DMS 2011

Proceedings of the Seventeenth International Conference on Distributed Multimedia Systems

Florence, Italy
August 18-20, 2011

PROCEEDINGS

DMS 2011

The 17th International Conference on Distributed Multimedia Systems

Sponsored by

Knowledge Systems Institute Graduate School, USA

Co-Sponsored by

Distributed Systems and Internet Technology Lab, University of Florence, Italy
ICSRiM, University of Leeds, United Kingdom
Visual Design and Engineering Lab, Mechanical Engineering Department,
Carnegie Mellon University, USA
Eco Controllo SpA, Italy

Technical Program

August 18-20, 2011
Convitto della Calza Florence, Italy

Organized by

Knowledge Systems Institute Graduate School

ISBN 1-891706-30-6 (Paper)

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:office@ksi.edu
http://www.ksi.edu

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

Printed by Knowledge Systems Institute Graduate School

DMS 2011 Foreword

Welcome to DMS 2011, the 17th International Conference on Distributed Multimedia Systems.

With today's proliferation of multimedia data (e.g., images, animations, video, and sound), comes the challenge of using such information to facilitate data analysis, modeling, presentation, interaction, and programming. This is particularly important for the ever increasing number of end-users who are domain experts, but not IT professionals. Accordingly, the theme of this year's conference is "Multimedia Inspired Computing."

The conference is organized into seven primary distributed multimedia systems sessions, and eleven sessions on a variety of specialized themes, including: slow intelligence systems, data compression, visual languages and computing, distance education technology, geo-visual computing, and multimedia arts. The selection of papers to be presented at the conference was based upon a rigorous review process, with an acceptance rate of about 40% of the submissions received in the category of full research papers. The conference program also includes short papers that report on ongoing research activities and applications.

In addition, we are very pleased to have two invited keynote speakers at this year's conference: Dr. Stefano Levialdi (University of Rome, Italy), a distinguished researcher in visual languages and human-computer interaction, and Dr. Herman Helbig (University of Hagen, Germany), a world renowned expert in knowledge representation and the semantics of natural language.

The DMS Conference continues to be an internationally diverse research gathering. This year we are expecting over 66 authors and guests from 18 countries: Austria, Canada, China, Columbia, France, Germany, Greece, Iran, Italy, Japan, Pakistan, Singapore, Spain, Sweden, Taiwan, Tunisia, the United Kingdom, and the United States.

As program co-chairs, we deeply express our gratitude to the dedicated program committee members and conference support staff who have contributed to making DMS 2011 a success. We hope that you find this year's conference to be an invigorating exchange of research ideas, and that you include some time to enjoy the beautiful sights, sounds, and tastes of Florence, Italy!

Gennaro Costagliola and Jennifer Leopold
DMS 2011 Program Co-Chairs

The 17th International Conference on Distributed Multimedia Systems (DMS 2011)

August 18-20, 2011
Convitto della Calza Florence, Italy

Conference Organization

Steering Committee Chair

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

Conference Co-Chairs

Paolo Nesi, *University of Florence, Italy*
Kia Ng, *University of Leeds, UK*

Program Co-Chairs

G. Costagliola, *University of Salerno, Italy*
Jennifer Leopold, *Missouri University of Science and Technology, USA*

Program Committee

Arvind K. Bansal, *Kent State University, USA*
Pierfrancesco Belllini, *University of Florence, Italy*
Ing-Ray Chen, *Virginia Tech (VPI&SU), USA*
William Cheng-Chung Chu, *Tunghai University, Taiwan*
Gennaro Costagliola, *University of Salerno, Italy*
Alfredo Cuzzocrea, *University of Calabria, Italy*
Andrea De Lucia, *University of Salerno, Italy*
Sergio Di Martino, *Universita degli Studi di Napoli Federico II, Italy*
David H. C. Du, *University of Minnesota, USA*
Larbi Esmahi, *National Research Council of Canada, Canada*

Daniela Fogli, *Universita degli Studi di Brescia, Italy*
Farshad Fotouhi, *Wayne State University, USA*
Kaori Fujinami, *Tokyo University of Agriculture and Technology, Japan*
Ombretta Gaggi, *University of Padova, Italy*
Jivesh Govil, *Cisco Systems Inc., USA*
Angela Guercio, *Kent State University, USA*
Niklas Hallberg, *FOI, Sweden*
Hewijin Christine Jiau, *National Cheng Kung University, Taiwan*
Erland Jungert, *Linkoping University, Sweden*
Wolfgang Klas, *University of Vienna, Austria*
Yiannis Kompatsiaris, *Informatics and Telematics Institute - CERTH, Greece*
Yau-Hwang Kuo, *National Cheng Kung University, Taiwan*
Alan Liu, *National Chung Cheng Univeristy, Taiwan*
Vincent Oria, *New Jersey Institute of Technology, USA*
Sethuraman Panchanathan, *Arizona State University, USA*
Antonio Piccinno, *University of Bari, Italy*
Fabio Pittarello, *University of Venice, Italy*
Giuseppe Polese, *University of Salerno, Italy*
Buntarou Shizuki, *University of Tsukuba, Japan*
Peter Stanchev, *Kettering University, USA*
Christian Timmerer, *Klagenfurt University, Austria*
Genny Tortora, *University of Salerno, Italy*
Ellen Walker, *Hiram College, USA*
Chi-Lu Yang, *Institute for Information Industry, Taiwan*
Kang Zhang, *University of Texas at Dallas, USA*
Roberto Zicari, *Goethe University Frankfurt, Germany*

Publicity Chair

Kao-Shing Hwang, *National Chung Cheng University, Taiwan*

Proceedings Cover Design

Gabriel Smith, *Knowledge Systems Institute Graduate School, USA*

Conference Secretariat

Judy Pan, *Chair, Knowledge Systems Institute Graduate School, USA*
Norrjhan Ali, *Knowledge Systems Institute Graduate School, USA*
Dennis Chi, *Knowledge Systems Institute Graduate School, USA*
David Huang, *Knowledge Systems Institute Graduate School, USA*
Rachel Lu, *Knowledge Systems Institute Graduate School, USA*

International Workshop on Distance Education Technologies (DET 2011)

August 18-20, 2011
Convitto della Calza Florence, Italy

Conference Organization

Steering Committee Chair

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

Workshop Co-Chairs

Rita Francese, *University of Salerno, Italy*
Mahbub Syed, *Minnesota State University, USA*
Kazuo Yana, *Hosei University, Japan*

Program Co-Chairs

Oscar Fuhua Lin, *Athabasca University, Canada*
Teresa Roselli, *University of Bari, Italy*
Veronica Rossano, *University of Bari, Italy*

Program Committee

Giovanni Adorni, *University of Genova, Italy*
Tim Arndt, *Cleveland State University, USA*
Yuan-Sun Chu, *National Chung Cheng University, Taiwan*
Luigi Colazzo, *University of Trento, Italy*
Massimo De Santo, *Universita degli Studi di Salerno, Italy*
Pierpaolo Di Bitonto, *University of Bari, Italy*
Larbi Esmahi, *Athabasca University, Canada*
Rita Francese, *University of Salerno, Italy*
Robert Heller, *Athabasca University, Canada*
Kao-Shing Hwang, *National Chung Cheng University, Taiwan*

Publicity Committee

International Workshop on Visual Languages and Computing (VLC 2011)

August 18-20, 2011
Convitto della Calza Florence, Italy

Conference Organization

Steering Committee Chair

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

Workshop Chair

Giuseppe Santucci, *University of Rome, Italy*

Program Chair

Levent Burak Kara, *Carnegie Mellon University, USA*

Program Committee(tentative)

Christine Alvarado, *Harvey Mudd College, USA*
Dorothea Blostein, *Queen's University, Canada*
Jeff Browne, *UC Santa Barbara, USA*
Alfonso F. Cardenas, *University of California, USA*
Gennaro Costagliola, *University of Salerno, Italy*
Philip Cox, *Dalhousie University, Canada*
Sergiu Dascalu, *University of Nevada, USA*
Aidan Delaney, *University of Brighton, UK*
Vincenzo Deufemia, *University of Salerno, Italy*
Ellen Do, Georgia *Institute of Technology, USA*
Filomena Ferrucci, *University of Salerno, Italy*
Manuel J. Fonseca, *INESC-ID, Portugal*
Luoting Fu, *Carnegie Mellon University, USA*
Vittorio Fuccella, *University of Salerno, Italy*
Maolin Huang, *University of Technology, Sydney, Australia*

Table of Contents

Semantic Computing

Humanities, Emergency & Health Care Management

GeoVisual Computing

Intelligent Multimedia Computing

Collaborative & Social Multimedia

Multimedia Databases & Info Systems

Multimedia Software Engineering

Mobile Intelligent Applications

DET Workshop

Computer Supported Collaboration

Virtual World, Serious Game in Education

Personalized learning and Instructional Design

New learning supported by new technologies

VLC Workshop

Visual Languages & Computing I

Visual Languages & Computing II

Visual Languages & Computing III

Poster/Demo: Multimedia Arts & DMS/VLC Demos

Note: (S) indicates a short paper.
(P) indiecates a poster or demo, which is not a refereed paper.

Keynote I
Ambient Intelligence Today/Tomorrow

Stefano Levialdi
IEEE Life Fellow

Abstract

Some prior definitions are given as well as some examples of interesting, possible applications developed in different laboratories around the world. Seamless integration and the importance of using all the human senses are described as well as the two main views, an optimistic one and a pessimistic one. The big problem is: will it make our life a better one (or not?). This issue is analyzed through a number of different, well motivated positions taken by various researchers, the conclusion is left to the audience.

About the Speaker

Stefano Levialdi graduated from Buenos Aires University in Telecommunication Engineering (1959), taught a number of courses at the Universities of Genoa, Naples, Bari and Rome in Computer Science (1961-1983). He has worked over 20 years in parallel image processing, and lastly, since 1984, in visual languages and human/computer interaction. He has published over 250 papers with more than 150 co-authors. He won a Chair of Excellence at the Carlos III University in Madrid (2009/10). He is IEEE Life Fellow (1988), Pattern Recognition Fellow (1994), Visual Language/Human-Centric Computer Fellow (2008) and co-editor of the Journal of Visual Languages and Computing (Elsevier Press) since 1990.

Keynote II
Multilayered Extended Semantic Networks as a Knowledge Representation Paradigm and Interlingua for Meaning Representation

Hermann Helbig
University at Hagen, Germany

Abstract

The talk gives an overview of Multilayered Extended Semantic Networks (abbreviated MultiNet), which is one of the most comprehensively described knowledge representation paradigms used as a semantic interlingua in large-scale NLP applications and for linguistic investigations into the semantics and pragmatics of natural language. As with other semantic networks, concepts are represented in MultiNet by nodes, and relations between concepts are represented as arcs between these nodes. Additionally to that, every node is classified according to a predefined conceptual ontology forming a hierarchy of sorts, and the nodes are embedded in a multidimensional space of layer attributes and their values. MultiNet provides a set of about 150 standardized relations and functions which are described in a very concise way including an axiomatic apparatus, where the axioms are classified according to predefined types. The representational means of MultiNet claim to fulfill the criteria of universality, homogeneity, and cognitive adequacy. In the talk, it is also shown, how MultiNet can be used for the semantic representation of different semantic phenomena. To overcome the quantitative barrier in building large knowledge bases and semantically oriented computational lexica, MultiNet is associated with a set of tools including a semantic interpreter WOCADI for automatically translating natural language expressions into MultiNet networks, a workbench LIA for the computer lexicographer, and a workbench MWR for the knowledge engineer for managing and graphically manipulating semantic networks. MultiNet has been used in practical applications like Natural Language Interfaces to the Internet (NLI-Z39.50), Semantic Recognition of Duplicates and Plagiates (SemDupl), Readability Checkers (DeLite), and in the commercial search engine SEMPRIA Search. Among the newest projects, where MultiNet and its language technology are employed as cornerstones, is a semantically based translation system 'German . Chinese' (Europhon project).

About the Speaker

Prof. Hermann Helbig received in 1968 M.S. in Physics from the University of Leipzig (Diploma in Quantum Theory). From 1968 to 1969 he was research assistant at the University of Leipzig. From 1970 to 1989 he was researcher in the fields of Artificial Intelligence (AI) and Computational Linguistics (CL) (Robotron, Dresden, Head of the AI Laboratory). He has the following accomplishments: (1) Creation of the word-class controlled functional analysis (Development of a semantically driven parser); (2) Development of the Natural Language Interface NLI-AIDOS; (3) Development of the Question-Answering System FAS-80. In 1976 he received the Ph.D. in Computer Science

(Promotion, Dr.rer.nat., in the field of AI). In 1986 he completed Habilitation (Dr.rer.nat.habil.) in the field of Knowledge Representation. He accomplished the following: (4) Creation of the knowledge representation paradigm ‘Multilayered Extended Semantic Networks’ (MultiNet). From 1988 to 1992 he was Lecturer for Artificial Intelligence at the TU Dresden. From 1989 to 1992 he was Senior researcher at SRS Dresden and Siemens-Nixdorf. He accomplished the following: (5) Development of Man-Machine Interfaces and Geographic Information Systems. From 1992 to 2008 he received Full Professorship at the University in Hagen, and became Head of the Chair: Intelligent Information and Communication Systems. The main research results of the chair (http;//pi7.fernuni-hagen.de/research/) are: (6) Creation of the largest semantically oriented computational lexicon (HaGenLex) in Germany; (7) Logical foundations of the MultiNet paradigm; (8) Natural language interface NLI Z39.50 to the Internet; (9) Workbench for the knowledge engineer MWR+; (10) Workbench for the computer lexicographer LIA+; (11) Virtual electronic laboratory VILAB for supporting courses in AI and Computational Linguistics. Since 2008 he was Emeritus, Head of the Working Group .Intelligent Information and Communication Systems. From 1997 to 2002 he had sabbatical stays at ICSI and lectured in Berkeley, CA, at the University of New York at Buffalo (USA), and at the Universities of Edinburgh, Sheffield and London (Great Britain). In 2006 he had sabbatical visits and lectured at CMU Pittsburgh, Universites of Toronto (Canada) and Rochester NY, and at MIT.

Proceedings
International Conference on Distributed Multimedia Systems
(DMS 2011)

Co-Editors

G. Costagliola, *University of Salerno, Italy*
Jennifer Leopold, *Missouri University of Science and Technology, USA*

Design and Modeling of Topic/Trend Detection System By Applying Slow Intelligence System Principles

Ji Eun Kim, Yang Hu and Shi-Kuo Chang
Department of Computer Science, University of Pittsburgh, Pittsburgh, PA 15260, USA
{kim, huyang, chang}@cs.pitt.edu
and
Chia-Chun Shih and Ting-Chun Peng
Institute for Information Industry, Taiwan, Republic of China
{chiachun, markpeng}@iii.org.tw

Abstract

This paper first describes the characteristics of the Slow Intelligence System (SIS), and then provides an example of Petri Net modeling of the topic/trend detection system by applying the SIS principles. The Petri Net model is simulated using PIPE3 tool. Based upon the requirements for Slow Intelligence Systems, a survey of existing development framework leads to suggestions for a generic framework for SIS system development.

Keywords

Component-based software system, slow intelligence system, Petri net modeling, slow intelligent system, SIS framework

1 Introduction

With a trend of mass use of social media such as blogs, and social networking in Internet, the detection of trends and hot topics out of millions of heterogeneous data become important not only to the users of the systems but also commercial stakeholders around these kind of systems. For example, HP Labs have demonstrated that social media can be effective indicators for predicting movie revenues [2]. Most of social medial platform providers adopt the technologies from Information Retrieval research to provide customized and the most relevant services and advertisements to the users. However, the data generated by users through social media is growing exponentially, this requires new approaches to collect and process many heterogeneous data available on the web efficiently. The new approaches should reduce the search space by intelligently selecting the most relevant data sources on the web, and adapt data processing algorithms to handle varieties of the data automatically or semi-automatically [15].

A Slow Intelligence System (SIS) [16] is a general-purpose system characterized by being able to improve performance over time through a process involving *enumeration, propagation, adaptation, elimination and concentration,* which are the characteristics of the Slow Intelligence System. A SIS continuously learns, searches for new solutions and propagates and shares its experience with other peers. A SIS differs from an expert system in that the learning is not always obvious. A SIS seems to be a slow learner because it analyzes the environmental changes and carefully and gradually absorbs that into its knowledge base while maintaining synergy with the environment. A Slow System in general has two decision cycles – a quick decision cycle providing an immediate response to the environment and a slow decision cycle that tries to follow the gradual changes in the environment by analyzing the information acquired from experts and past experiences so that the performance of the system is improved over time.

The online topic/trend detection system in [15] adopted Slow Intelligence for their design so that computing resources are gradually concentrated on prospect solutions. We employ their design approaches and model an online topic/trend detection system in Petri Net model. In addition, we recommend the SIS generic framework, which integrates the subsets of selected existing software frameworks in order to support the entire lifecycle of the SIS.

The remainder of this paper is organized as follows. Section 2 illustrates Petri Net modelling of topic/trend detection system as a case study and shows the simulation results. Section 3 provides a survey of existing software frameworks with respects to requirements for SIS systems. Finally, recommendations for the generic SIS framework are given in Section 4.

2 Modelling of Topic/Trend Detection System in SIS

The online topic/trend detection system (TDT) proposed by [15] is to detect current hot topics and to predict future hot topics based on data collected from the Internet. Since it is unlikely to collect all data on the Internet, the system requires users to provide their information needs, including their concerned keywords and their concerned websites.

Furthermore, since hot topics change quickly, the system requires periodical updates in hourly or daily intervals. The system first collects latest data from Internet based on users' information needs by Crawler & Extractor, then adopts TDT techniques to discover current hot topics by Topic Extractor, and finally applies trend estimation algorithms [3] to predict hot topics by Trend Detector.

The *Crawler & Extractor* building block is for the restriction of web data in the topic trend detection. It is modelled as SIS system principle. Currently *Topic Extractor* building block is a black box component, which can be later modelled and refined. *Trend Detector* building block is modelled as SIS system.

2.1 Crawler & Extractor

The responsibility of *Crawler* is to collect web pages from Internet. Crawler needs to be selective, that is, only collect web pages that satisfy predefined requirements. Extractor is responsible to extract information from web pages.

Crawler

Several approaches have been proposed in the literature to restrict the media resources to crawl. The approaches include focused crawling [5], contextual crawling, semantic web and genetic-based crawling [7]. We observed that most of the existing approaches are close to the core concept of the SIS, which can be modelled as *enumerator*, *adaptor*, *eliminator* and *concentrator*.

We select the focused crawler as a basic model for the crawlers in the SIS in this project. However, the focused crawler can be replaced with another crawler depending on the different requirements and constraints for the target topic trend detection system.

The focused crawler consists of two phases: classifier and distiller. The first state of the classifier is to come up with the most relevant URLs for the topic trend detection. The list of URLs from the focused crawler can be different for each user because the SIS interacts with each user to enumerate example URLs and then recommend relevant URLs to the user. Then the focused crawler in SIS trains the knowledge base to get the most relevant URLs. In other words, we can map the classification phase of the focused crawler approach to *enumerator* and *adaptor*. Once training is done, SIS will identify the most relevant hubs by running a topic distillation algorithm. The results of this run raise the priorities of hubs and immediate neighbours. This distillation phase is mapped to eliminator in the SIS concept. The last phase is to report the most popular sites and resources to the users. The users finally mark the results as useful or useless, and send feedback to classifier and distiller. Thereby classifier and distiller can concentrate their crawling data. This whole cycle of *enumerator*, *adaptor*, *eliminator* and *concentrator* can be repeated.

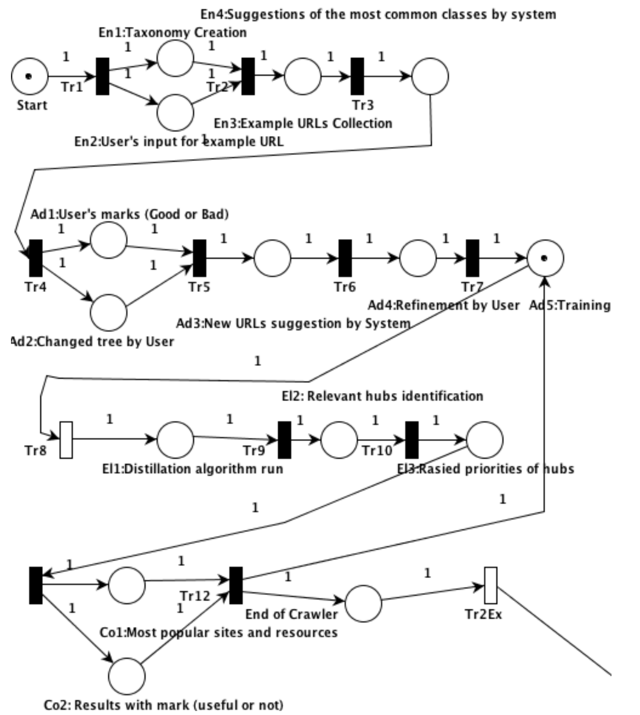


Figure 1. Petri Net of Focused Crawler

Extractor

Information extraction resources categorized into three folds: unstructured resources such as free text written in natural language, semi-structure resources such as HTML tables and structured resources including XML and relational database. The role of *Extractor* in the topic trend detection is to build the HTML tag tree, mine data regions, identify data records from each data region, learn the structure of a general data record and then extract the data.

