [20] R. Frost, “Jazz and the Eclipse way of collaboration,” *IEEE Software*, vol.24, no.6, 2007, pp. 114-117.
- [21] P. Maresca, A. Guercio, L. Stanganelli, “Building wider team cooperation projects from lessons learned in open communities of practice,” *Proceedings of the 18<sup>th</sup> International Conference on Distributed Multimedia Systems, Workshop on Distance Education Technologies*, pp. 144-149, 2012.
- [22] A. Guercio, P. Maresca, L. Stanganelli, “Modeling multiple common learning goals in an ETC-plus educational project,” *Proceedings of the 19<sup>th</sup> International Conference on Distributed Multimedia Systems, Workshop on Distance Education Technologies*, pp. 122-128, 2013.
- [23] K. Kobylinski, J. Bennett, N. Seto, G. Lo, F.Tucci. “Enterprise application development in the cloud with IBM Bluemix,” *Proceedings of the 24<sup>th</sup> International Conference on Computer Science and Software Engineering*, pp. 276-279, 2014.

## Authors' Index

Amato, Flora	62, 191
Anna Maria, Fanelli	23
Ardito, Carmelo	115
Bao, Egude	199
Bellini, Pierfrancesco	221
Benatan, Matt	72
Bruno, Ivan	221
Caggiano, Sonia	284
Cai, Guoqiang	49
Cai, Yuanyuan	241
Caruccio, Loredana	274
Chang, Maiga	186
Chang, Shi-Kuo	212
Che, Xiaoping	241
Coccoli, Mauro	300
Colace, Francesco	62, 191
Costabile, Maria Francesca	115
Costagliola, Gennaro	14, 29, 257
Cuzzocrea, Alfredo	85
D'Apice, Ciro	179
De Angeli, Antonella	115
De Marsico, Maria	284
De Rosa, Mattia	14, 29
Del Fatto, Vincenzo	39, 124
Delaney, Aidan	108
Desolda, Giuseppe	115
Deufemia, Vincenzo	39, 274
Di Bitonto, Pierpaolo	148
Ding, Yaoming	264
Distasi, Riccardo	284
Dodero, Gabriella	124
Eloe, Nathan	250
Francesse, Rita	292
Fucella, Vittorio	14, 29, 257
Garn, Kristin	186
Giovanna, Castellano	23
Gravino, Carmine	292
Greco, Luca	191, 364



Grieco, Claudia	179
Guercio, Angela	300
Hammond, Tracy	101
Jungert, Erland	212
Kapetanakis, Stelios	108
Kara, Levent	1
Kuo, Rita	186
Lanzilotti, Rosa	115
Lena, Roberta	124
Leopold, Jennifer	250
Li, Bo-Shi	186
Li, Zhao	199
Liscio, Luca	179
Liu, Weibin	78, 172, 231
Lu, Wei	199, 241
Maresca, Paolo	300
Maria Alessandra, Torsello	23
Minuto, Andrea	131
Moissinac, Jean-Claude	162
Moro, Alessandro	85
Moscato, Vincenzo	62, 191
Mouratidis, Haralambos	108
Mumolo, Enzo	85
Nappi, Michele	284
Nesi, Paolo	155, 221
Ng, Kia	72
Nijholt, Anton	131
Pantaleo, Gianni	155
Paolino, Luca	39
Pascuccio, Fernando Antonio	257
Pesare, Enrica	148, 204
Picariello, Antonio	62, 191
Piscopo, Rossella	179
Pittarello, Fabio	59, 131
Polese, Giuseppe	274
Ren, Jinchang	65
Riccio, Daniel	284
Risi, Michele	292
Rodriguez, Bertha Helena	162
Roselli, Teresa	148, 204

Rossano, Veronica	148, 204
Sabharwal, Chaman	94
Sanesi, Gianmarco	155
Scanniello, Giuseppe	292
Shei, Shaun	108
Shi, Kailun	241
Stanganelli, Lidia	300
Taele, Paul	101
Torre, Ilaria	141
Tortora, Genoveffa	292
Tumiati, Sara	39
Ulu, Nurcan Gecer	1
Umeda, Kazunori	85
Vercelli, Gianni	85, 141
Wang, Guangwei	264
Wei, Ruxiang	78
Xing, Weiwei	78, 172, 199, 231
Xiong, Zenggang	264
Xu, Fang	264
Ye, Conghuan	264
Yin, Ruixue	172
Zhang, Xiaohui	231
Zhang, Xuemin	264
Zhao, Danyang	65
Zheng, Jiangbin	65