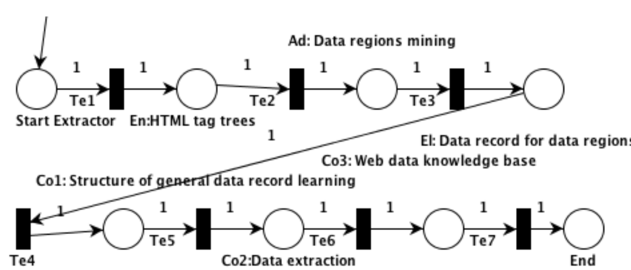


Figure 2. Petri Net of Extractor

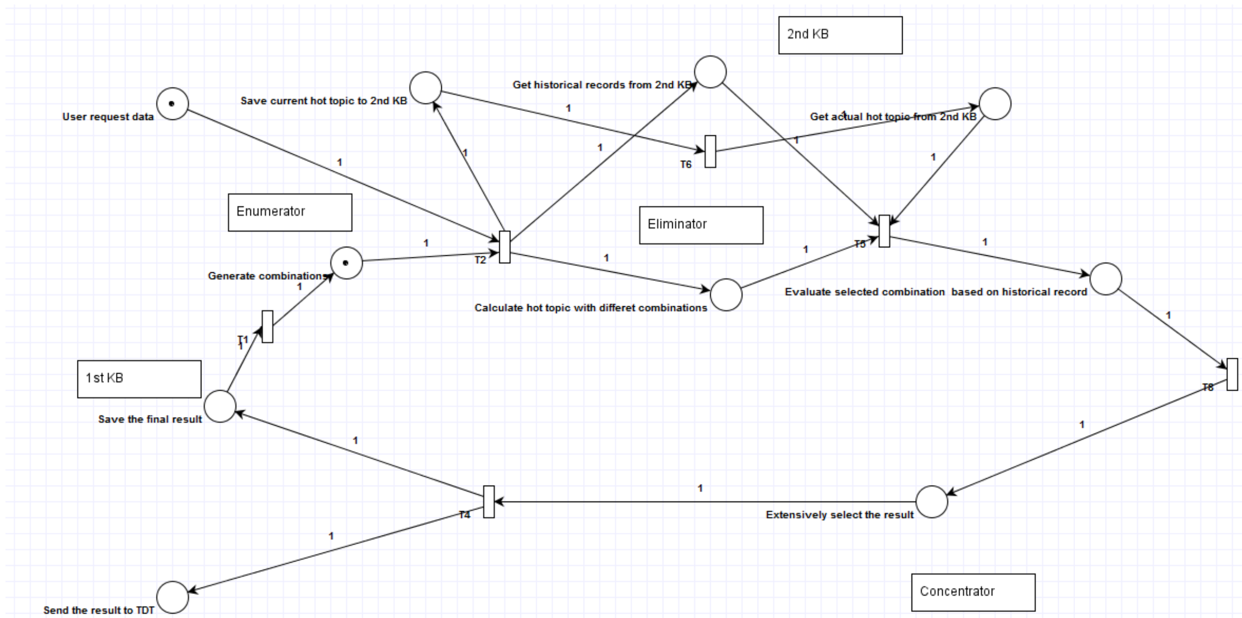


Figure 3. Petri Net of SIS

The extractor is also modelled in SIS. *Enumerator* can include building the HTML tag tree. *Adaptor* mines data regions. A data region is a collection of two or more generalized nodes, which have the same parents nodes and are adjacent. The regions are found using tree edit distance algorithm. Identifying the data records from data regions can be a role of *eliminator*. Learning the structure and extracting of the data can be done in the *concentrator* in SIS. For each data region we need to understand the structure of the data records in the region. Therefore, partial tree algorithm [19] is used to gather the structure. Extraction of data from single and multiple pages often introduces a lot of noise (e.g., advertisement and non-relevant data). During the *concentration* phase, a sample page is taken as the wrapper. Then this wrapper is refined to solve mismatches. The literature [19] found different types of mismatches, which include text, string mismatches and tag mismatches. By using partial alignment and wrapper solution, the different structures are merged and build a generic template for a data record. Finally the extracted data is stored in the web data as a knowledge base.

2.2 Topic Extractor

The responsibility of *Topic Extractor* is to detect hot topics from a set of text documents. The process of topic detection can be divided into the following steps, which adopt some state-of-the-art techniques: 1) topic word extraction: TF-IDF [14] scheme is applied to measure the importance of terms in a given text document and generates top-N topic word candidates for each text document. 2) topic word clustering: single-pass clustering [1, 17], a popular topic detection approach, is adopted to cluster related documents into associated topic groups. The centroid topic word of cluster with highest weighting score is treated as the representative name of each generated cluster, which represents an extracted "topic". 3) extract hot topics: hot topics derive from hot events in a particular timeline [1, 6].

The Topic Extractor is a black box component in our model. It connects *Crawler & Extractor* to *Trend Detector.*

2.3 Topic Detector

The responsibility of trend detector is to detect trends (future hot topics) based on currently available data. Figure 4 illustrates the subcomponents of the topic detector component, which consists of user request, SIS based TDT, knowledge base and dispatcher. Figure 5 illustrates how SIS principle applied to the topic detector. We model each subcomponent in Petri net.

Figure 4. Components in Topic Detector

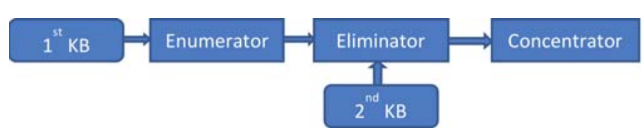


Figure 5. Hierarchy of the SIS based TDT

<u>User request handler</u>

This subcomponent (Figure 6) is responsible for collecting the entire user request from the web server then sends the collected data to the TDT component.

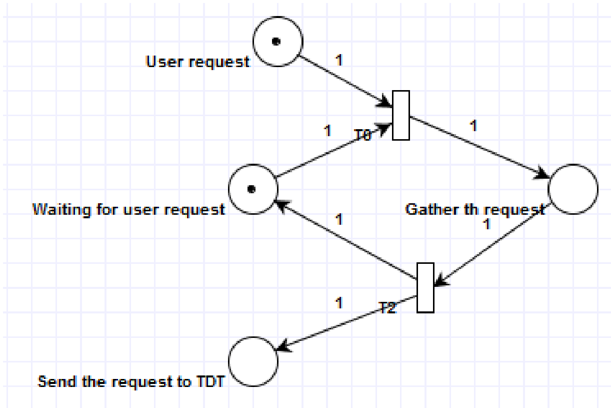


Figure 6. Petri Net of User Request Handler

Topic Detection and Tracking

TDT component (Figure 7) is responsible for finding the trend in the stream of user requests, and makes decisions in order to tune up the server cluster. Examples include sending the reducing power supply instruction to the group of servers that working on one topic when TDT component detect that this topic may cold in the future [8].

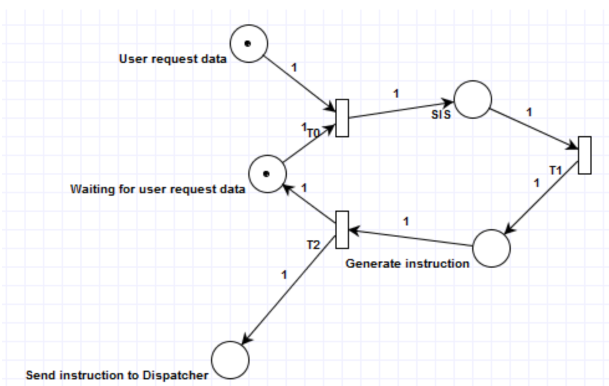


Figure 7. Petri Net of Topic Detection and Tracking

a. Slow Intelligence System

SIS is the fundamental framework for TDT who is going to use different methods to predict the trend. With its help, TDT can gradually improve performance and accuracy over time.

The input to this framework is the user request, which will be treated as the criteria of the current hot topic. Then the eliminator will calculate the current hot topic using different method's combination generated by the enumerator and compare the result with the user request. Finally, the concentrator will select the best combination according to the result generated by the eliminator to help SIS in predicting the trend.

In the meantime, the intermediate result will be kept in two knowledge bases.

b. 1st Knowledge Base

1st KB (Figure 8) is the core database, which contains two knowledge bases that can help SIS to select the best method over time.

1) Evaluation result knowledge base

The result about what method we have choice last time.

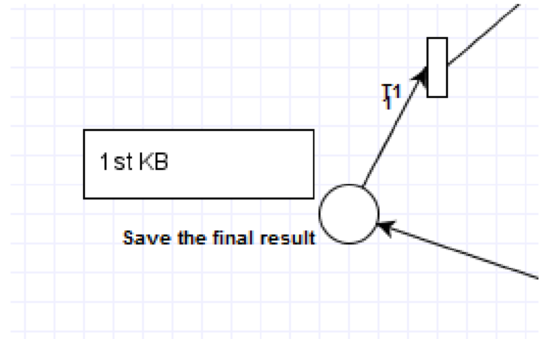


Figure 8. Petri Net of 1st KB

2) Sampled historical record knowledge base

2nd KB (Figure 9) saves the record for previous predicted trend and the actual hot trend at that time. So that the Eliminator can pick up the method and compare with these records to see which algorithm can generate the best result.

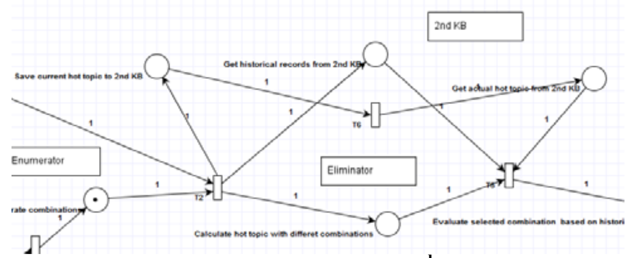


Figure 9. Petri Net of 2nd KB

Dispatcher

Dispatcher gets the instruction from the TDT server and the topic (marked with hot or cold), then sends the associated instruction to the group that dealing with this topic.

2.4 Composition of other components

Since the topic/trend detection systems are sequentially ordered with *Crawler &Extractor*, *Topic Extractor* and *Trend Detector*, the output from the Petri-Net model of the crawler and extractor is the input to the Topic Extractor component.

2.5 PIPE 3 Simulation Results

PIPE 3 [13] is an open source, platform independent tool for creating and analyzing Petri nets. We employ PIPE 3 and extend it to support timing model. Our extended PIPE 3 enables us to specify immediate execution, which is shown as -1 in Execution Time in simulation results or the static time for the component execution.

Name	Execute Time	Script Path	Source Code Path
T0	-1	D:\DropBox\T0.jar	
T10	-1	D:\DropBox\T10.jar	
T11	-1	D:\DropBox\T11.jar	

T12	-1	D:\DropBox\T12.jar
T13	-1	D:\DropBox\T13.jar
T14	-1	D:\DropBox\T14.jar
T15	-1	D:\DropBox\T15.jar
T16	-1	D:\DropBox\T16.jar
T2	-1	D:\DropBox\T2.jar
T20	-1	
T3	-1	D:\DropBox\T3.jar
T37	-1	
T4	-1	D:\DropBox\T4.jar
T5	-1	D:\DropBox\T5.jar
T6	-1	D:\DropBox\T6.jar
T7	-1	D:\DropBox\T7.jar
T8	-1	D:\DropBox\T8.jar
T9	-1	D:\DropBox\T9.jar
Te1	130	D:\DropBox\Te1.jar
Te2	140	D:\DropBox\Te2.jar
Te3	150	D:\DropBox\Te3.jar
Te4	160	D:\DropBox\Te4.jar
Te5	170	D:\DropBox\Te5.jar
Te6	180	D:\DropBox\Te6.jar
Te7	-1	D:\DropBox\Te7.jar
Tr1	1	D:\DropBox\Tr1.jar
Tr10	10	D:\DropBox\Tr10.jar
Tr11	110	D:\DropBox\Tr11.jar
Tr12	120	D:\DropBox\Tr12.jar
Tr2	10	D:\DropBox\Tr2.jar
Tr2Ex	20	D:\DropBox\Tr2ex.jar
Tr3	20	D:\DropBox\Tr3.jar
Tr4	40	D:\DropBox\Tr4.jar
Tr5	50	D:\DropBox\Tr5.jar
Tr6	60	D:\DropBox\Tr6.jar
Tr7	70	D:\DropBox\Tr7.jar
Tr8	10	D:\DropBox\Tr8.jar
Tr9	90	D:\DropBox\Tr9.jar

Figure 10. PIPE3 Simulation Results

3 Towards a Framework for Slow Intelligent System Development

As a part of the development of this novel software design methodology, we define a general framework for developing Slow Intelligence Systems. This paper discusses the development time and runtime aspect of SIS framework.

3.1 Requirements for SIS generic framework

We identify the six important requirements for the software frameworks for Slow Intelligent Systems (SIS) as follows:

ID	Description
R1	Java support is required to provide interoperability amongst different hardware and operating systems. Note that SIS shall provide different levels of interoperability by employing web services for distributed systems later. However, the current SIS framework requires platform independent environment to run components without supporting web services.
R2	Dynamic lifecycle management of building blocks during runtime is required because SIS principle allows a software component to be created, modified and removed dynamically.
R3	XML message based communications. The SIS components use XML messages which specify input and output components for each message type.
R4	IDE provides Code frame generation out of meta-models in a form of class diagram.
R5	Scalability of components up to 1,000 -10,000 components
R6	Coherent Integrated Development Environment (IDE) support to facilitate seamless development, simulation and deployment of the software components

3.2 Survey of existing frameworks

This section discusses the existing frameworks from development environment to runtime framework, as well as the relevant concepts for solving similar problems.

3.2.1 OSGi framework

The OSGi (Open Service Gateway Initiative) [12] service platform is a service platform for the Java programming language that implements dynamic component model. The bundle (software package with meta information) life cycle management such as start, stop, install and update is done via APIs that allow for remote downloading of management policies. The service registry allows bundles to detect the addition of new services, or the removal of services, and adapt accordingly. The OSGi specification that defines a dynamic component system for Java enables a development model where applications are dynamically composed of many different reusable components. The OSGi specifications enable components to hide their implementations from other components while communicating through services, which are objects specifically shared between components.

SIS can employee many interesting concepts from OSGi for the SIS system. The software component life-cycle management (e.g., install, start, stop, update software bundles) feature can be a solution for our requirement of

R2 (Dynamic lifecycle management of software building blocks during runtime) and OSGi framework is directly usable to the SIS framework. Distributed *EventAdmin* service is applicable to realize the requirement of R3 (XML message based communication). In addition, OSGi's dynamic service handling concept and declarative service concept can be employed to SIS framework in order to support R2.

3.2.2 Eclipse EMF

EMF [10] is an Eclipse plug-in, which supports the java code frame generation based on *Ecore* model, which is a UML representation of classes. Once users specify an EMF model, the EMF generator can create a corresponding set of Java implementation classes. Users can edit these generated classes to add methods and instance variables and still regenerate from the model as needed: users' additions will be preserved during the regeneration. If the code the user added depends on something that users changed in the model, users will still need to update the code to reflect those changes; otherwise, users' code is completely unaffected by model changes and regeneration.

Therefore, it can support R4 (Code generation) and can be used together with Eclipse IDE. In order to make EMF useful, the SIS model should be available in a form of *Ecore* model or other supported UML formats such as Rational Rose.

3.2.3 Eclipse ECF

Eclipse ECF (Eclipse Communication Framework) [9] is an Eclipse plug-in, which provides a communications container. ECF containers represent access to a protocol-specific communications context. ECF can be integrated into SIS framework's communication concept realization. ECF supports both *point-to-point* and *publish-and-subscribe* communication methods, which can be important realization of SIS's message based communication for the distributed building blocks. ECF generic container can be usable for the SIS domain specific message contents.

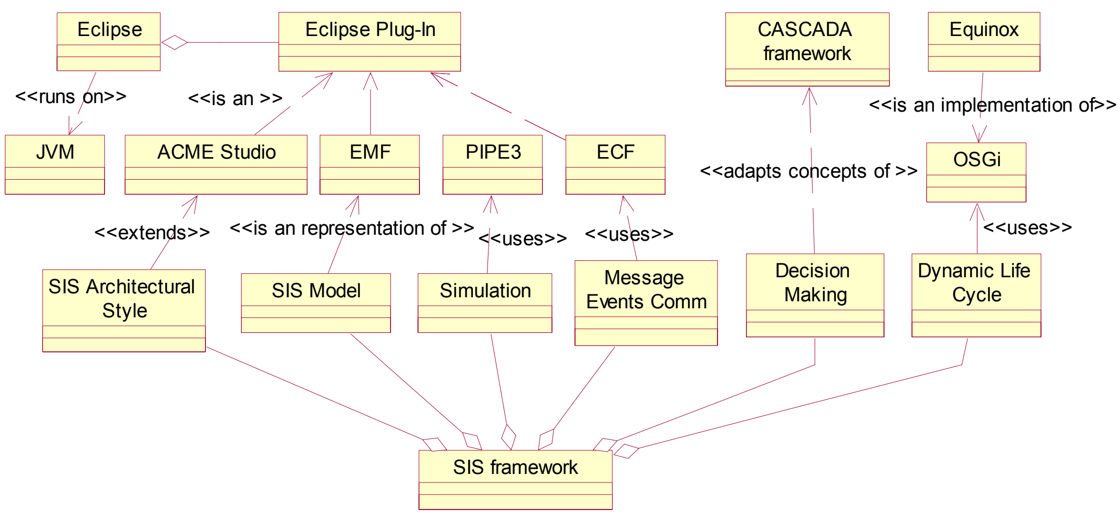


Figure 11. Relationship of the investigated technologies with respect to tool-chains

3.2.4 ACME

ACME [8] is a software architecture description language to support software architecture, design and analysis. It provides an architectural ontology consisting of seven basic architectural design elements, a type mechanism for abstracting common, reusable architectural idioms and styles and an open semantic framework for reasoning about architectural description. ACME language and the Acme Studio tool are possibly integrated into the SIS framework. ACME supports the extensibility of the domain specific architectural styles. SIS may provide a catalog of architectural styles, which are commonly used for the SIS components. Depending on what properties we want to assure during the design time, we can extend ACME to enable SIS specific analysis.

3.2.5 CASCADA

CASCADA [11] aims to provide an autonomic component-based framework to support the deployment of a novel set of services through development of distributed applications capable of coping well with uncertain environments by dynamically adapting their plans as the environment changes in uncertain ways. To enable dynamic software component/services management over time, their architectural concept to support plan, execution, and self-* model can be adaptable to SIS purpose.

Figure 11 illustrates the relationships between different technologies and how each technology can be applied to the SIS framework. We envision developing SIS systems using Eclipse IDE. The core features of the SIS framework shall be realized by extending existing technologies, as Eclipse plug-ins. SIS systems are compliant to SIS architectural styles, which is extendable in ACME. The meta model for SIS components shall be represented using Eclipse EMF. Dynamic lifecycle management of SIS components can be realized by using OSGi framework. SIS system supports message based communication through Eclipse ECF. The existing PIPE 3 tool shall be also ported to an Eclipse plug-in. The SIS approaches may adapt CASCADA's self-* model to add context-aware decision making mechanism in addition to the current SIS principle.

4 Conclusion

We recommend the SIS generic framework use Eclipse as an IDE and build the SIS software framework by employing existing framework such as OSGi. For the robust development environment support, we need more research on the meta-model description for the SIS, and integrate the meta-model to the target SIS architecture. To support architectural design, we recommend adding SIS architectural styles to the existing ACME. The concept of the CASCADAS project is also recommended to take consideration to support decision-making process of SIS. In addition, implementing the PIPE 3 tool as an Eclipse plug-in can provide seamless and coherent development, simulation and deployment environment by leveraging Eclipse ecosystem.

Acknowledgements

This study is conducted under the "Innovative and Prospective Technologies Project" of the Institute for Information Industry, which is subsidized by the Ministry of Economy Affairs of the Republic of China.

References

[1] Allan, J., Carbonell, J., Doddington, G., Yamron, J., and Yang, Y., "Topic detection and tracking pilot study: Final report," *Proceedings of the DARPA broadcast news transcription and understanding workshop*,1998

[2] Asur, S. and Huberman, B., "Predicting the Future With Social Media," Arxiv preprint arXiv:1003.5699, 2010

[3] Bianchi, M., Boyle, M., and Hollingsworth, D., "A comparison of methods for trend estimation," *Applied Economics Letters*, vol. 6, pp.103-109, 1999

[4] Belady, Christian. "In the Data Center, Power and Cooling Costs More Than the IT Equipment it Supports." *Electronics Cooling*. Vol. 23, No. 1, February 2007.

[5] Chakrabarti, S., Van den Berg, M., and Dom, B., "Focused crawling: a new approach to topic-specific Web resource discovery," *Computer Networks*, vol. 31, pp. 1623-1640, 1999.

[6] Chen, K. Y., Luesukprasert, L., and Chou, S., "Hot topic extraction based on timeline analysis and multidimensional sentence modeling," *IEEE transactions on knowledge and data engineering*, vol. 19, pp.1016-1025, 2007

[7] Chia-Hui Chang, Mohammed Kayed, Moheb Ramzy Girgis, Khaled Shaalan., "A survey of web information extraction systems," *IEEE transactions on knowledge and data engineering*, vol. 18, pp.1411-1428, 2006

[8] David Garlan, Robert T. Monroe, and David Wile, *Acme: Architectural Description of Component-Based Systems*, Foundations of Component-Based Systems, Gary T. Leavens and Murali Sitaraman (eds), Cambridge University Press, 2000, pp. 47-68.

[9] Eclipse ECF, Available at: http://www.eclipse.org/ecf/

[10] Eclipse EMF, Available at: http://www.eclipse.org/modeling/emf/

[11] Luciano Baresi, Antonio Di Ferdin and Antonio Manzalini, *The CASCADAS Framework For Autonomic Communications*, Autonomic Communication, Springer, 2009

[12] OSGi Alliance, Available at: http://www.osgi.org

[13] PIPE 3, Available at: http://code.google.com/p/pipe-3-development/

[14] Salton, G. and Yang, C. S., "On the specification of term values in automatic indexing," *Journal of documentation*, vol. 29, pp. 351-372, 1973.

[15] Shi and Peng, "Building Topic/Trend Detection System based on Slow Intelligence," *Proceedings of the 16th International Conference on Distributed Multimedia Systems (DMS)*, pp. 53-56, 2010.

[16] Shi-Kuo Chang, "A General Framework for Slow Intelligence Systems," *International Journal of Software Engineering and Knowledge Engineering*, Volume 20, Number 1, February 2010, 1-16.

[17] Smith, D. A., "Detecting and browsing events in unstructured text," *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval*, 2002, pp. 73-80

[18] U.S. Environmental Protection Agency. 2007. *EPA Report to Congress on Server and Data Center Energy Efficiency*.

[19] Yanhong Zhai, "Web data extraction based on partial tree alignment," *Proceedings of the 14th international conference on World Wide Web*, 2005, p.85

High Dimensional Feature Selection via a Slow Intelligence Approach

Yingze Wang and Shi-Kuo Chang

Department of Computer Science, University of Pittsburgh,
Pittsburgh, PA 15260, USA
{yingzewang,chang}@cs.pitt.edu

***Abstract*— Feature selection in high dimensional space is the hot topic in contemporary machine learning area. In the past decade, a lot of effort has been devoted in developing various feature selection algorithms. However, each feature selection method has its own advantage for different datasets or applications. Therefore, one might face the difficulty of choosing the most suitable feature selection method in practice. Reference [1] proposed a new framework called Slow Intelligence System, which can continuously learn, search the appropriate method and propagate information according to the environment to improve the performance over time. In this paper, we adopt this slow intelligence idea in the application of high dimensional feature selection. We propose a new framework which could dynamically choose feature selection algorithm from a pool of methods according to the status quo. We apply our method to cancer datasets with the expression levels of thousands of genes as features. The experimental result shows that our method is superior to individual state-of-the-art feature selection methods.**

Keywords-Feature selection, machine learning, slow intelligence system

I. INTRODUCTION

High-dimensional feature selection is a hot topic in statistics and machine learning areas. It has many important applications in supervised learning. For example, in gene selection problems, we have gene expression levels of tens of thousands of genes as features for a number of patients as samples. A typical classification task is to separate cancer patients from the healthy ones or to distinguish different types of patients for helping regimens and treatment. There are two important issues of this application that motivate the efficient and scalable feature selection algorithm. On one hand, we may only have fewer than 100 patients (samples), it is important to eliminate those irrelevant genes (features) to obtain a robust classifier. On the other hand, from biological point of view, one wants to know which genes are the most critical factors to the cancer.

In the past decade, a lot of feature selection algorithms have been proposed. These feature selection algorithms can be roughly classified into two categories: convex regularization approach and stepwise greedy pursuit approach. The first approach regularizes the generalized linear models by adding a sparsity-inducing penalty function. The most representative algorithms include L1-norm regularized generalized linear model, e.g. lasso for linear model [2] and sparse logistic regression [3]; and elastic-net regularized generalized linear model [4]. The second approach, the stepwise greedy pursuit approach, iteratively selects the current optimal feature according to some criteria, leading to the method as forward stepwise regression, or orthogonal matching pursuit. However, it is hard to compare all these different methods. One method may be superior to another for different datasets or in different applications. In practice, one might face the difficulty of choosing the appreciate feature selection algorithm. Another problem is that for the convex optimization approach, when the dimension is ultra-high, it may not be able to provide the solution due the computational limitations.

Recently, Chang et. al [1] proposed a very general learning framework called slow intelligence approach/system (SIS). They also presented the design of component based SIS [13] and the user interface design to produce and manage the generic SIS system [14]. In this paper, we use the idea of SIS and specialize it to the feature selection task. We propose a new feature selection system, which uses different feature selection algorithms as the candidate methods. Our system selects the feature using different methods, searches for the best method according to the status quo and propagates the learned features over iterations. As we show, it gives the superior feature selection performance and can handle ultra-high dimensional data.

The paper is organized as follows. In Section II, we discuss related works about the slow intelligence system and the state-of-the-art feature selection algorithms. In Section III, our feature selection system using slow intelligence approach is presented. In Section IV, the comprehensive algorithm for whole system is shown. Also we discuss the detailed issues of time controller and knowledge base design. In Section V, we demonstrate the superiority of our system by applying it to the real gene selection problems for leukemia and colon cancer datasets. In Section VI, we give the conclusion of our work and point out some possible future works.

II. RELATED WORK

A. *Slow Intelligence System*

We will first introduce the concept of slow intelligence and

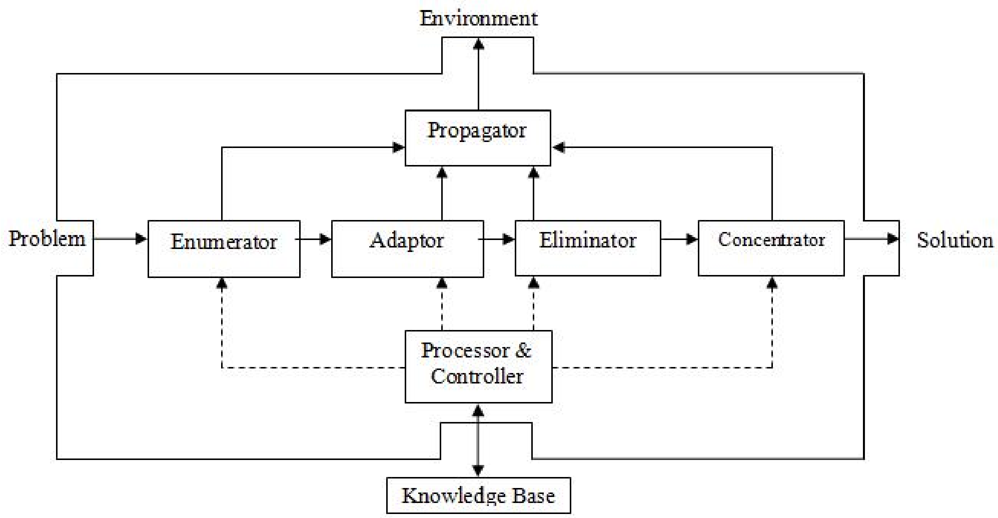


Figure 1. The advanced building block (ABB)

present the framework of slow intelligence system (SIS) [1]. SIS system is not really slow, but continuously learns, searches for solutions evolutionarily and propagates its information with environment. The traditional expert system uses a knowledge base of human expertise for problem solving and learning in expert system is explicit. However, learning in SIS system is implicit and not always obvious. In [1], the author indicated that a slow intelligence system holds six typical characteristics, including Enumeration, Propagation, Adaptation, Elimination, Concentration, two complementary decision cycles.

Enumeration: Given a problem, the system firstly enumerates different solutions until the appropriate one can be found.

Propagation: The system analyzes the environmental changes and constantly exchanges information with the environment.

Adaptation: Solutions are enumerated and adapted according to the environment. Sometimes the adapted solutions are mutations that exceed the previous ones.

Elimination: Unsuitable solutions are eliminated, so that only the appropriate ones are further considered.

Concentration: Among the suitable solutions left, one or most a few solutions are selected. Those solutions are the best ones for the problem.

Slow decision cycles to complement quick decision cycles: Slow intelligence system at least has two decision cycles: quick and slow decision cycles. The quick decision cycles provide an instant response or solution to the environment and problem while the slow decision cycles evolutionarily select and adjust the solutions by following the gradual environmental changes and analyzing the information acquired by experts and past experiences. Usually, slow decision cycles may perform more poorly in the short run but better in the long run than quick decision cycles.

Reference [1] proposed the structure of SIS by introducing of the basic building block (BBB) and advanced building block (ABB). The major difference between an ABB and a BBB is the induction of knowledge base. In this paper, we will design the feature selection system by using a knowledge base, thus we need to use the ABB structure shown in Fig. 1.

B. Feature selection

We introduce the formulation of feature selection problem and some contemporary algorithms. Let $Y = (Y_1, \ldots, Y_n)^T$ be a vector of responses and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be their associated covariate vectors where $\mathbf{x}_i = (X_1^i, X_2^i, \ldots, X_p^i)^T \in \mathbb{R}^p$. The vector pairs are assumed to be independent and identically distributed. Thus the designed matrix is denoted as $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ with size of $n \times p$ and each column/feature of X is denoted as $X_j = (X_j^1, \ldots, X_j^n)^T$.

When Y are real numbers, we assume a linear model: $Y_i = \beta_0 + \mathbf{x}_i^T \beta + \epsilon$, ($1 \le i \le n$), where ϵ is the noise following a zero-mean Gaussian distribution. Then we estimate the regression coefficient $\boldsymbol{\beta}$ and the bias β_0 by minimizing the following loss function:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} L(Y_i, \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{1}{2} \|Y_i - \beta_0 - \mathbf{x}_i^T \boldsymbol{\beta}\|_2^2 \qquad (1)$$

When $Y \in \{0,1\}^p$ for the classification problem, we assume a logistic model:

$$Pr(Y = 1|\mathbf{x}_i) = \frac{\exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta})} \quad 1 \le i \le n.$$

We estimate the regression coefficient $\boldsymbol{\beta}$ and the bias β_0 which minimizes the following loss function:

$$
\begin{aligned}
L(\boldsymbol{\beta}) &= \sum_{i=1}^{n} L(Y_i, \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) = \sum_{i=1}^{n} - \log Pr(Y_i|\mathbf{x}_i) \\
&= \sum_{i=1}^{n} \log(1 + \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta})) - Y_i(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) \quad (2)
\end{aligned}
$$

One of the most widely used feature selection methods is the L1-regularized regression [2] [3], which estimates a sparse $\boldsymbol{\beta}$ by minimizing:

$$
\min_{\beta_0, \boldsymbol{\beta}} \frac{1}{2N} \sum_{i=1}^{n} L(Y_i, \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) + \lambda ||\boldsymbol{\beta}||_1 \qquad (3)
$$

where $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{p} |\beta_j|$ is the L1-norm of $\boldsymbol{\beta}$ and λ is the regularization parameter . The above formulation gives a sparse estimator of $\boldsymbol{\beta}$, i.e. many entries in $\boldsymbol{\beta}$ will be zeros. And the important features are those with non-zero entries in the $\boldsymbol{\beta}$.

Based on L1-regularized approach, a more robust feature selection algorithm is proposed by using an elastic-net regularization [4] which minimizes:

$$
\min_{\beta_0, \boldsymbol{\beta}} \frac{1}{2N} \sum_{i=1}^{n} L(Y_i, \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) + \lambda \left(\alpha ||\boldsymbol{\beta}||_1 + \frac{1-\alpha}{2} ||\boldsymbol{\beta}||_2^2 \right) \quad (4)
$$

where α is the weight for $||\boldsymbol{\beta}||_1$. As we can see when $\alpha = 1$, it reduces to the standard L1-norm regularized approach. For both L1-norm and elastic-net regularization, a simple coordinate descent method [5] has been proposed for efficient optimization of (3) and (4).

Both (3) and (4) are based on convex optimization. For ultra-high dimensional problem with tens of thousands features, they cannot be easily applied since it leads to very slow optimization procedure. In contrast to the convex regularization method, the forward greedy pursuit approach does not suffer from such problems. Instead of trying to formulate the whole learning task to be a global convex optimization, the forward stepwise regression adopts iterative algorithm with a local view: during each iteration, it selects the best feature given a small set of selected features. Therefore, in the whole process, only a small number of variables are actually involved in the model fitting so that the whole inference only involves low dimension models. However, the main drawback of forward stepwise regression is that once an irrelevant feature is selected, it can never be removed later.

In practice, it is often hard to say which individual feature selection algorithm is better than the others. In fact, a clever synthesis of the features obtained from several algorithms may lead to the best result. But currently, there is no method can do that. This drawback motivates our work that can automatically choose the right feature selection algorithm over time and the final selected features are the results taking the advantages of multiple feature selection algorithms.

III. FEATURE SELECTION SYSTEM VIA SIS

In our feature selection process, we use the L1-regularized regression, elastic-net penalty regularized regression and forward stepwise regression as the candidate feature selection algorithms, which are stored in a knowledge base. While in principle, any existing feature algorithm can be put into the knowledge base. The whole process for our feature selection system contains one main SIS procedure and a sub SIS procedure. The main system contains five phases of slow intelligence system: enumeration, elimination, adaptation, propagation, concentration. The sub system contains two phases of SIS with the knowledge base: enumeration, concentration. The main system has a particular time controller. Next we illustrate the feature selection process step by step.

- **Main_Enumerator:** Enumerate p features $X_1, \ldots, X_p$. Among these features, some are relevant to the responses while others not.
- **Main_Eliminator:** Compute pearson correlation between each feature X_j and response Y using (5).

$$
R_j = \frac{\sum_{i=1}^{n} (X_j^i - \bar{X}_j)(Y_i - \bar{Y})}{(n-1) S_{X_j} S_Y}, \qquad (5)
$$

where $\bar{X}_j$ and S_{X_j} are the sample mean and correlation for X_j and $\bar{Y}$ and S_Y are the sample mean and correlation for Y. We rank the absolute value of R_j from high to low and eliminate the lowest $(p - C)$ features where C is a pre-defined constant. We denote the selected top C features as $\mathcal{A}_1$.

As we can see, the **Main_Eliminator** procedure enables our system to analyze the ultra-high dimensional data. For an ultra-high dimensional data where convex optimization feature selection algorithm cannot be directly applied, we first eliminate $(p - C)$ irrelevant features and keep only a small amount of features, so that the convex optimization approach can then be applied.

After that, we iterate the following process from $k = 1$ to a certain number of iterations which is defined by the time controller introduced later.

Next, we start our Sub SIS procedure that uses the feature selection algorithms in the knowledge base as the candidate methods. It automatically chooses the best method and selects the relevant feature set $\mathcal{S}_k$ from $\mathcal{A}_k$.

- **Sub_Enumerator:** Inquire the knowledge base that stores the existing feature selection algorithms, e.g., L1-regularized regression, elastic-net penalty regularized regression and forward stepwise regression, etc. We enumerate all these feature selection algorithms by applying them to our data with the feature set $\mathcal{A}_k$. Suppose for the l-th algorithm, we denote the selected top Q feature set as $\mathcal{S}_k^l$. Note that we always have $\mathcal{S}_k^l \subset \mathcal{A}_k$.
- **Sub_Concentrator:** For each selected feature set $\mathcal{S}_k^l$, we compute the loss function that is regressed on $\mathcal{S}_k^l$ as defined (1) or (2) :

$$L^l = \sum_{i=1}^{n} L(Y_i, \beta_0 + \sum_{j \in \mathcal{S}_k^l} \beta_j X_j^i),$$

and choose the best algorithm l^* with the minimum loss:

$$l^* = \mathrm{argmin}_l L^l$$

Now the Sub system selects Q features $\mathcal{S}_k^{l^*}$ from $\mathcal{A}_k$ by the best algorithm l^*. For simplicity we denote the selected feature set as $\mathcal{G}_k \equiv \mathcal{S}_k^{l^*}$ with $|\mathcal{G}_k| = Q$. Then the process goes back to main process's adaptation.

- **Main_Adaptator:** For all other features in the total p features, $X_h \in \{X_1, \ldots, X_p\} - \mathcal{G}_k$, we add each one to $\mathcal{G}_k$ and compute the loss function:

$$L_h = \sum_{i=1}^{n} L(Y_i, \beta_0 + \sum_{j \in \mathcal{G}_k} \beta_j X_j^i + \beta_h X_h^i)$$

- **Main_Concentrator:** rank all L_h with $X_h \in \{X_1, \ldots, X_p\} - \mathcal{G}_k$ from low to high, and select the top features $C - Q$ with the smallest L_h.
- **Main_Propagator:** add these features to $\mathcal{G}_k$ to form the new feature set $\mathcal{A}_{k+1}$. Note that the size of $\mathcal{A}_{k+1}$ is C.

In this entire process, the time controller controls the termination of whole process. It sets a threshold K, then checks whether the total number of cycles is equal to or larger than threshold. If yes, then it stops the process and outputs the feature selection result. That is, if$k \geq K$, it stops after sub concentration process and outputs the selected Q features; or if $k < K$, the process continues to main adaptation. The advantages of time controller design will be discussed in Section IV.

IV. SYSTEM ARCHITECTURE

In this section, we firstly show the whole algorithm of our feature selection system via SIS introduced in Section III. Then we discuss the advantages of time controller and knowledge base design issues, which characterized the SIS system uniquely.

A. General algrithm of feature selection system via SIS

The comprehensive algorithm for feature selection system vis SIS is shown below. Note that classical SIS propagates the information with an environment. In our feature selection system, the "environment" means entire features and responses in the dataset.

Sub enumeration phrase inquires knowledge base and uses those stored algorithms to compute the top Q feature set $\mathcal{S}_k^l$, where l denotes the number of different algorithms. Time controller is the decision phrase which controls the termination of whole process. Thus, knowledge base and time controller design are important issues for whole system architecture.

Algorithm: Feature selection via SIS

enumerate p features $X_1, \ldots, X_p$
compute R_j, eliminate the ($p - C$) features with lowest R_j, denote $\mathcal{A}_1$
for each k cycle **do**
  for each algorithm l in knowledge base **do**
    compute top Q features set $\mathcal{S}_k^l$ from $\mathcal{A}_k$, $\mathcal{S}_k^l \subset \mathcal{A}_k$
    compute L^l as defined in (1) or (2)
  choose the best algorithm l^* with minimum loss, set $\mathcal{G}_k \equiv \mathcal{S}_k^{l^*}$
  if $k \geq K$ **then**
    output results $\mathcal{G}_k$
  else
    for all other features in $X_h \in \{X_1, \ldots, X_p\} - \mathcal{G}_k$ **do**
      add each one to $\mathcal{G}_k$ and compute the loss function L_h
    rank L_h, select the top $C - Q$ features with the smallest L_h
    add these features to $\mathcal{G}_k$ to form new $\mathcal{A}_{k+1}$

B. Time controller design

In [1], the author pointed out that problem and solution are both functions of times in SIS system. The time controller is also a time function to control the two decision cycles. In our proposed feature selection system, we use a time controller by defining a threshold K to check the loop condition in whole process. If $K = 1$, the whole process doesn't contain main adaptation, main concentration and propagation phrases. It just outputs sub concentrator's results of the initial run. If $K > 1$, the system iteratively selects features and outputs the results after K cycles. The larger the K is, the more accurate feature selection solution is, which will be shown in the experiment. Thus, we can identify that our feature selection system possesses the main characteristic of SIS system: slow decision cycles complement quick decision cycles. When K is smaller, the quick decision cycle outputs less accurate result. When K is larger, the slow decision cycle can result in better performance. Also, the quick decision cycle's result can be used and adapted into the slow cycle to enable the system having great long-term goals. That is one of the important reasons that why our system can outperform individual state-of-the-art feature selection method.