## Program Committee's Index

Bilal Alsallakh	Vienna University of Technology
Timothy Arndt	Cleveland State University
Mario Arrigoni Neri	
Danilo Avola	Sapienza University of Rome
Arvind Bansal	Kent State University
Andrew Blake	University of Brighton
Paolo Bottoni	Sapienza University of Rome
Paolo Buono	Dipartimento di Informatica - Università degli Studi di Bari Aldo Moro
	Athabasca University
Maiga Chang	University of Pittsburgh
Shikuo Chang	University of Brighton
Peter Chapman	
Ing-Ray Chen	
Shu-Ching Chen	Florida International University
Yuan-Sun Chu	National Chung Cheng University
Mauro Coccoli	DIBRIS - University of Genoa, Italy
Francesco Colace	Università degli studi di Salerno
Luigi Colazzo	Università di Trento
Gennaro Costagliola	Dipartimento di Informatica, Università di Salerno
Alfredo Cuzzocrea	ICAR-CNR and University of Calabria
Andrea De Lucia	Department of Mathematics and Informatics - University of Salerno
	University of Brighton
Aidan Delaney	Department of Management & Information Technology, Uni- versity of Salerno
Vincenzo Deufemia	Bonn-Aachen International Center for Information Technol- ogy B-IT
	University of Minnesota
Tiansi Dong	Athabasca University
	Università di Salerno
David Du	University of Brighton
Larbi Esmahi	Università di Brescia
Filomena Ferrucci	University of Salerno
Andrew Fish	
Daniela Fogli	
Rita Francese	
Vittorio Fuccella	
Kaori Fujinami	Tokyo University of Agriculture and Technology
David Fuschi	Bridging Consulting Ltd
Ombretta Gaggi	Dept of Mathematics, University of Padua
Angelo Gargantini	University of Bergamo
Nikolaos Gkalelis	ITI
Angela Guercio	Kent State University at Stark
Bob Heller	Athabasca University
Carlos A. Iglesias	Universidad Politécnica de Madrid
Pedro Isaías	Universidade Aberta
Erland Jungert	FOI (Swedish Defence Research Agency)
Levent Kara	CMU
Yau-Hwang Kuo	National Cheng Kung University

Robert Laurini	INSA Lyon
Jennifer Leopold	
Fuhua Lin	Athabasca University
Alan Liu	National Chung Cheng University
Jonathan Liu	
Luana Micallef	Helsinki Institute for Information Technology HIIT
Nikolay Mirenkov	University of Aizu
Andrea Molinari	University of Trento
Paolo Nesi	
Kia Ng	
Max North	Southern Polytechnic State University
Paolo Maresca Paolo Maresca	University Federico II
Ignazio Passero	Dipartimento di Matematica ed Informatica, Università degli Studi di Salerno, Italy
	New Mexico State University
Joseph Pfeiffer	University of Bari
Antonio Piccinno	University of Salerno
Giuseppe Polese	University of Salerno
Michele Risi	University of Kent
Peter Rodgers	Department of Computer Science - University of Bari
Teresa Roselli	Department of Computer Science - University of Bari
Veronica Rossano	Dipartimento di Matematica, Informatica ed Economia
Giuseppe Scanniello	Dipartimento di Management and Information Technology (DISTRAT) - Università di Salerno
Monica Sebillio	Arizona State University
	University of Tsukuba
Panchanathan Sethuraman	Kettering University
Buntarou Shizuki	
Peter Stanchev	University of Brighton
Lidia Stanganelli	
Gem Stapleton	Wellesley College
Mahbubur Syed	Dipartimento di Matematica e Informatica- Università di Salerno
Franklyn Turbak	Japan Advanced Institute of Science and Technology
Giuliana Vitiello	Czech Technical University in Prague
	University of Texas at Dallas
Atsuo Yoshitaka	
Tomas Zeman	
Kang Zhang	

## External Reviewers' Index

Caruccio, Loredana  
Cheng, Bo-Chao  
Di Nucci, Dario  
Fasano, Fausto  
Guan, Sheng  
Jiang, Jung Yi  
Li, Yi-Na  
Palomba, Fabio  
Tsai, Jen-Sheng  
Tsai, Wen-Hao