C. Knowledge base design

In section III, we introduce a knowledge base in the sub enumeration phrase, which stores the existing feature selection algorithms e.g., L1-regularized regression, elastic-net penalty regularized regression and forward stepwise regression, etc. When we design such knowledge base, we can add any individual feature selection method which is generic or specialized candidate algorithm. Generic one means a method template with different parameters to generate different candidate algorithms. Specialized one means the particular algorithm with no parameters. For example, we note that in (4), elastic-net regularized regression reduces to the standard L1-norm regularized regression when $\alpha = 1$. It means that elastic-net regularized regression (glmnet) can be a generic candidate algorithm with parameter α. L1-norm regularized approach is one candidate elastic-net method with parameter $\alpha = 1$. In the experiment, we also use another candidate elastic-net method with parameter $\alpha = 0.5$. But forward stepwise regression is a specialized candidate algorithm with no parameter.

V. EXPERIMENT

In this section, we evaluate the performance of our feature selection system on two cancer datasets [6]. Both datasets are binary classification problems where responses $Y \in \{0,1\}^p$. We first describe the datasets and the experimental step, and then analyze the feature selection results via our system.

A. Dataset description

Firstly, we use the leukemia data set [7]. Leukemia is a type of cancer of the blood or bone marrow characterized by an abnormal increase of white blood cells. This dataset consists of 72 samples including 47 acute myeloid leukemia (ALL) ($Y_i = 1$) and 25 patients with lymphoblastic leukemia (AML) ($Y_i = 0$). For each sample, the dataset contains the expression levels of 7129 human genes (X) measured by Affymetrix high-density gene chips. The data is separated to training set with 38 samples (27 ALL and 11 AML) and testing set with 34 samples (20 ALL and 14 AML).

The second dataset we use is the colon-cancer dataset [8]. It consists of 62 samples including 40 tumor colon tissues ($Y_i = 1$) and 22 normal colon tissue ($Y_i = 0$). For each sample, the dataset contains intensities of 2,000 genes (X) analyzed with an Affymetrix oligonucleotide array. Similarly, the data is separated into 32 training cases and 30 testing cases. As we can see, both datasets are ultra-high dimensional in the sense that the number of features are hundred times of the sample size.

B. Experimental protocol

We compare our feature selection system via SIS with the three individual feature selection algorithms in our knowledge base, namely, forward regression, L1-regularized logistic regression and elastic-net regularized logistic regression with $\alpha = 0.5$. In our system, we set the constant *C*=500 in the main elimination phrase and report the results by varying the threshold K for the time controller and the number of selected features Q in the sub system. More specifically, given the confusion matrix as in Table I, where a, b, c and d represent the number of examples falling into each possible outcome, we report the number errors out of the total testing samples:

$$(b+c)/(a+b+c+d)$$

and the balanced error rate:

$$0.5\left(\frac{b}{a+b}+\frac{c}{c+d}\right)$$

The results for leukemia and colon-cancer datasets are shown in Table II and III. In addition, we list the selected genes description of leukemia data by SIS ($K = 5$) and SIS ($K = 10$) in Table IV.

C. Experimental results analysis

As we can see from Table II and Table III, for both datasets, our method greatly out-performs individual feature selection algorithm. For individual feature selection algorithms, L1-regularzied Logistic regression and glmnet achieve better classification results as compared to forward stepwise regression. It might because the forward stepwise regression is ``too'' greedy. However, it is hard to say which one is better between L1-regularized logistic regression and glmnet. Our system utilizing all these three methods is superior to any individual one. In addition, as we can see, when we increase K, the number of cycles defined by time controller, the accuracy of our system improves. In fact, using a larger K needs more time to run, but can provide us a more accurate solution. In other words, slow cycles result in better performance in long run. It is a natural tradeoff between the running time and the performance of our feature selection system.

In addition, we present the selected genes of our system by SIS ($K = 5$) and ($K = 10$). It is known from biological background [9] that these genes are critical for the leukemia disease. For example, Zyxin selected by both SIS ($K = 5$) and SIS ($K = 10$) processes LIM domain which is known to interact with leukemogenic bHLH proteins [10]. For SIS ($K = 10$), it finds very important two genes that are not discovered by SIS ($K = 5$). The first one is Cystatin C (CST3), measured by reverse transcription polymerase chain reaction (RTPCR), and confirmed that this gene is significantly increased in AML patients [11]. The other one is Cystatin A, a

TABLE I. CONFUSION MATRIX FOR THE TEST DATA WHERE $a+b+c+d$ IS THE NUMBER OF TESTING SAMPLES

Testing Data		**Predicted Class**	
		Class 0	***Class 1***
True Class	***Class 0***	a	b
	Class 1	c	d

TABLE II. PERFORMANCE OF THE LEUKEMIA DATASET: NUMBER OF ERRORS AND THE BALANCE OF ERROR RATE

Methods	**$Q=10$ features**		**$Q=20$ features**		**$Q=30$ features**	
Forward Regression	8/34	0.2857	8/34	0.2857	8/34	0.2750
L1 Logistic Regression	5/34	0.1786	2/34	0.0714	4/34	0.1429
Glmnet with $\alpha = 0.5$	6/34	0.2143	3/34	0.1071	1/34	0.0357
SIS ($K = 5$)	2/34	0.0714	2/34	0.0714	1/34	0.0357
SIS ($K = 10$)	1/34	0.0357	1/34	0.0357	1/34	0.0357

TABLE III. PERFORMANCE OF COLON CANCER DATASET: NUMBER OF ERRORS AND THE BALANCE OF ERROR RATE

Methods	**$Q=10$ features**		**$Q=20$ features**		**$Q=30$ features**	
Forward Regression	8/30	0.3175	8/30	0.3175	11/30	0.4206
L1 Logistic Regression	5/30	0.1825	8/30	0.3175	8/30	0.2857
Glmnet with $\alpha = 0.5$	5/30	0.2143	6/30	0.2381	10/30	0.4206
SIS ($K = 5$)	4/30	0.1587	8/30	0.3175	8/30	0.2857
SIS ($K = 10$)	4/30	0.1587	5/30	0.1825	7/30	0.2619

TABLE IV. THE SELECTED GENES DESCRIPTION OF LEUKEMIA DATA BY SIS ($K = 5$) AND SIS ($K = 10$).

leukaemia SIS gene selection ($K = 5$)
INDUCED MYELOID LEUKEMIA CELL DIFFERENTIATION PROTEIN MCL1
LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog
CD33 CD33 antigen (differentiation antigen)
FAH Fumarylacetoacetate
PEPTIDYL-PROLYL CIS-TRANS ISOMERASE, MITOCHONDRIAL PRECURSOR
DF D component of complement (adipsin)
Leukotriene C4 synthase (LTC4S) gene
PRG1 Proteoglycan 1, secretory granule
Zyxin
LEPR Leptin receptor
leukaemia SIS gene selection ($K = 10$)
CYSTATIN A
LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog
CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)
FAH Fumarylacetoacetate
Leukotriene C4 synthase (LTC4S) gene
RETINOBLASTOMA BINDING PROTEIN P48
Zyxin
C-myb gene extracted from Human (c-myb) gene, complete primary cds, and five complete alternatively spliced cds
ELA2 Elastatse 2, neutrophil
TCF3 Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)

natural inhibitor of cysteine proteinases, which has been suggested that an inverse correlation exists between cystatin A and malignant progression [12]. Therefore, SIS ($K = 10$) is leading to more meaningful feature selection results. We do not present the selected genes by other methods due to the limitation of the space. But all other methods miss these three important genes and the selected genes are less relevant to leukemia.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we propose a new feature selection system using the framework of slow intelligence system. It iteratively selects the best individual feature selection algorithm and propagates the learned features over iterations. It leads a superior feature selection performance and can handle high dimensional data. The experimental results show that our method greatly out-performs individual feature selection algorithm, e.g., L1-regularzied Logistic regression, glmnet and forward stepwise regression. As for the future work, we will design more sophisticate mechanism which can dynamically update the knowledge base and design the user-friendly interface to put our system in more real world applications.

REFERENCES

[1] Shi-Kuo Chang, "A General Framework for Slow Intelligence Systems", International Journal of Software Engineering and Knowledge Engineering, Vol. 20, No. 1, February 2010, 1-16.

[2] Tibshirani, R. (1996). “Regression shrinkage and selection via the lasso”. J. Royal. Statist. Soc B., Vol. 58, No. 1, pages 267-288

[3] Balaji Krishnapuram, Lawrence Carin, Mario A.T. Figueiredo, “Sparse Multinomial Logistic Regression: Fast Algorithms and Generalization Bounds”, IEEE Transaction on PAMI, VOL. 27, NO. 6, JUNE 2005

[4] H. Zou and T. Hastie. “Regularization and variable selection via the elasticnet”. J. Royal. Stat. Soc. B., 67(2):301{320, 2005.

[5] Jerome Friedman Trevor Hastie, Rob Tibshirani, “Regularized Paths for Generalized Linear Models via Coordinate Descent”. Journal of Statistical Software, 33(1), 2010

[6] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines, 2001.

[7] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. “Molecular classification of cancer: class discovery and class prediction by gene expression monitoring”. *Science*, 286(5439):531, 1999.

[8] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S.Ybarra, D.Mack, and A. J. Levine. “Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays”. *Cell Biology*, 96:6745-6750, 1999.

[9] Jianwen Fang and Jerzy W. Grzymala-Busse, “Leukemia Prediction from Gene Expression Data—A Rough Set Approach, Artificial Intelligence and Soft Computing” – ICAISC 2006 Lecture Notes in Computer Science, 2006, Volume 4029/2006, 899-908

[10] Wadman I., Li J. X., Bash R. O., Forster A., Osada H., Rabbitts T. H., Baer R, “Specific in-vivo association between the Bhlh and Lim proteins implicated in human T-cell leukemia”, EMBO Journal 13 (1994) 4831–4839.

[11] Sakhinia E., Faranghpour M., Yin J. A. L., Brady G., Hoyland J. A., Byers R. J, “Routine expression profiling of microarray gene signatures in acute leukaemia by real-time PCR of human bone marrow”. British Journal of Haematology 130 (2005) 233–248.

[12] Kuopio T., Kankaanranta A., Jalava P., Kronqvist P., Kotkansalo T., Weber E., Collan Y.: “Cysteine proteinase inhibitor cystatin A in breast cancer”. Cancer Research 58 (1998) 432–36.

[13] Shi-Kuo Chang, Yao Sun, Yingze Wang, Chia-Chun Shih and Ting-Chun Peng, “Design of Component-based Slow Intelligence Systems and Application to Social Influence Analysis”, Proceedings of 2011 International Conference on Software Engineering and Knowledge Engineering, Miami, USA, July 7-9, 2011, 9-16.

[14] Shi-Kuo Chang, Yingze Wang and Yao Sun, “Visual Specification of Component-based Slow Intelligence Systems”, Proceedings of 2011 International Conference on Software Engineering and Knowledge Engineering, Miami, USA, July 7-9, 2011, 1-8.

Word Expert Translation from German into Chinese in the Slow Intelligence Framework

Tiansi Dong and Ingo Glöckner
Department of Mathematics and Computer Science
University of Hagen
Email: {tiansi.dong|ingo.gloeckner}@fernuni-hagen.de

***Abstract*—This paper presents a novel approach to translating German sentences into Chinese using word expert translators, thereby extending the application area of the slow intelligence architecture. The word expert perspective to natural language understanding is reviewed. The motivation to word expert translation is presented – It is shown in detail that the Chinese language depends crucially on the topic-comment relation and is more suitable to be understood from the word expert perspective. Five principles for communication among Chinese word experts are proposed. Main activities of word expert translators consist of enumerating possible Chinese lexemes of a German lexeme, determining linear ordering among Chinese lexemes within one word expert translator, determining topic-comment relations among word experts, constructing nested topic-comment relations among word expert translators, choosing the possible Chinese lexemes. All these fit well within the slow intelligence framework.**

I. Introduction

Machine translation (MT) is to transform texts from one natural language into another by computers. The worldwide market for machine translation is large and steadily increasing with a growth rate of around 20% per year, in part because the manual translation by human translators is expensive and slow. Main technologies for MT are the interlingua method, the statistical method, and the rule-based approach. The rule-based method assumes that languages can be governed by rules. However, every rule has exceptions; and the mapping between rules of different languages can be complicated, sometimes even impossible. The statistical method [1] focuses on the probability of translating a sentence in the source language into a sentence in the target language, therefore, it requires a highly qualified and extensive sample translation corpus. Even if such sample corpus is available, what the statistical method guarantees is only a probability. The interlingua approach ideally assumes a common semantic representation for all natural languages. Translation is a process of transforming text in the source language into the interlingua, and a process of transforming the interlingua into the target language. The difficulty is that such interlingua is hard to develop, at least at the time when the ALPAC report was made. Recent research on the semantic representation of natural language has made fruitful progress. It is time now to reconsider MT with the interlingua approach, e.g. [8]. The present paper adopts the MultiNet representation [5] as the interlingua for the semantic representation of natural languages, and applies the word expert perspective for machine translation.

The rest of the paper is structured as follows: Section 2 reviews the word expert perspective, and the successful application of the WOCADI parser developed by [6]; Section 3 presents the topic-comment structure of the Chinese language, and motivates the method of word expert translator; Section 4 presents the main activities of word expert translators and shows how these activities are carried out in the slow intelligence framework.

II. Word Expert Perspective

A. The Perspective

The traditional perspective views words as passive data (with knowledge of part of speech and meaning), and languages as an infinite set of sequences of words (satisfying grammatical rules). The word expert perspective pioneered by Rieger [9] views individual words as active procedures – *Each word of language is seen as an active lexical agent called a word expert, which participates in the overall control of the parsing process by its internal actions and its interactions with other such agents* [10, p.1]. The word expert view advocates the integrated syntax-semantics coupling approach to language understanding. The traditional syntax is viewed as an artifact describing patterns of lexical interactions, and cannot be used to model comprehension due to the rich semantic particularities of lexemes [10, p.3]. The word expert perspective views text understanding as a process of interactions among word exports that results in a disambiguation. The work of parsing is designed as the decision-making process of each word expert to choose one suitable interpretation for the word that it represents. Comprehension is therefore simulated as an activity of looking for the best possible fit among word experts. This view not only differs from the traditional rule-based approach (in which words are passive), but also from the statistical-based approach (in which *best fit* is guessed by looking back into the past existing sample corpus).

B. Word Expert Parser for English

Following Wilks' parsing system [13], Small [10] developed one of the most influential word expert parsers for English. His theoretical position is that *words have no meaning per se, but rather that fragments of lexical items mean something through their interrelationships* [11, pp.70]. That is, each lexical item is viewed as having certain interactions with its neighborhood items, and produces meanings.

C. Word Class Expert Parser for German

Based on the ideas of Word Class Functions [7], Helbig and Hartrumpf [6] developed the first semantically oriented word class expert parser for the German language – WOCADI. In contrast to Small's distributed-interaction approach, WCFA *describes the grammatical functions of whole classes of words* [6, pp.313]. The parser transforms German sentences into the MultiNet formalism [5]. It has been tested with all of the texts in the German Wikipedia. The MultiNet formalism and the WOCADI parser have been successfully applied in the LogAnswer question answering (QA) system which scored second among non-English QA systems in the CLEF competition [4].

D. Word Experts for Translation

As pointed out by Small, the word expert perspective suggests a new way to look at translation [10, p.15]. The generation step, in particular, requires word experts to arrange themselves into a meaningful sequence by communicating with each other.[1] For example, let the three word experts represent *drive*, *Joanie* and *car*. The *drive* expert would send out the message: in front of me there shall be someone, behind me there shall be a vehicle, who fits? The *Joanie* expert would reply: I can stay in front of you; the *car* expert would reply: I can stay behind you. A sequence of *Joanie drives car* is therefore formed.

Translation from the word expert perspective consists of two processes: one process is word expert parsing from the source language into a meaning representation, and the second process is word expert generation from the meaning representation into the target language. We focus on the translation from German into Chinese. The first process is carried out by the WOCADI parser. The second process transforms the MultiNet semantics representation [5] into Chinese.

III. The Chinese Language and Chinese Word Experts

The grammar for Chinese is totally different from the grammars of German or English: Most parts of speech can serve as both the subject and the predicate in a Chinese sentence. Therefore, an attempt to map grammar rules between German and Chinese only makes things complicated. We will show that the word expert perspective is a very suitable way to explain the Chinese language.

A. Subject and Predicate as Topic and Comment

Chao [3] studied the Chinese language and concluded that the relation between the subject and the predicate in a Chinese sentence is a topic-comment relation. This relation holds in all Chinese dialects [12], e.g., Mandarin (used in Peking area), Wu (used around Shanghai area), Cantonese (used in and around Canton area), and WenYan (used in ancient China). For example, in 约翰死了(约翰/John, 死了/dead), 约翰/John is the topic, 死了/dead is the comment, which presents a comment on 约翰, which means *he is dead*. In 约翰死了父亲(约翰/John, 死了/dead, 父亲/father), 约翰/John is the topic, 死了父亲(死了/dead, 父亲/father) is the comment, which presents a comment on 约翰, which means *his father is dead*. In 水开了(水/water,开了/boil), 水/water is the topic, 开了/boil comments on the water; in 书看完了(书/book, 看/read, 完了/finish), 书(book) is the topic, 看完了(看/read, 完了/finish) is a comment – (I) have finished reading the book.

The topic-comment relation introduces a question-answer relation between the subject and the predicate [3, p.81]. Imagine a man returns home after work, and asks his wife, 饭呢? (*where (is the) rice?*) his wife answers, 都吃完了(*all eaten*). The man introduces a topic 饭/rice in a question; his wife comments on the topic by answering *all eaten*.

B. Full Chinese Sentences

A full Chinese sentence has a topic and a comment.

1) nominal expressions as comments: In 他美国人(he is an American), 他/he is a pronoun, 美国人/American is a noun. There is no link verb 是/is between 他/he and 美国人/American. The <u>Chinese Word Expert 他</u> (shortened as 'CWE他') asks its surrounding experts, for example, who can be my property? The CWE美国人 answers: I can be your property of nationality. If we view Chinese words as such active agents, instead of passive data as in the traditional rule-based grammar, the link verb 是/be is not necessary. In 屋里许多蚊子(屋里/inside of the room, 许多/many, 蚊子/mosquito), the nominal expression 许多蚊子 is the predicate. The CWE屋里 asks: what is inside of the room? The CWE许多蚊子 answers: many mosquitoes. The meaning of the sentence is *there are many mosquitoes inside of the room*.

2) active verbs as comments: In 这件事早发表了(this matter has long been published), 这件事(this matter) is the topic, 发表/publish is a verb in the active form. In Chinese the passive form 被发表(be published) may not be used to mark the passive action – a construction taken for granted in English or German. However, from the word expert perspective, Chinese has more pragmatic efficiency in expressing meanings: the CWE发表 asks: what can be published? the CWE这件事 answers: this matter. So, in Chinese, the passive form is indeed not necessary. In 酒不喝,烟抽(wine, (I) do not drink, tobacco, (I) smoke), 酒/wine and 烟/tobacco are subjects, 喝/drink and 抽/smoke are verbs in the active form. The CWE不喝 asks: what not to drink? the CWE酒 answers: wine. The CWE抽 asks: what to smoke? the CWE烟 answers: tobacco.

3) adjective as comments: In 我穷(我/I, 穷/poor), the whole comment is one adjective word 穷. The meaning of the sentence is *I am poor*. The CWE我 asks: how about me? The CWE穷 answers: poor. In 菜咸(菜/dish, 咸/salty), the whole comment is one adjective word 咸. The CWE菜 asks: who can serve my property? The CWE咸 answers: salty. The meaning of the sentence is *the dish is salty*.

4) full sentences as comments: In 这个人耳朵软(这个/this, 人/man, 耳朵/ear, 软/soft), the comment is a full sentence 耳朵软. The direct translation is *(as for) this man, the ear is soft*. The CWE这个人 and the CWE耳朵 ask: what is my property?

[1] For simplicity, we adopt Small's individual word experts view here though an abstraction of word class experts would again make sense.

The CWE软 answers to both: soft. As the CWE软 is nearer to the CWE耳朵 than to the CWE这个人, its answer was first accepted by the CWE耳朵. As a result a new word expert CWE耳朵软 is formed which means gullible. This new expert answers the question raised by CWE这个人. The meaning of the sentence is *this man is gullible*. From this example we propose two principles for communications among Chinese word experts as follows:

Chinese WE Principle 1: Neighborhood word experts have the priority in communication.

Chinese WE Principle 2: New word experts may appear after successful communications and play roles in communication with other experts.

In 这个人脑筋简单(这个/this, 人/man, 脑筋/mind, 简单/simple), the comment is a full sentence 脑筋简单. The direct translation is *this man (is such that his) mind is simple*. The meaning of the sentence is *the mind of this man is simple*. With the two principles, the CWE简单 first communicates with the CWE脑筋, and forms a new word expert CWE脑筋简单, which answers the question of the CWE这个人. The whole process can be simulated by two question-answer rounds: -how about the mind? -simple -how about this man? -mind is simple.

5) verbal expressions as topics: In 走行,不走也行(走/go, 行/all right, 不/not, 也/also), verbal expressions 走 and 不走 are topics. The meaning of the sentence is *to go is all right, not to go is also all right*. Primitive verbs as subjects are neither allowed in English nor in German, but very normal in Chinese, and can be easily explained from the word expert perspective. The CWE走 asks: shall I perform? The CWE行 answers: all right. The CWE不走 asks: can I not perform? The CWE行 answers: all right.

6) Spatial-temporal expressions as topics: In 今儿冷(今儿/today,冷/cold), the temporal expression 今儿is the topic. The meaning of the sentence is *today is cold*. The CWE今儿 asks: how is today? The CWE冷 answers: cold. In 这儿是哪儿? (这儿/here,是/is,哪儿/where), the spatial expression 这儿is the whole topic. The meaning of the sentence is *where is here?* The CWE这儿 asks: where is here? The CWE哪儿 will communicate with word experts in the next sentence for an answer.

7) conditional expressions as topics: In 他死了的话简直不堪设想了(他/he,死了/dead,的话/if,简直/simply,不堪设想了/unthinkable), the topic is the conditional expression 他死了的话. The meaning of the sentence is *the supposition that he should die is simply unthinkable*. The CWE他 asks: how is he? The CWE死了 answers: dead. The CWE的话 asks: what will be the result under what condition? The CWE他死了 serves the condition, the CWE不堪设想 serves the result.

8) prepositional expressions as topics: In 由主席召集会议(由/through,主席/chairman,召集/convene,会议/meeting), the topic is a prepositional expression 由主席. The meaning of the sentence is *the meeting is convened through the chairman*. The CWE由 asks: through what? by whom? The CWE主席 answers: chairman. A new word expert CWE由主席 is formed. The CWE召集 asks: who convenes what? how to convene? The CWE由主席 answers the how question, the CWE主席 answers the who question, and the CWE会议 answers the what question.

9) full sentences as topics and comments: In 他死了我真难受(他/he, 死了/dead, 我/I, 真/awfully, 难受/feel bad), the topic is a full sentence 他死了(he is dead), the comment is also a full sentence 我真难受(I feel awfully bad). The direct translation is *he is dead, I feel awfully bad*. The meaning of the sentence is *that he is dead is something about which I feel awfully bad*. The CWE他 asks: how is he? the CWE死了 answers: dead. The CWE他死了 asks: how? what is the result? The CWE我 asks: how about myself? The CWE真难受 answers: awfully bad. The CWE我真难受 answer the question: what is the result? If the sentence is 他死了, 车祸(他/he, 死了/dead, 车祸/traffic accident), the CWE车祸 will answer the question of the CWE他死了: how come?

We conclude that structures of full Chinese sentences violate many important grammar rules of Western languages, and that understanding full Chinese sentences can be achieved by communications among Chinese word experts following some communication principles. The meaning of a full Chinese sentence can be represented by a dialog process among the word experts.

C. Minor Sentences

In conversations, it is normal that one speaker introduces a topic and the other makes a comment. This introduces the term of *minor sentence* [3, p.60]. In contrast to *full sentence*, a *minor sentence* does not have both topics and comments. The above conclusion is supported by the structure of Chinese minor sentences in that a speaker in a conversation may only say a few words to answer the questions raised by word experts of the other speaker.

D. Compound and Complex Sentences

By parallelizing two or more sentences, we can construct compound Chinese sentences. For example, 你不认得我, 我不认得你(you do not know me and I do not know you) is a compound sentence by parallelizing two ‘A 不认得 B’ (*A does not know B*) sentences. By nesting one full sentence either in topic or in comment, we can construct complex sentences. For example, 我死了丧事从简(我/I,死了/dead,丧事/funeral,从简/simple)is a complex sentence, meaning *when I die, the funeral should be simple*, constructed by using a full sentence 我死了(*I die*) as the topic. Both of the two structures can be easily explained in the word expert framework with the second communication principle (*Chinese WE Communication 2*) and the third communication principle as follows.

Chinese WE Principle 3: A newly formed word expert has priority over old ones in communication.

With this principle, although both CWE死了 and CWE我死了 answer the CWE丧事从简’s conditional question, CWE我死了 is newly formed and has the priority.

E. Pivotal Constructions

A Chinese sentence with a pivotal construction has normally two verbs and a nominal expression. This nominal expression serves as the object of the first verb and as the subject of the second verb [3, p.124-125]. For example, in 我们派他做代表(我们/we, 派/order, 他/he, 做/serve, 代表/representative), there are two verbs 派/send and 做/serve, a pronoun他/he is the object of 派/send and the subject of 做/serve (in Chinese 他/he has the same form as nominal case and as dative case). This direct translation of the sentence is that 'we order he serve as representative'. The meaning is that *we delegate him to be representative*. Within the word expert perspective, we can explain this special construction without introducing any new terminologies (as done in the rule-based grammar theories). CWE派 asks: whom is sent? CWE做 asks: who serves? CWE他 answers to both: 他/he.

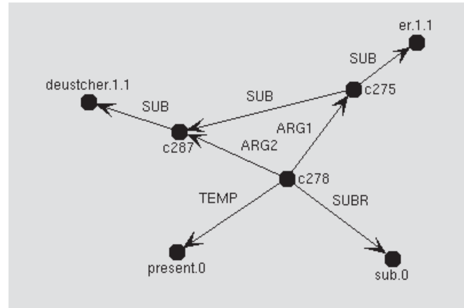


Fig. 1. MultiNet representation of the sentence *Er ist ein Deutscher* (he is a German)

IV. German-Chinese Word Expert Translators

Given a sentence in German, the WOCADI parser can be used to generate a corresponding semantic representation in the MultiNet formalism. We need to design word expert translators that communicate with each other, linearize themselves to form a (nested) topic-comment structure, and transform the MultiNet representation into Chinese sentences.

We start with a simple example to introduce the main idea. Suppose the German sentence is *Er ist ein Deutscher* (He is a German). The WOCADI parser delivers the MulitNet semantic representation as illustrated in Figure 1: `er.1.1` is the word sense[2] of the lexeme *er/he*; `c275` represents the word expert for a concrete individual[3] subordinate (`SUB`) to `er.1.1`; similarly, `c287` represents the word expert whose concept is subordinate to `deutsche.1.1`/*German*. The word expert `c278` has two arguments: one is the topic, `c275`, pointed by `ARG1`; the second is the comment, `c287`, pointed by `ARG2`; the temporal status (`TEMP`) of `c278` is present (`present.0`).

The word expert `c275` posts a message: if I am the topic, who can be my comment? The word expert `c278` answers: as far as I know, the word expert `c287` is your comment. If word expert `c275` knows that his Chinese lexeme is 他/he, and word expert `c287` knows that his Chinese lexeme is 德国人/German, they will know that in the linearization of the Chinese sentence `c275` is before `c287`. The simplest case is: 他德国人(He German), which is indeed a valid Chinese sentence with the same meaning as *Er ist ein Deutscher*. If word expert `c278` knows that his Chinese lexeme is 是/is, and knows that I shall stay between the topic and the comment in the Chinese sentence, the Chinese sentence will be 他是德国人(he is German). If `c278` knows that 现在/now is the Chinese lexeme meaning *present* and decides that his temporal knowledge shall also be encoded in the Chinese sentence, the Chinese sentence could be 现在他是德国人, 他现在是德国人, 他是现在德国人, 他是德国人现在. If he knows that his temporal label shall be in front of him, the two linearizations will be 现在他是德国人 and 他现在是德国人, both are valid Chinese sentences. These word experts are now also experts for translating. Activities of each word expert translator comprise: to transform its German lexeme into possible Chinese lexemes, to determine a linear ordering relation among Chinese lexemes of a single word expert translator, to determine whether its Chinese lexemes shall appear in the Chinese translation, to establish a linear ordering relation with other word expert translators by communication, and to choose most appropriate Chinese lexemes.

[2]Our lexicon uses a double indexing scheme to distinguish word senses
[3]Entities mentioned in the text are represented by constants c1, c2 etc.

A. Transforming into Chinese Lexemes

This task is to find Chinese lexemes for a given German lexeme such that they represent the same concept. This may not be always feasible. Some German lexemes may not have a corresponding native Chinese lexeme, e.g. names of Cheeses, beers, and chocolates – in German, *Franziskaner* can refer to a kind of beer, and there is no corresponding native Chinese lexeme. For those having corresponding native Chinese lexemes, these may be different in Chinese dialects. For example, German lexeme *wir* (*we*) can be mapped to 我们 in Mandarin, 阿拉 in Shanghai dialect, 我哋 in Canton dialect. Chao [3] suggest that a complete lexicon shall be constructed *to make selection applicable in grammar*. For the translation from German into Chinese, we need to embed the Chinese lexical ontologies system into the German lexical ontologies system, and mark each Chinese lexeme with its dialect group, while neglecting fine Chinese ontologies lexemes which do not have corresponding German lexemes. German lexemes which do not have native Chinese lexemes will be translated separately.

B. Lexeme Ordering within a Word Expert Translator

A word expert in the analysis of a German sentence may have a temporal property representing the tense. In Chinese tense is achieved by particles: 了、过、着, such as 他吃过饭了 (er hat gegessen/he has eaten). The word expert essen.1.1/eat has a temporal property `past.0`, which shall be translated into the Chinese particle 过 and 了. The corresponding Chinese lexemes of word expert essen.1.1 are 吃、过、了, whose linear ordering in Chinese sentence is stated as follows.

Chinese WE Principle 4: Let L be a word expert which may have 过 or 了as particle. (1) if 过 is the only particle, it must occur directly after L; (2) if 了 is the only particle, it

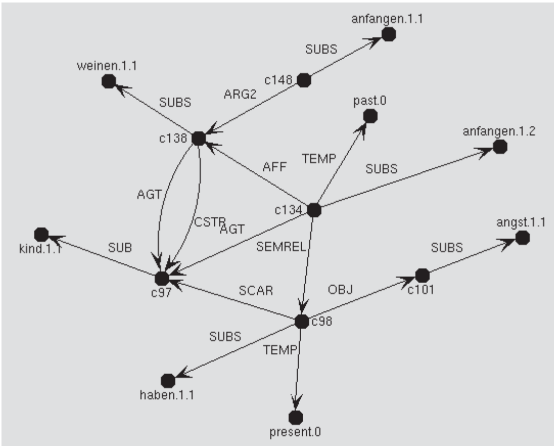


Fig. 2. MultiNet representation of the sentence *Das Kind hat Angst und fängt an zu weinen* (The Child is scared and begins to cry)

shall be after L; (3) if both 过 and 了are particles, 过 shall be before 了besides obeying rules (1) and (2); (4) if 了 is used twice, besides obeying rule (2), they must be separated by lexemes of another word expert, and one must directly follow L.

With the above principle, the following orderings are all understandable translations: 吃过饭– (1), 吃饭了– (2), 吃了饭– (2), 吃过饭了– (1)(2)(3), 吃过了饭– (1)(2)(3), 吃了饭了– (2)(4).

C. Determining to Appear or Not

Not all the word experts of a German sentence shall appear in the translated Chinese sentence. For example, the German sentence *Das Kind hat Angst und fängt an zu weinen* (*The Child is scared and begins to cry*) shall be translated into Chinese 孩子吓哭了(孩子/Kind/Child, 吓/Angst/scare, 哭/weinen/cry, 了/particle), as illustrated in Figure 2. The following word experts appear in the translated Chinese sentence: kind.1.1/child (`c97`), weinen.1.1/cry (`c138`), angst.1.1/scare (`c101`), and haben.1.1/have (`c98`). The carrier (`SCAR`) of haben.1.1/have is kind.1.1/child, which is the actor (`AGT`) of both weinen.1.1/cry and anfangen.1.2/begin. The affected object (`AFF`) of anfangen.1.2/begin is weinen.1.1/cry, which is the second argument (`ARG2`) of anfangen.1.1.[4]

Considering the topic-comment relations among the word experts, we find that the comment of angst.1.1/scare is weinen.1.1/cry, and that the two word experts anfangen.1.2/begin and anfangen.1.1/begin simply duplicate this topic-comment relation. Therefore, they shall not appear in the translated Chinese sentence – the redundancy is only an artifact of our deep semantic analysis. A principle is stated as follows.

Chinese WE Principle 5: A word expert shall not appear in the translated Chinese sentence, if it duplicates an existing topic-comment relation.

[4]This analysis reflects that if someone starts something (anfangen.1.1), then this starting actions causes something to start (anfangen.1.2).

D. Ordering of Word Expert Translators

For those word experts that will appear in the translated Chinese sentence, a nested topic-comment ordering shall be constructed. As we have semantic representations from WOCADI and lexemes tagged with semantic roles, topic-comment relations among word experts are not difficult to obtain – we only need to examine the description of the lexeme and the possible MultiNet constructions, to see which elements can be the comment of which other elements in the meaning representation. In Figure 2, angst.1.1/scare is the topic for weinen.1.1/cry, so we have a list (WE-angst.1.1/scare WE-weinen.1.1/cry), kind.1.1/child is the topic for angst.1.1/scare, so we have (WE-kind.1.1 (WE-angst.1.1 WE-weinen.1.)). By flattening this nested structure, we have a linear ordering of the word experts: (WE-kind.1.1 WE-angst.1.1 WE-weinen.1.1). Chinese sentences can be obtained by replacing each word expert with their Chinese lexeme(s). For example, 孩子吓<u>了</u>哭, 孩子吓哭<u>了</u>, and 孩子吓<u>了</u>哭<u>了</u> are all understandable translations, where 了 is the particle of 吓, whose ordering follows *Chinese WE Principle 4*.

E. Choosing Chinese Lexemes

A word expert may have more than one corresponding Chinese lexeme even in one Chinese dialect. Communications among word experts are required to select the most suitable one, or to delete incompatible ones. For example, the word expert ein.1.1/a in the German phrase *ein Baum* (*a tree*) can be mapped to 一只、一头、一张、一顿. . . . By communicating with the word expert Baum.1.1/tree, word expert ein.1.1/a knows that it can only be mapped to 一棵. This requires word experts of countable objects to have measurement information.

F. A Slow Intelligence (SIS) Workflow

The word expert translation system can be organized in the slow intelligence architecture [2]. Each word expert that occurs in the German parse is a unit slow intelligence system. It enumerates possible Chinese lexemes, determines possible linear orderings among themselves, and the topic-comment relation of its arguments. Word experts communicate among themselves and form a larger slow intelligence system: nested topic-comment structures are firstly enumerated, then duplicated topic-comment relation will be removed, incorrect lexemes will be pruned, ordering of particles will be linearized with lexemes of other word experts. At last possible Chinese sentences will be produced.

For the German sentence *Ich fällte einen Baum mit einer Axt* (*I cut down a tree with an ax*), the MultiNet analysis results in four word experts: WE-ich.1.1/I (abbreviated WE-I in the following), WE-baum.1.1/tree (WE-tree), WE-axt.1.1/ax (WE-ax), and WE-fällen.1.1/cut-down (WE-cut-down). The system first creates one CWE for each word expert WE. All the CWEs are unit SIS, searching for possible Chinese lexemes and enumerating linear orders based on CWE principles. In our current German-Chinese dictionary CWE-I has one Chinese lexeme: 我; CWE-cut-down has five Chinese lexemes: 伐、做出、砍、砍伐、采伐 and two particles 过 and 了;

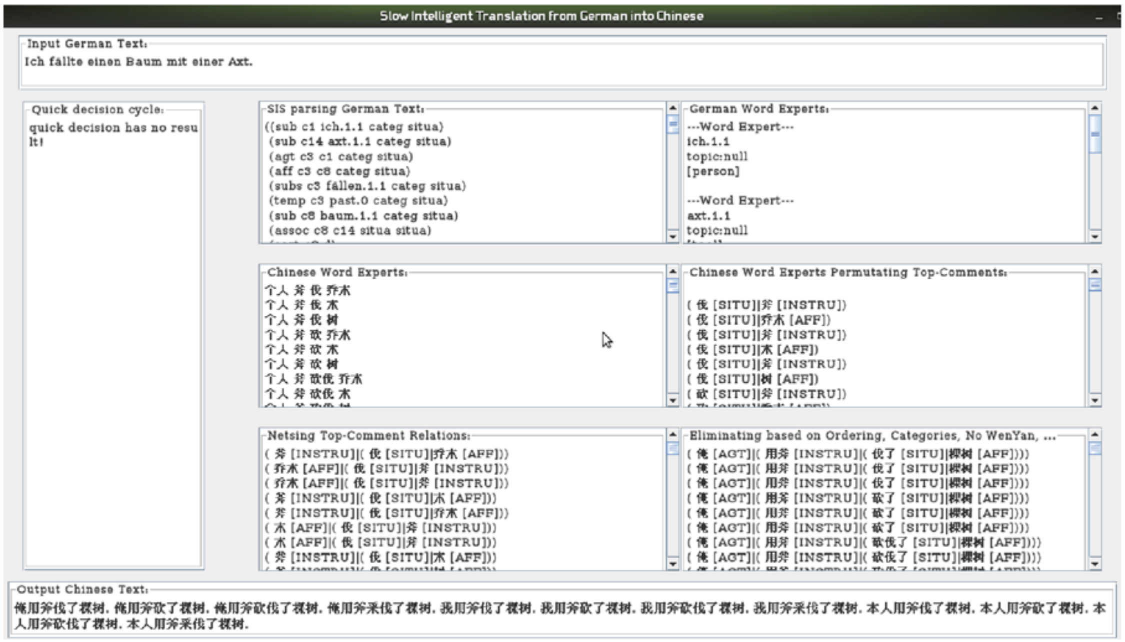


Fig. 3. Word expert translation results for the input sentence *Ich fällte einen Baum mit einer Axt*

CWE-tree has one Chinese lexeme: 一棵树; CWE-ax has one Chinese lexeme: 斧. Only CWE-cut-down need to carry out an enumeration-elimination process to linearize orders of each Chinese lexeme and particles. The orders among the five CWEs will be enumerated based on the topic-comment relations as follows: (1) CWE-I is the actor of CWE-cut-down, therefore, CWE-I is the topic, CWE-cut-down is the comment; (2) CWE-tree is the affected object of CWE-cut-down, therefore, CWE-cut-down is the topic and CWE-tree is the comment; (3) CWE-ax is analysed as the association of CWE-tree, whose features delivered from WOCADI parser are different: CWE-ax is an instrument, while CWE-tree is a non-instrument. Therefore, CWE-ax is the instrument commented both on the action and on the actor. This introduces the particle 用/using before the lexeme of CWE-ax: 用斧(using an ax). Two nested topic-comment structures are (CWE-I ((CWE-cut-down CWE-tree) CWE-ax) and (CWE-I (CWE-ax (CWE-cut-down CWE-tree)). In the adaptation phase CWEs will communicate with each other to prune lexemes. 做出 and 采伐 of CWE-cut-down are removed, as they incompatible with the instrument CWE-ax. A demo translation system has been developed. Its translation results are illustrated in Figure 3.

V. Conclusions

The work reported in this paper is mainly based on a classic research work on Chinese grammar [3]. We argue that the Chinese language structure is more suitable to be understood from the word expert perspective than from the traditional rule-based perspective. The machine translation from German into Chinese is outlined from the word expert perspective. A slow intelligence workflow for machine translation is illustrated with an example.

Acknowledgments

Hermann Helbig introduced us the topic and critically commented on the draft paper; Peiling Cui criticized the Chinese translations. Financial support from DFG is gratefully acknowledged.

References

[1] P. F. Brown and S. A. Della Pietra J. Cocke. A Statistical Approach to Machine Translation. *Computational Linguistics*, 16(2):79–85, 1990.

[2] S. K. Chang. A general framework for slow intelligence systems. *International Journal of Software Engineering and Knowledge Engineering*, 20(1):1–18, 2010.

[3] Y. R. Chao. *A Grammar of Spoken Chinese*. University of California Press, 1968.

[4] I. Glöckner and B. Pelzer. The LogAnswer project at ResPubliQA 2010. In *CLEF 2010 Working Notes*, September 2010.

[5] H. Helbig. *Knowledge Knowledge Representation and the Semantics of Natural Language*. Springer-Verlag, 2006.

[6] H. Helbig and S. Hartrumpf. Word class functions for syntactic-semantic analysis. In *Proceedings of the 2nd International Conference on Recent Advances in Natural Language Processing*, pages 312–317. Tzigov Chark, Bulgaria, 1997.

[7] H. Helig. Syntactic-semantic analysis of natural language by a new word-class controlled functional analysis. *Computers and Artificial Intelligence*, 5(1):53–59, 1986.

[8] V. Mihailevschi. *Machine Translation Interlingua Based on MultiNet*. VDM Verlag Dr. Müller, 2008.

[9] C. Rieger. Viewing Parsing as Word Sense Discrimination. In Dingwall, editor, *A Survey of Linguistic Science*. Greylock Publishers, 1977.

[10] S. Small. *Word Expert Parsing: A Theory of Distributed Word-Based Natural Language Understanding*. PhD thesis, Department of Computer Science, University of Maryland, 1980.

[11] S. Small. Viewing Word Expert Parsing as Linguistic Theory. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 70–76. Erlbaum, Hillsdale, 1981.

[12] L. Wang. *History of Chinese Language*. Zhonghua Book Co. Press, 1980. 王力, 漢語史稿, 中華書局出版.

[13] Y. Wilks. Making Preferences More Active. AI Memo 206, Artificial Intelligence Laboratory, Stanford University, 1973.

Beyond Macroblocks in Lossy Video Compression

Claudio Cappelli[1]

[1] Eco Controllo SpA
Centro Direzionale, Isola E5, 80143 Napoli (NA), Italy
claudio.cappelli@ecocontrollo.com

Abstract

In this paper we describe the performances of a new proprietary lossy video compression technique with respect to those of main competitors. Since our codec is covered by industrial secret, we only sketch the basic principles of the compression algorithms underlying it, while discussing the key advantages with respect to macroblocks based techniques, on which most current codecs rely.

1. Introduction

Recent years have witnessed a remarkable growth of digital image processing applications. This is mainly due to several factors, including the development of Internet, digital TV, the increasing prevalence of mobile devices with high multimedia capabilities, and so forth. Although seemingly different, these application domains have a common denominator: the need to transmit still and / or moving images through different types of network infrastructures, wired or wireless that, given the average size of objects to be transmitted, has yielded an exponential growth of bandwidth needs.

The problem can be faced by first acting on the increase in bandwidth and on the other hand on the compression of multimedia objects to be transmitted. In the first case, however, we often clash with budget constraints, long realization time (reduced in part by the adoption of wireless network infrastructures), as well as maintenance problems, which have often led to delays in development or simply in the worldwide diffusion of certain applications. Thus, although the development of network links is a key aspect for the success of these multimedia applications, the pursuit of satisfactory targets with adequate times and costs is undoubtedly linked to the development of efficient compression techniques.

In the last decade, many video codecs have been created. Among these, *MPEG-4* and *H.264* have received particular attention [10], and many applications have exploited them. More recently, particular interest is devoted to VP8, a video compression format owned by Google and originally created by On2 Technologies.

Codecs have played a central role in competitions among big commercial brands (think about recent contests between Google and Apple, as well as Microsoft commercial strategies to include Windows Media Player within its operating systems), all focusing on the promotion of proprietary codecs as a means to build customer loyalty and to gain market shares. This yields a trend of these brands to highlight slight technological differences with respect to their competitors, claiming the exclusivity of their products to customers, and focusing on innovation features that should be properly advertised.

All these codecs rely on the concept of macroblock, a square portion of a video frame whose size depends on the codec and is usually a multiple of 4, and in modern codecs is about 16x16 pixels. Usually, codecs analyze a macroblock trying to guess the macroblock in the next frame that best approximates it [10]. This might have different coordinates due to the dynamics within the scene. If such an approximate search succeeds, the codec might avoid to fully encode the macroblock twice for the two frames. Rather it might store it once for the first video frame, together with a motion vector to reconstruct the position that the macroblock has on the next frame, plus the pixels that have changed. For instance, *MPEG-4* might fully encode a macroblock, or just store the information enabling its reconstruction from a macroblock in a previous video frame, or in combination from a previous and a following video frame. This mechanism is particularly sophisticated. For some video contents it might be difficult encoding macroblocks in terms of motion vectors and differences with respect to other macroblocks, which might entail fully encoding of video frames, yielding a tremendous increase in the required bit rate.

In this paper we describe the basic principles and performances of a new proprietary codec. The codec has been produced within a research project at Eco Controllo s.p.a., and it is based on a completely new idea. In particular, it does not rely on macroblocks, rather it exploits the grayscale component of images to partition them and to guide the prediction of basic objects in a scene and their motion. In the past, it has already been experimentally shown that this prediction method yields more efficient video coding, allowing to reach a better tradeoff between *bit rate* and quality of the compressed images [1]. Due to industrial secret, in this paper we only provide some hints on the ideas underlying our compression algorithms. However, we provide further experimental evidence of their potential advantages. In particular, we describe results of recent tests we entrusted to external research institutions, in which the latest version of our codec has been compared to VP8 [3]. The choice of VP8 has been motivated by the increased interest on this codec after Google purchasing its commercial rights from On2 Technologies. In such tests our codec outperformed VP8 both in terms of objective video quality metrics applied to compressed videos and in terms of compression time.

The paper is organized as follows. Section 2 provides a sketch description of the ideas underlying our compression algorithms. In Section 3 we provide test results. Finally, Discussion is reported in Section 5.

2. Principles of the Eco Controllo's codec

Although Eco Controllo's codec is covered by industrial secret [2], in this section we try to give an idea of its strength points.

Eco Controllo's compression algorithms gain main advantages by using alternative techniques with respect to macroblocks. This is because macroblocks tend to partition images in a way that is independent of their contents, yielding an artificial partitioning method that might be more or less suitable depending on video contents, possibly making motion prediction a hard process. Conversely, Eco Controllo's video compression technique exploits the grayscale component of images to partition them and to guide the prediction of basic objects and their motion. The idea stems from the experimental observation that the colors in a picture tend to be distributed in a layered fashion, starting from the basic grayscale component up to fine color tones.

Also MPEG somehow takes into consideration the layered nature of the colors in an image, although it exploits this characteristic to derive a better sampling of the image, processing images represented in YUV rather than in the RGB color model [10].

Like MPEG, also Eco Controllo's video codec relies on YUV color model. In particular, it exploits the separation of the luminance component (Y) from the two chrominance components (UV). The former alone is capable of producing grayscale images, whereas the second one measures the color values to be added to the gray pixels in order to derive a colored image. Thus, while it will always be possible to express a picture by only employing the Y component, achieving a grayscale picture, the converse will not be true. In other words, the U and V values only make sense if they are associated to some luminance value expressed by the Y component. This means that the 8 most significant bits of the YUV representation are the ones providing the skeleton of the picture (Figure 1(a)), whereas the other two less significant bytes, corresponding to the U and V components, respectively, add color details (Figure 1(b)).

(a) Grayscale skeleton of an image.

(b) True color representation of image (a).

Figure 1. Grayscale and true color pictures in the YUV model.

Proceedings
International Conference on
Distributed Multimedia Systems
(DMS 2011)

Co-Editors

G. Costagliola, *University of Salerno, Italy*
Jennifer Leopold, *Missouri University of Science and Technology, USA*

3. Experimental Comparison

Codecs are continuously compared to test their capabilities to preserve proper quality of video sequences while reaching high compression ratios. Most modern codecs have been experimentally compared at the CS MSU GRAPHICS & MEDIA LAB VIDEO GROUP [12].

In this section we describe the results of comparative tests that Eco Controllo has committed to external research organizations, in order to compare their codec with respect to VP8, the codec used on YouTube.

Codecs might be tested by means of both subjective or objective metrics [17]. The latter try to simulate the visual quality as it is perceived by humans [4][5][6][10][11][13][14][15][16], whereas in the formers the evaluation is accomplished by a group of individuals, according to a given methodology [9].

The results we describe in this paper have been measured by comparing Eco Controllo's codec and VP8 through the Peak Signal-to-Noise Ratio (*PSNR*) and the Structural Similarity Index (*SSIM*) objective metrics [17]. In particular, Eco Controllo requested external research organizations to estimate the perceived "fidelity" of the compressed video (by means of the different codecs being compared) with respect to the original video, given a fixed value of bit rate. Moreover, it has been requested to repeat such tests for different frame rates (25 and 50 Hertz), and different values of bit rate, with respect to which compress each selected video, by using each of the codecs being compared.

Test video sequences were selected by choosing typical high-definition sequences, many of which had been previously used in the context of similar experiments described in literature, including heterogeneous video sequences commonly used in television programs, hence realized with professional quality. In particular, test cases were selected among the following video test sets:

- HDTV (720p-50Hz and 25 Hz) "SVT High Definition Multi Format Test Set" [7] - Video sequences produced by the Swedish television channel SVT, also available on the 'Video Quality Experts Group (VQEG)' web site ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/.
- CIF - Test Set: Xiph.org. In particular the "Derf" collection, available at http://media.xiph.org.

In order to execute tests, a sample of 20 video sequences (in what follows named *3, 4,…17, parkjoy, ducks, shields, parkrun,* and *mobcal*) in the two different formats (720P@25Hz, 720P@50Hz) were selected.

Finally, Tests were accomplished by using Linux OS (Ubuntu) on a Siemens Celsius V830 Workstation, RAM 8 GB, CPU 2 AMD Opteron 240, HD 2 HD SataII 400GB, VID NVIDIA Quadro FX 3400 - 256 MB. The computation of PSNR and SSIM was performed by using a workbench running on Windows 7.

Video sequences named *3, 4,….17,* were available in 720P@25Hz, and 720P@50Hz formats, respectively, whereas those named *parkjoy, ducks, shields, parkrun,* and *mobcal* were only available in 720P@50Hz format. Each available video sequence was compressed at 1000, 2000, 3000, and 4000Kbps, using both compared codecs, yielding a total of *280* different compressed files. Among these, only those having size ± 5\% than *F* have been considered,

$$F = \frac{br \times s}{8}$$

where

br: *bit rate* per second in *Kbps* (1000 bit/sec)
s: video duration in seconds
F: file size in *KBytes*

The 280 different compressed video files were successively evaluated through the *PSNR* and *SSIM* metrics, by using the *MSU Video Quality Measurement Tool* rel. 1.4, produced by the *Graphics & Media Lab Video Group* of *Moscow State University* [12].

Table 1 and Figure 2 show the SSIM averaged over the above mentioned video resolutions and bitrates. Table 2 and Figure 3 show average PSNR. It can be noticed that Eco Controllo's codec reaches better objective quality values. In particular, Figure 2 shows how the gap between the two codes is broader on the last five test video sequences. This is due to the fact that such video sequences have characteristics that stress video codecs, creating the conditions in which Eco Controllo's codec performs better. We noticed similar trends when using low bitrates, where Eco Controllo's codec degraded performances less than VP8.

Such results are not achieved by devoting more time for compression, since Eco Controllo's codec has turned out to execute much faster than VP8 on the selected test video sequences. This is also shown in Table 3 and Figure 4, reporting average time in seconds to compress the 20 video sequences with the two compared codecs.

Average SSIM		
Video Sequence	**VP8(IVF)**	**EcoControllo**
03	0,92441	0,94386
04	0,82851	0,86318
05	0,81535	0,84829
06	0,84205	0,85305
07	0,91652	0,93210
08	0,89555	0,90676
09	0,82534	0,84722
10	0,98139	0,98394
11	0,93429	0,94878
12	0,93717	0,94712
13	0,87540	0,91955
14	0,95164	0,96137
15	0,93231	0,94719
16	0,95868	0,96286
17	0,93164	0,93480
parkjoy	0,70267	0,79144
ducks	0,73882	0,80566
shields	0,91064	0,92825
parkrun	0,76141	0,84549
mobcal	0,92867	0,94903
Comprehensive Average	**0,89201**	**0,91270**

Table 1. Average *SSIM* comparison between EcoControllo (ECO10) and VP8(IVF).

Average PSNR		
Video Sequence	**VP8(IVF)**	**EcoControllo**
03	30,45273	31,07028
04	30,97422	31,42202
05	31,36040	32,15858
06	33,24558	33,31683
07	34,04667	35,36670
08	33,85275	34,59936
09	33,39754	34,15639
10	42,28937	42,66007
11	32,38916	32,88479
12	35,66304	35,89919
13	34,61392	35,78846
14	37,25661	38,12034
15	34,97223	36,20449
16	39,97067	40,22078
17	39,28353	39,36191
parkjoy	24,38476	25,75025
ducks	24,85895	26,24344
shields	33,99838	35,29050
parkrun	24,89755	26,58961
mobcal	34,16519	36,41307
Comprehensive Average	**33,94480**	**34,76193**

Table 2. Average *PSNR* comparison between EcoControllo and VP8(IVF).

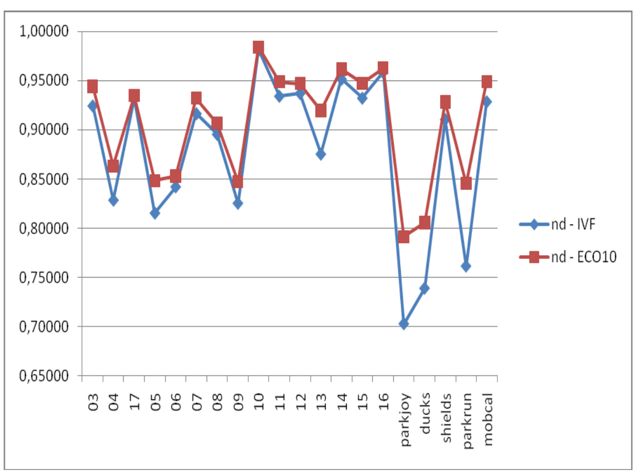


Figure 2. Average *SSIM* comparison between EcoControllo (ECO10) and VP8(IVF).

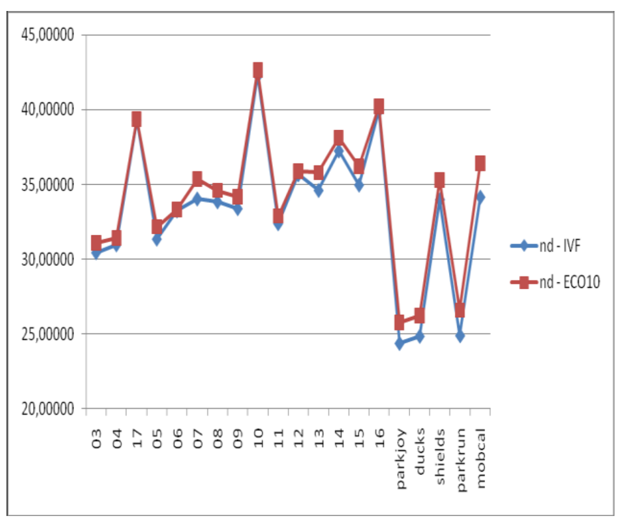


Figure 3. Average *PSNR* comparison between EcoControllo (ECO10) and VP8(IVF).

Average Compression Time (Seconds)		
Video Sequence	**VP8(IVF)**	**Eco-2010**
03	921,3333333	92
04	414,6666667	86
05	568,3333333	90
06	346,3333333	107
07	574,3333333	87,66666667
08	472	81,33333333
09	438,6666667	91,66666667
10	237,3333333	66,33333333
11	487,3333333	80,66666667
12	322,6666667	85,66666667
13	483	93,66666667
14	277,6666667	66,33333333
15	306	87,66666667
16	235,6666667	79
17	319	99,33333333
parkjoy	988	96,66666667
ducks	816,6666667	114,3333333
shields	574	88,66666667
parkrun	1093	79,66666667
mobcal	694,3333333	89,66666667
Comprehensive Average	**528,5166667**	**88,16666667**

Table 3. Average times (seconds) to compress 50Hz video sequences.

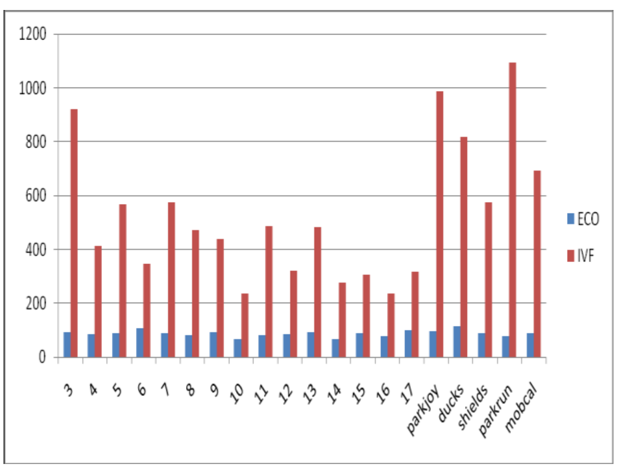


Figure 4. Average times (seconds) to compress 50 Hz video sequences.

4. Discussion

We have discussed the principles of a new codec for lossy video compression, and have provided some experimental evidence of its promising performances. In 2008 we had already provided a performance comparison of a former version of it with respect to main commercial implementations of H.264 and MPEG4 [1], also including results of subjective evaluations. Here we have compared the latest version of our codec with respect to VP8. Test results described in Section 3 reveal that Eco Controllo's video codec has been competitive with respect to VP8. In particular, although the performance advantage with respect to VP8 does not seem to appear relevant averaging over bitrates, it reaches higher gaps in critical situations, such as those in which video contents are particularly difficult to compress, or when particularly low bit rates are used. This is not a negligible issue, especially in those application contexts, such as HDTV, where critical situations can be frequent and it is not acceptable that codecs react by extremely reducing quality of videos or by loosing precious information.

In addition to technological and innovation aspects, Eco Controllo's research project has significant strategic potentials. We have already remarked that big commercial software brands consider proprietary codecs strategic to increase their market shares. In this scenario, Eco Controllo's codec can yield better quality/compression ratio, potentially opening new perspectives and drawing particular attention on this new compression technology.

Bibliography

[1] C. Cappelli, *Comparing the "Eco Controllo's video codec with respect to MPEG4 and H264"*, in proceedings of DMS09, "15th International Conference on Distributed Multimedia Systems", San Francisco, CA, USA, September 10-12, 2009.

[2] C. Cappelli, *Cenni sul codec video Eco Controllo*, Technical Report, Eco Controllo s.p.a., 2008.

[3] S.K. Chang, G. Polese, G.Tortora. *Performance Evaluation of Eco Controllo's Video Codec*,

Technical Report, University of Salerno, November 2010.

[4] M. P. Eckert and A. P. Bradley, `Perceptual quality metrics applied to still image compression', Signal Processing, vol. 70, pp. 177-200, Nov. 1998.

[5] A. M. Eskicioglu and P. S. Fisher, `Image quality measures and their performance', IEEE Trans. Commun., vol. 43, pp. 2959-2965, Dec. 1995.

[6] B. Girod, `What's wrong with mean-squared error', in Digital Images and Human Vision, A. B. Watson, Ed. Cambridge, MA: MIT Press, pp.207-220, 1993.

[7] L. Haglund, `The SVT High Definition Multi Format Test Set', Sveriges Television, 2006: ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/SVT_MultiFormat_v10.pdf.

[8] S. Ishihara. `Tests for colour-blindness', Handaya, Tokyo, Hongo Harukicho, 1917.

[9] ITU-R. "Methodology for the subjective assessment of the quality of television pictures, RECOMMENDATION ITU-R BT.500-11, 1-48, 2002.

[10] I.E.G. Richardson, ' H.264 and MPEG-4 video compression - Video Coding for Next-generation Multimedia', Wiley 2003.

[11] P. C. Teo and D. J. Heeger, `Perceptual image distortion', in Proceedings of SPIE, vol. 2179, pp. 127-141, 1994.

[12] D. Vatolin, D. Kulikov, A. Parshin, "MPEG-4 AVC/H.264 Video Codecs Comparison", Report of GRAPHICS & MEDIA LAB VIDEO GROUP at Moscow State University (MSU), http://www.compression.ru/video/codec_comparison/index_en.html.

[13] Z. Wang, `Rate scalable Foveated image and video communications', Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Texas at Austin, Austin, TX, Dec. 2001.

[14] Z. Wang. `Demo Images and Free Software for ``a Universal Image Quality Index'''. Available: http://anchovy.ece.utexas.edu/~zwang/research/quality_index/demo.html.

[15] Z. Wang and A. C. Bovik, `A universal image quality index', IEEE Signal Processing Letters, vol. 9, pp. 81-84, Mar. 2002.

[16] S. Winkler, `A perceptual distortion metric for digital color video', in Proceedings of SPIE, vol. 3644, pp. 175-184, 1999.

[17] Z. Wang, A. C. Bovik, and L. Lu, `Why is image quality assessment so difficult', in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 4, Orlando, FL, pp. 3313-3316, May 2002.

Collaborative and Assisted SKOS Generation and Management

Pierfrancesco Bellini, Antonio Cappuccio, Paolo Nesi

DISIT, Distributed Systems and Internet Technology, Department of Systems and Informatics,
University of Florence, Firenze, Italy, tel: +39-0554796523, fax: +39-055-4796363
email: pbellini@dsi.unifi.it, cappuccio@dsi.unifi.it, nesi@dsi.unifi.it , www: http://www.disit.dsi.unifi.it

Abstract— Despite to the presence of many systems for developing and managing structured taxonomies and/or SKOS models for a given domain, the production and maintenance of these domain knowledge bases is still a very expensive and time consuming process. This paper proposes a solution for assisting expert users in the collaborative development and management of knowledge SKOS. The proposed solution accelerates the SKOS production by crawling and exploiting different kinds of sources (in multiple languages and with several insistencies among them). The proposed WEB based tool supports the experts in defining relationships among the most recurrent concepts reducing the time to SKOS production and allowing collaborative production. The solution has been developed for Open Mind Innovative Space project, with the aim of creating a portal to allow industries at posing semantic queries about potential competences in a large institution such as the University of Florence.

Keywords-SKOS, Semantic Web, Skills Management System

I. Introduction

In the Semantic Web era, representing knowledge in the form of ontologies, thesaurus, taxonomies and other type of semantic data has became increasingly mandatory. The semantic markup is widely available, both to enable sophisticated interoperability among agents and to support human web users in locating and making sense of information.

Among the models for content semantic classification and enrichment the simplified knowledge organization system, SKOS is probably the most diffused [1]. It is a data model for sharing and representing knowledge organization systems (KOS) such as thesauri, classification schemes, term lists, controlled vocabulary and taxonomies within the framework of the Semantic Web. The SKOS model is typically defined using the Resource Description Framework (RDF) [2] and it allows information to be machine-understandable and processable by automatic software agents. Thus, for content classification, the adoption of a SKOS is one of the first step to enter in the semantic world.

The adoption of a knowledge representation structures for content involves many advantages:

- the semantic information is available in machine readable format and can benefit from the emergent technologies of the semantic web;
- the increasing spread of the semantic search engines like Swoogle [3], Watson [4] and SIndice [5] helps to find information with a high degree of precision;
- the users are helped in tagging their content using a predefined set of terms (also called vocabulary or terms, which can be multilingual and supported by a thesauri) belonging to the SKOS that should be validated and accepted by the community. The terms may be also provided in different languages and their translations validated as well.

The SKOS can be also a model at which one may tend after the collection of free tags from the users, the so called folksonomy. The free tags could be statistically analyzed to build a taxonomy and finally a SKOS adding related relationships. Tools for SKOS editing are quite diffuse, such as thManager [7] (a simple SKOS editor), SKOSEd [8], a Protégé plug in for editing SKOS.

However, despite the presence of emerging systems for developing and managing structured taxonomies and/or SKOS models for a given domain, the automated production and maintenance of these domain knowledge bases is still a process often very expensive and time consuming. In the literature, there exist few tools that support the process to pass from text to ontologies and/or SKOS, among them: PoolParty [9], TextToOnto [10], other methods [13] for semi-automatic conversion from well-defined thesaurus like MESH [14] or NCI [15] thesaurus in SKOS format.

In point of fact, software products on the market still require an important phase of manual collection of information from the domain and do not provide satisfactory mechanisms for the coordination of a collaborative production process performed by several groups. Thus, the modeling of domain-based SKOS often turns out to be a manual process and, in most case, it is time consuming and hard. It may involve a large number of personnel that are not easy to be coordinated to collaborate to a unique SKOS.

These problems become very critical when a large knowledge modeling is needed. For example, when one has to SKOSify the activities/documents of a local govern, or of a large University, or of a commercial/industrial district, or of a large editor/press such as IEEE, ACM, Springer, etc. The already in place standard classifications for companies (e.g., European and national codes, for example the ATECO or ISTAT); or for institutions or researchers (e.g., in Italy SSD, and CUN areas) are not suitable to match documents and content with the definitions of their "terms" and "key phrases". Most of those classifications have been frequently produced years ago and with the purpose of manual fitting of activities into them, and not with the purpose of using them for automated classification and reasoning with semantic tools. Moreover, in most cases the manual production of SKOS leads to stress some knowledge area/subarea depending on the knowledge of the experts involved, without taking into account the effective distribution of data to be ingested in the knowledge base.

As a matter of fact, the task of creating a SKOS requires a deep knowledge of the specific domain, and implies:

- the precise understanding of the semantic model behind a SKOS to avoid the production of terms which are not related each other by a specialization/generalization and/or relation relationships;
- the adoption of skilled personnel in both modeling knowledge and application domain or sub-domain;
- the domain analysis and thus the collection of terms in an organized form and relationships;
- a mechanism for coordination of activities in the various stages of the production task;
- the adoption of rules to avoid over-classification (over specialization in the SKOS hierarchy) and under-classification in some areas;

The previous description put in evidence why the knowledge base production process is a time and resource consuming task and not free of errors, even if the target is only the production of a SKOS.

A solution could be to start from the data/content to be classified, and directly extract from these sources the SKOS with an automated or semi-automated process and tool.

This paper proposes a novel solution for assisting expert users in the collaborative development and management of a knowledge modeled as a SKOS. The main idea is to realize a solution and tools to strongly accelerate the process of SKOS production exploiting the real documents/content and web pages to be indexed, and involving the experts in creating relationships among the most recurrent concepts. The solution proposed has been developed in a wider project called Open Mind Innovative Space, which has been founded by ECRF. Open Mind Innovative Space project has as main objective the realization of a portal on which the industries can pose questions with the aim of identifying the competences in terms of researchers and groups in the large knowledge of the University of Florence. In the literature, there is a number of systems that have been proposed to solve the above described problem of helping modeling knowledge bases, may be matching the *demand* (semantic query) against the *offer* (knowledge about domain).

Thus, the problem is still there; how to accelerate the production of SKOS when the domain knowledge is very large and the amount of information to be processed is huge.

The rest of paper is organized as follow. In Section 2, an overview of the Open Mind Innovative Space is provided. Section 3 reports the requirements of the identified tool for collaborative and assisted generation and management of SKOS. Section 4 shows the software architecture that implemented the solution in the global project framework, putting particular emphasis on multi language RDF-SKOS editor. Finally, in Section 5, the achieved experimental results are reported. Conclusions are drawn in section 6.

II. Open Mind Innovative Space Overview

As previously stated, the main goal of the Open Mind Innovative Space project is to realize a service to industries on which they can pose questions with the aim of identifying researchers and groups with the needed competences, knowledge among those of the University of Florence. The University of Florence includes more than 50 different departments belonging to all the scientific sectors areas, and hosting about 2000 researchers and more than 400 labs with their web pages. Each researcher may also teach at 2-3 courses; thus about 6000 course programs that may be considered competence descriptors as well. Moreover, the several research departments and researchers participate to research projects, for a total of about other 20.000 descriptors, etc. In such context it is very hard to identify a manageable number of people that could be reasonably entitled in terms of skill to create a shared common SKOS. This is due to the fact that the whole knowledge model has to be extracted from a huge amount of information, ranging from health care to geometry and math, from engineering to agriculture, from mechanics to statistic and pharmaceutics, etc. And, the sources of this knowledge may change quite dynamically, every year the courses are updated, the CV of people change, other publications and projects arrive, etc.

On the basis of the above description, the available information can be ingested from a large amount of different sources. This highly dynamic collection of sources may be automatically gathered through the use of software agents and crawling tasks. The information gained can be used by a semantic search engine to answer user queries with a high degree of precision. For example, by using an assisted semantic query interface with natural language query engine such as Aqualog [11]. This scenario is shown in Figure 1.

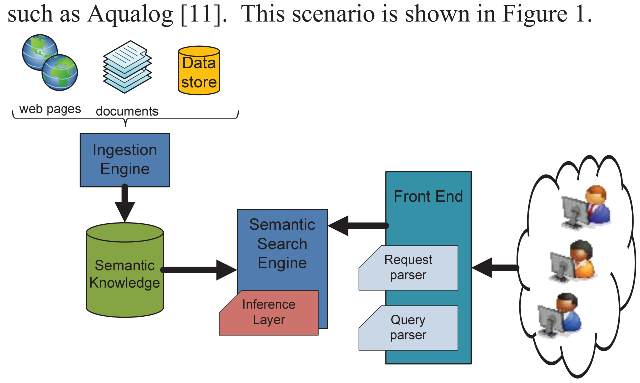


Figure 1. Ingestion and querying semantic knowledge base

This general model, is similar to those used in other mentioned state of the art solutions, and poses its bases on the availability of a good ontology, may be including a SKOS of the domain knowledge. The general ontology can be composed of dynamic blocks or subsections that describe the various features of the domain. They include both concepts that describe the physical structure of the domain and concepts that describe relations about personnel.

As a first step we had taken into account the fact that, the huge amount of data to be processed belong to governmental institutions such as the university and thus also on people. More specifically, the domain knowledge is composed by three self-supporting ontologies which are related by semantic relationships. Therefore, the basic elements of the knowledge base are those regarding:

- **Friend of a Friend (FOAF) ontology** [18] used to model many properties about Person and Organization class (professor, phd students, students, researchers, contractors, their relationships, research classification as SSD, CUN, etc.): the name, the surname, the e-mail properties and the *knows* relationship (applicable to individual belonging to the Person class).
- **Academic life ontology** is an ontology, we developed specifically for the Italian University case structure and terminology, that defines elements for describing universities and the activities that occur at them (labs, departments, faculties, research centers, groups, projects, courses, curricula, matter, projects, integrated labs, etc.). The main OWL entities and classes described by ontology are:
    - **Organization** class describes physical structures of university like research center, departments and laboratories;
    - **People and role** describe instances like full-professors, researchers and PhD students, related and derived from FOAF concepts;
    - **Activity entities** that cover concepts like past projects, ongoing projects and academic publications; To each person the specific publications are added as well, establishing in this way also relationships among the different authors.
- **Competences SKOS:** it is the SKOS ontology that describes the hierarchy of the technical skills of structures and people belonging to the given application context. This part of the knowledge is the most dynamic.

The components related to the **Academic life ontology** and to the **FOAF** are initialized and directly populated by gathering information from the central database of the University and of other institutions. Among them the central CINECA servers [19]. This operation is performed with a set of crawling tasks realized by using SOAP Client implemented in JAVA making use of JAX-WS [20].

On the basis of the described project, the most critical aspect is the modeling and population of the above mentioned **Competence SKOS** for the whole university area. Typically, in these cases the solution proposed is to manually produce a coarse classification. On the other hand, what it is really needed is to arrive at a SKOS strongly related to the real sources of descriptors to allow the automated classification and reasoning.

For these reasons we started with the idea of producing a solution for assisting expert users in the collaborative development and management of a **Competence SKOS**, the ***Collaborative SKOS Accelerator and Manager,*** **CoSKOSAM**. With the aim of accelerating the process of SKOS production and population. In the next section the identified requirements are presented.

III. REQUIREMENTS OF COSKOSAM

The CoSKOSAM is not only an editor, its requirements put in evidence that the aim was to create a collaborative environment in which several experts can contribute to the production and management of the same SKOS. And the systems may help them in identifying the keywords and concepts which are located in the real documents and sources. The main functional requirements for the CoSKOSAM have been to provide the capability of:

- ingesting and analyzing content from a large amount of different sources (web pages, cv, documents, etc.) to extract keywords and concepts and keep them linked with the original context/source;
- updating the crawling and ingestion of the content and thus the update of the semantic structures related to the identified keywords and concepts;
- helping the area editors with suitable tools that allow them to identify the most relevant keywords and concepts;
- managing content coming from different languages so that to map the concepts into a multilingual knowledge base, also providing/exploiting translation services and utilities;
- creating and editing a multi language SKOS about the identified concepts/competences/skills of personnel and research centers of the academic structure;
- supporting collaborative crawling, management and editing of the SKOS structures;
- allowing the incremental and distributed production of knowledge;
- supporting the integration of produced SKOS in terms of related terms and synonymous.

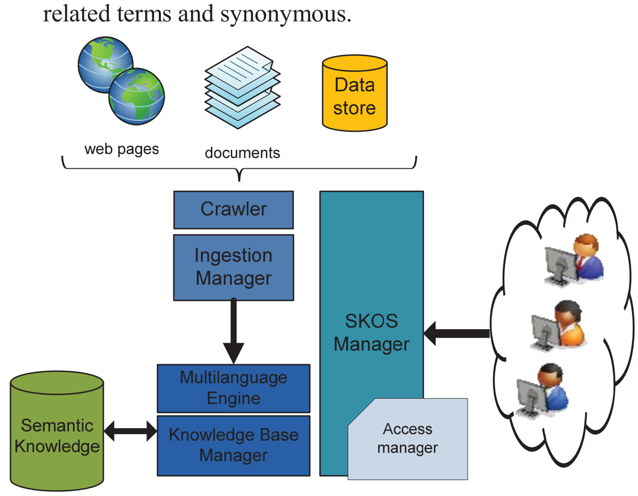


Figure 2. Architecture of Knowledge Base Management System

The benefits of creating a **Competence SKOS** via a WEB-based collaborative environment and management are many. Using CoSKOSAM skilled staff, such as directors of departments, or sector referent people, are enabled to build and manage the knowledge base incrementally and collaboratively confronting with their colleagues toward the common goal of realizing a high quality product.

IV. CoSKOSAM Architecture

The CoSKOSAM architecture is shown in Figure 2. It has been implemented in JAVA as a web application, using a client server architecture.

The architecture consists of some primary building blocks:

- ***SKOS Manager*** provides services for creating and maintaining a SKOS of keywords and concepts formalized according to a SKOS vocabulary.
- ***Access Manager*** offers services for the security login and grant check. It helps to synchronize and coordinate the access to the various sub-sections of the SKOS among different users. The privileges pyramid provisioned by system can be grouped in the following three main categories:
    - *Administrators users*: users who have the highest privileges on semantic store.
    - *Writer users:* users who can edit, read and write the working knowledge base but cannot execute backup and crawling activities;
    - *Reader users:* guest users. They can only read the knowledge base but cannot change it in any way;
- ***Crawler and Ingestion Manager*** offer the ability for crawling and ingesting sources to feed the knowledge base. The crawling task performs a breadth-first search on graph of the web structure of University and ingests some kind of domain-information like personnel, courses, staff curriculum, advertising, research centers, etc. Specifically, it makes use of the GATE NLP Platform [16] to implement the crawler architecture;
- ***Multilanguage Engine*** provides services for the management of multilingual SKOS taxonomy. The information ingested by crawling task is automatically translated by Multilanguage Engine in the languages handled by module. The sources are frequently produced in only one language while the keywords and concepts have to be declined in multiple language (e.g., in Italian and English at least).
- ***Knowledge Base manager*** provides the software API interface for manipulating the domain ontology. The knowledge base has been built by using the API provided by the Sesame framework [17].

A. *SKOS Manager*

The SKOS Manager provides services for creating, managing, and maintaining a multilingual SKOS model as comprised of concepts compounded from keywords extracted from several kinds of sources by the Ingestion Manager. The approach allows the organization of concepts into concept schemes where it is possible to indicate semantic relationships between terms. The SKOS Manager enables the complete development of a SKOS by providing a range of services, for the incremental, collaborative and multilingual development of the common knowledge via WEB.

Each information about Competence SKOS that is ingested by previous crawling processes is stored in the knowledge base as an instance of the skos class: *Concept*. A user of the SKOS Manager, which is expert in a given sub-domain such as a single department, has the chance to change the knowledge base by adding new concepts and inserting semantic relationships among already existing concepts.

The relations between concept over the SKOS vocabulary allow to add semantic information to the knowledge base.

The allowed relations are:

- **skos:conceptSchema** relation provides the ability to express the origin of a concept in a concept scheme;
- **skos:hasTopConcept** relation provides the ability to express the mayor topics that are wrapped up into a concept scheme;
- **skos:broader** relation must be used to express the fact that a concept is in some way less general than another.
- **skos:narrower** is the inverse relation of skos:broader;
- **skos:related** relation provides the ability to create associative links between concepts. The property carries weak semantics and it expresses the fact that two concepts are in some way related, and that the relationship should not be used to create a hierarchy but for create links between branches of a hierarchy of concepts;
- **skos:prefLabel** is the preferred label associate to concept in a given language. A label is any word, phrase or symbol that can be used to refer to the concept by people. When a user adds a label to SKOS the system provides to automatically translate the text-label in the right language by exploiting an external service (in any case the translation may be corrected by the user).

A graphic explanation of generated SKOS vocabulary is shown in Figure 3.

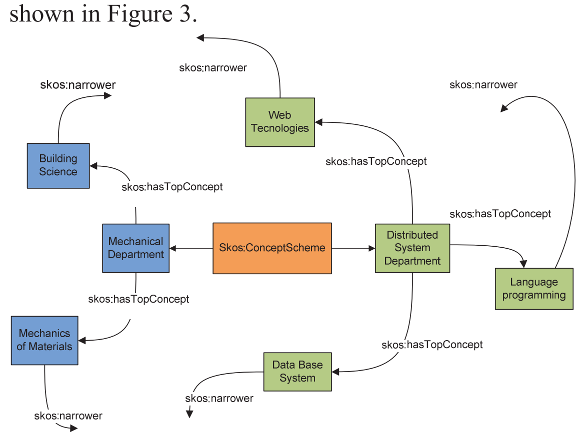


Figure 3. Generated SKOS Knowledge Base

Multilingual and multicultural issues are dealt to assure a wider and more effective exploitation of data beside the background of the operator and their location. Furthermore a multilingual approach helps to improve *precision and recall* of popular search engine, which are very good at retrieving the accurate information.

Other features provided by the system are:

- searching by label in the working taxonomy and discovery branches that contain the target label;
- view the semantic information about personnel or structures related to each *skos:Concept*; that is the link to the original sources;
- view the frequency with which the concepts occur in source web documents;
- filtering the concepts by frequencies. In this way the user has the opportunity to discard some labels that may be considered statistically less remarkable or highly specializing;
- ability to edit concepts in multilingual mode taking advantage of automatic translations. An user can also manually re-translate a label for achieving an higher quality result;
- visual system log for provide additional information about working knowledge base;
- more than a graphical view for displaying the SKOS concepts:
  - only label: displays the label in the current language
  - label with frequencies: displays next the label, the frequencies of the current concept;
  - label with language: displays next to the label, the dual language to the current one.

V. EXPERIMENTAL RESULTS

In this section, the achieved results and some user experiences are presented. The proposed tool has been used to develop an experimental knowledge base related to DSI Department, one of the most active departments of the University of Florence [21]. It is under application by 3 different departments, while the development of the Competence SKOS for each department is autonomously developed as separate SKOS branches. They are lately merged together in an unique SKOS by joining concepts and adding skos:related relationships.

The aim of this experimentation has been: (i) the ingestion of the whole sources related to DSI department; and (ii) the building of a multilingual Competence SKOS about department and its personnel. The department included about 62 researchers (28 full professors, 13 associate professors, 21 researchers), and about 160 courses and programs, related publications and web pages, for more than 2000 publications. All staff accessible via web.

The department includes 5 different sectors/research-areas such as math, operating research, computer science, computer engineering, automated control. The people and activities of those areas have different competences and skills and are difficult to be represented by a unique person.

The CoSKOSAM started crawling and ingesting all data and producing the list of basic concepts as presented in Figure 4 in which the SKOS Manager is shown.

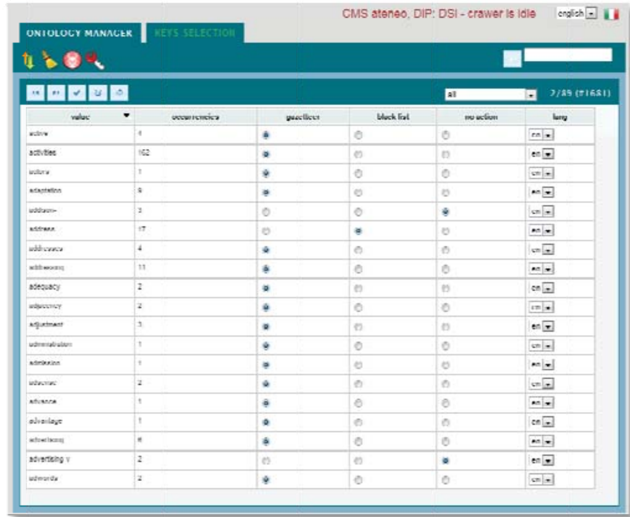


Figure 4. SKOS Manager Web user interface – setting up crawling task stage for basic concepts

The extracted basic concepts/keywords has been browsed by the reference domain expert to identify those that have been erroneously classified to the wrong language. The list of them can be filtered to present only those that have a frequency greater than a given value. This allows the expert to focus its time on the most relevant one. At the end of this process, the selected keywords are put in the gazetteer of the semantic Ingestion Manager based on GATE [16]. In the case of DSI, the process extracted about 1600 terms aligned by the system in both languages, revised by the expert in about 1 hour of work.

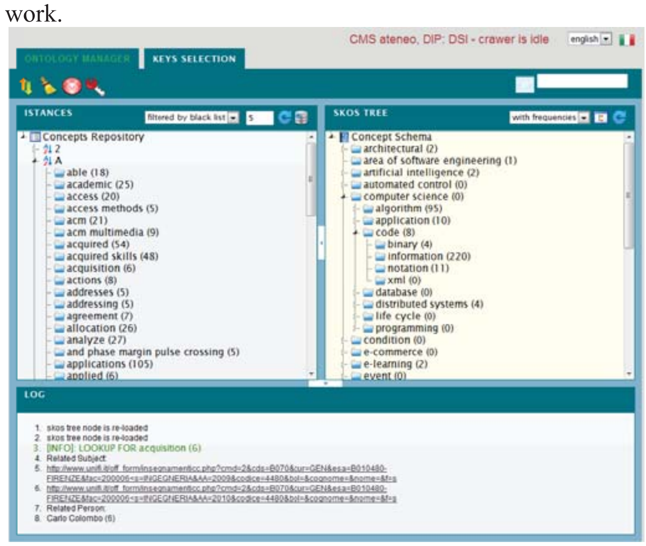


Figure 5. SKOS Manager Web user interface – setting up Conpetences SKOS from the bsic concepts.

After this phase, the basic concepts and keywords can be used to create the Competence SKOS passing to a different interface of the SKOS Manager, as shown in Figure 5. On the left side, the list of basic concepts with their frequency (alphabetically ordered), while on the right side; the produced Competence SKOS are reported. By using the mouse, the user of the SKOS Manager may look up at the original sources referred by both basic concepts or SKOS concepts and reported in the log box below.

The SKOS on the right side can be manually created by adding new concepts or by using the drag and drop paradigm by exploiting and arranging the basic concepts on the left side.

The produced Competence SKOS for the DSI consists of different classes and has required 1 day of work for its completion.

At the end of the Competence SKOS production for DSI, the remaining basic concepts not allocated were 1937 and the produced SKOS was based on 786 (for each language) concepts organized in 27 levels. The domain OWL ontology for the DSI only has 3270 individuals and 45 OWL classes.

The Competence SKOS produced has been validated by experts of the sectors by using the look up facility and in short by observing the sources identified and connected a large sample of broader concepts. Another form of validation has been the production of simple SPARQL queries with and without the usage of the inferential engine. The inferential engine, in this case, exploited both the hierarchy of concepts and the related relationships the SKOS.

VI. CONCLUSIONS AND FUTURE WORK

This paper proposed CoSKOSAM, a web based solution for accelerating the production and management of SKOS derived from sources that can be located on Web. The solution and proposed tool has been used for the automated ingestion and analysis about the university life, including afferent organization and technical skills. The hierarchical structure of the competences and the semantic relation among them have been formalized using the SKOS vocabulary, providing a developing method collaborative, computer-aided and coordinated activity.

The methodology greatly reduces the time spent in the development process aiding the users in all stages of the production process. Furthermore, the ontology is produced according to the OWL/RDF/SKOS rules and can benefit from emerging technologies and innovations offered by the semantic web. The generated ontology can be used as information domain by a demand and supply system about academic skills. It is currently in connection with a semantic database for allowing performing SPARQL queries allowing:

- semantic search engine to retrieve ranked information. For computing ranking it is possible making use of term frequency as a factor weighting within the ranking algorithm;
- semantic indexing for search engine optimization and fuzzy queries.
- exploiting inferential engine to increase the system intelligence,
- improving the engine for providing results to the users and permitting them to navigate in the mesh of relationships among FOAF entities and results;

In addition of this, some guide lines for short-term future work have been already planned, in particular regarding usability and user support.

Acknowledgements: The authors would like to thank all contributor to the Open Mind project and in particular Prof. M. Lombardi and Prof. P. Rissone.

REFERENCES

[1] SKOS: Simple Knowledge Organization for the Web. [Online]. Available: http://www.w3.org/2004/02/skos/

[2] RDF – Vocabulary Description Lanaguage 1.0: RDF Schema. [Online]. Available: http://www.w3.org/TR/rdf-schema/

[3] D. Li; T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. C Doshi, and J. Sachs (2004). "Swoogle: A Search and Metadata Engine for the Semantic Web". Proc. of the Thirteenth ACM Conference on Information and Knowledge Management (CIKM). ACM. pp. 652–659.

[4] Mathieu d'Aquin and Enrico Motta, "Watson, more than a Semantic Web search engine", in press.

[5] G. Tummarello, E. Oren, and R. Delbru. Sindice.com: Weaving the open linked data. In Proceedings of the 6th International Semantic Web Conf and 2nd Asian Semantic Web Conf (ISWC/ASWC2007), Busan, South Korea, vol.4825 of LNCS, Berlin, Nov. 2007. Springer Verlag.

[6] M. d'Aquin. Building SemanticWeb Based Applications with Watson. In WWW2008 - The 17th International World Wide Web Conference - Developers' Track, 2008.

[7] ThManager – an Open Source Tool for creating and visualizing RDF SKOS vocabularies

[8] SKOSEd-Thesaurus editor for the Semantic Web. [Online]. Available: http://code.google.com/p/skoseditor/.

[9] Pool Party - SKOS Thesaurus Management. [Online]. Available: http://poolparty.punkt.at/

[10] Alexander Maedche and Steffen Staab. The TEXT-TO-ONTO Ontology Learning Environment. Institute AIFB, University of Karlsruhe, 76128 Karlsruhe, Germany.

[11] V. Lopez, E. Motta, E., V. Uren, and M. Pasin, (2007) AquaLog: An ontology-driven Question Answering System for Semantic intranets, Journal of Web Semantics, 5(2), p.72-105.

[12] Protégé, Protégé enviroment. [Online]. Available: http://protege.stanford.edu.

[13] M. van Assem, V. Malais´e, A. Miles, and G.Schreiber. A method to convert thesauri to SKOS. In Proceedings of the third European SemanticWeb Conference (ESWC 2006), pages 95–109, Budva, Montenegro, June 11-14 2006.

[14] Soualmia L.F., C. Goldbreich, and S.J. Darmoni. Representing the mesh in owl: Towards a semi-automatic migration. InProc. of the 1st Int'l Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004), pages 81–87, Whistler, Canada, 2004.

[15] J. Goldbeck, G. Fragoso, F. Hartel, J. Hendler, B. Parsia, and J. Oberthaler. The National Cancer Institute's Thesaurus and Ontology.Journal of Web Semantics, 1(1), Dec 2003.

[16] Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.:GATE: A Framework and Graphical Development Enviroment formRobust NLP Tools and Applications. In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL '02), Philadelphia (2002).

[17] Arjohn Kampman, Frank van Harmelem, Jenn Broekstra: Sesame: Ageneric architecure for storing and querying rdf and rdf schema: in proceedings of ISWC 2002, October 7-10, Sardinia, Italy (2002).

[18] The Friend of a Friend (FOAF) Project. [Online]. Available: http://www.foaf-project.org/.

[19] Centro di supercalcolo, Consorzio di Università. [Online]. Available: http://www.cineca.it/page/chi-siamo

[20] JAX-WS. [Online]. Available: http://jax-ws.java.net/.

[21] DISIT Lab – Department of Systems and Informatics, University of Florence. [Online]. Available:http://www.disit.dsi.unifi.it/.

MAMA: A Novel Approach to Ontology Mapping

Shi-Kuo Chang[1], Francesco Colace[2], Massimo De Santo[2], Emilio Zegarra[1], YongJun Qie[1]

[1]Department of Computer Science, University of Pittsburgh Pittsburgh, PA 15260 USA
[2]DIEII - Department of Electronic and Information Engineering, University of Salerno, Italy
{chang, ezegarra}@cs.pitt.edu, yongjun@pitt.edu,{fcolace, desanto}@unisa.it

Abstract

Ontology is a key factor for enabling interoperability across heterogeneous systems, services and users. One of the most challenging tasks in its use is ontology construction. However, experts usually represent the same knowledge domain by the use of different ontologies that can differ in structure, concepts, relations and attributes. Therefore, ontology could lose its main feature: allowing the semantic interoperability among actors working on the same knowledge domain. An effective solution to this problem is the introduction of methods for finding mapping among the various components of ontologies. In this paper, a novel approach to the ontology mapping is proposed. The proposed approach, named MAMA, investigates the possibility of finding overlaps among different ontologies that describe the same knowledge domain through the combined use of syntactic, semantic and topological similarity indexes. The indexes are combined to define the degree of similarity between the various components of ontologies by introducing a combining rule. This rule is adaptive and automatically emphasizes the contributions of those indexes that provide better results in certain operative conditions. The proposed approach has been tested on standard datasets and the obtained results are very promising. Moreover, we are currently exploring the application of the ontology mapping approach to software component reuse.

1. Introduction

Distributed systems are by their nature heterogeneous and so the various players in the system need tools that allow them to share services and resources in it. In this complex scenario, an effective solution is the adoption of the ontology formalism. The term 'ontology' was first used in the computer science field by Gruber referring to an explicit specification of a conceptualization [13]. So an ontology is defined as a formal, explicit specification of a shared conceptualization [11] and plays an important role in the domains of software engineering, artificial intelligence, knowledge management, information integration and semantic web [2]. Ontology is playing a significant role in information systems, semantic web and knowledge-based systems, where 'ontology' refers to "the representation of meaning of terms in vocabularies and the relationships between those terms" [1]. The role of ontology, then, is to represent the main components of a knowledge domain according to the view that a generic user has about it. However, for the same knowledge domain each user may create a different ontology, which differs in structure, concepts, attributes and relationships from the other ones. In other words, an ontology, while describing the same domain of knowledge, may be different from another because both contain the point of view of a particular user, and how it was built. Therefore, ontology could lose its main feature: allowing the semantic interoperability among actors working on the same knowledge domain [12].

An effective solution to this problem is the introduction of methods for finding mapping among the various components of ontologies. The main approaches to the ontology mapping can be grouped into three categories according to the pursued approach, namely: the lexical category, the semantic category and the structural category. In the literature, many approaches in these three categories have been developed, but none is appropriate for ontology mapping. Each of the proposed approach, in fact, shows good performance in certain situations, while exhibiting weaknesses in others. An approach in the structural category, for example, works well when ontologies shows similar structure, but is unable to recognize semantic similarities related the name of the concepts.

In this paper, a novel methodology for ontology mapping is proposed. The starting point of the proposed methodology is the assumption that the previous approaches can be complementary and so they can reach better results when working together. The contributions of various approaches are combined by an adaptively weighted rule in order to emphasize those that work better in the various scenarios. In Section 2 the proposed ontology mapping approach, named MAMA, will be described. In Section 3 the initial experimental results obtained by the use of the proposed approach on standard datasets will be discussed. In Section 4 we explore the application of ontology mapping to software component reuse, which is an important issue in software engineering. Some initial experimental results in software component reuse by ontology mapping will be presented.

2. The Proposed Ontology Mapping Approach

What is an ontology? This question could seem trivial but it is still difficult to give a unique definition, An ontology's formal definition could be the following [14]: O={C, H_C,H_R,A,I}, where

- C is the set of concepts in a domain
- H_C is the set of generic relations among concepts
- H_R is the set of the hierarchical relations among concepts
- A is the set of axioms
- I is the set of concepts' instances

Keeping in mind the above definition, ontology mapping is an operation that can be described as follows: given two ontologies A and B, mapping one ontology to another means that for each concept (node) in ontology A, a corresponding concept (node) in ontology B is found, which has the same or similar semantics in ontology B and vice versa [15].

According to this definition, an ontology mapping function can be defined as follows: MAP: O_1 in O_2: MAP(e_1) in e_2 if sim(e_1, e_2)>s with sim(e_1, e_2) being the similarity coefficients between e_1 and e_2 , s being the threshold, e_1 and e_2 being the entities(concept, instance, attribute etc.) in two ontologies. This process can be iterated for any number of ontologies to obtain the mapping functions among them.

In this paper, a novel methodology for ontology mapping is proposed: this approach, named MAMA (Mapping And Merging Approach) is based on the combination of various indexes working at lexical, semantic and structural level to find equivalence among the various components of the ontologies. The proposed mapping approach could be so described:

```
// create mapping groups
for (c ∈ O_1,...,O_N) do
  if (type(c)="class") then
        add c to groups_cls
  else if (type(c)="property") then
        add c to groups_props
      end if
  end if
end for
classify (groups_cls)
classify (groups_props)
```

The generic classification phase `classify(groups_in)` can be so described:

```
//classify(group_in)
if(size(group_in)>1) then
  remove concept c_r from group_in
for (∀c ∈ group_in) do
  if (c_r ∈ O_i _and _c ∈ O_j with _i ≠ j) then
    if(type(c_r) = "class" and type(c)="class")then
    index = calculateIndexSimilarityClasses(c_r, c)
    else
      if(type(c_r) = "property" and type(c)="property") then
      index = calculateIndexSimilarityProperties(c_r, c)
      end if
    if(index>threshold) then
      add mapping found between c_r and c
    end if
  end if
end for
  classify(group_in)
end if
```

Therefore the mapping process could be seen as a classification task. In particular, the proposed algorithm classifies the similarities among the classes, the properties and the relationships and then creates a new ontology that will be the common layer capable of bridging the various ontologies.

As previously said, various indexes are introduced in order to calculate the similar components among the ontologies. Some of them are adopted without change from the literature, while others have been improved and modified [3]-[10]. The introduced indexes are:

Editing Distance (ED): This index is so defined:

$$sim_{ed}(x,y) = \max\left(0, \frac{\min(|x|,|y|) - ed(x,y)}{\min(|x|,|y|)}\right) \in [0,1]$$

It aims to calculate the likelihood between two words that label concepts in the ontology.

Trigram Function (TF): This function aims to measure the number of similar trigrams that are in the words that label the concepts in the ontologies.

$$TF(x,y) = \frac{1}{1 + |tri(x)| + |tri(y)| - 2 * |tri(x) \cap tri(y)|} \in [0,1]$$

The function tri(x) gives the set of trigrams that are in the in the word x.

Semantic similarity index (SS): This index is so defined

$$SS(w_1,w_2) = \frac{1}{sim_{jc}(w_1,w_2)} \in [0,1]$$

where

$$sim_{jc}(w_1,w_2) = 2 * \log P\big(LSuper(c_1,c_2)\big) - (\log P(c_1) + \log P(c_2))$$

This index aims to compare from a semantic point of view two words according to the taxonomy defined in Wordnet [16]. In particular this index measures the distance of the words in the taxonomy defined in Wordnet.

Granularity (GR): This index aims to measure the mutual position of the words representing the concepts of the ontology in the WordNet. This index is so defined:

$$GR(c_1,c_2) = \frac{\min\big(Dens(c_1) * path(c_1,p), Dens(c_2) * path(c_2,p)\big)}{\max\big(Dens(c_1) * path(c_1,p), Dens(c_2) * path(c_2,p)\big)} \in [0,1]$$

where dens(c) is the function representing the density of the concept c. This function is defined as E(c)/E where E is the ratio between the number's arc of the concept and the numbers of its parents while E(c) is the number of the sibling of the concept c. The function path(c_1, p) is the shortest path from c_1 to p that is the first parent common to c_2

Attribute Index: This index aims to measure the numbers of similar attributes between the two nodes. In particular it is so defined:

$$sim_{att} = \frac{|X \cap Y|}{|X \cap Y| + \alpha(x,y) * |X/Y| + (1 - \alpha(x,y)) * |Y/X|} \; con \; 0 \leq \alpha \; \leq 1$$

Synonym Index (SI): This index aims to verify if in Wordnet there are synonyms of the word related to the concept in an ontology that label a concept in another ontology. This index can assume value 0 or 1.

Derived Index (DE): This index aims to find in WordNet an adjective, representing a node of an ontology, derived from the label of a concept that is in the other ontology. This index can assume value 0 or 1.

Property Similarity Index (ISP): This index has the aim to verify the equality between the nodes evaluating their properties. In particular the following indexes are introduced

- Equality Indexes among superclasses (ISC): this index verifies if the superclasses of the comparing classes are similar. This index can assume value 0 or 1.

- Equality indexes among equivalent classes (IEC): this index compares all the classes that are equivalent to the comparing classes. This index assumes value 1 if all the classes are equivalent and 0 otherwise.

- Equivalent Indexes of inherited classes (IIC): this index assumes value 1 if all the inherited classes of the comparing nodes are similar. Also in this case this index assumes value 1 if all the inherited classes are equivalent and 0 otherwise.

- Equivalent Indexes of disjoint classes (IDC): this index evaluates the similarity among the disjoint classes of the comparing nodes. This index assumes values 1 if all the disjoint nodes are similar, and 0 otherwise.

The full similarity index (ISP) is obtained by the following formula:

ISP = ISC*IEC*IIC*IDC

This index can assume value 0 or 1.

Similarity Index for entities (ISI): This index evaluates if the entities derived from the comparing nodes are equal. The comparison is made by evaluating both the number of entities and their typology. This index can assume value 0 or 1.

Acronym (AC): This index aims to verify if in the two comparing nodes one word is the acronym of the other. If it is true this index is 1, otherwise it is 0.

Fingerprint Index (IM): This index verifies if the word that describes a comparing node is in the other word that describes the other nodes. If the word is contained in the other one this index is 1, otherwise it is 0.

Abbreviation (AB): This index measures if a word that describes a node is an abbreviation of the other that describes the other comparing node. If the word is an abbreviation of the other one this index is 1, otherwise it is 0.

Label (LA): This index measures if the two labels of the comparing nodes are equal. Also, in this case the index assumes value 1 if the nodes have the same label, and 0 otherwise.

The introduced indexes can be grouped in three sets:

Syntactic indexes: These indexes aim to detect the syntactical similarities among the various components of the ontology. The following indexes are syntactic indexes:

- Editing Distance (ED)
- Trigram Function (TF)
- Acronym (AC)
- Fingerprint (IM)
- Abbreviation (AB)
- Label (LA)
- Attributes (ATT)

Semantic indexes: These indexes aim to compare the ontologies from a semantic point of view. As previously said these indexes can use structured knowledge as in the WordNet. The set of semantic indexes includes:

- Semantic Similarity (SS)
- Granularity (GR)
- Synonym Index (SI)
- Derived (DE)
- Label (LA)

Structural indexes: The indexes belonging to this set aim to compare the ontologies from a structural point of view. The indexes of this set are:

- Attributes (ATT)
- Similarity Index for properties (ISP)
- Similarity Index for Entities (ISI)

The three sets are so combined in order to map the various nodes that are in the ontologies:

$$Mapping(X,Y) = \theta * IndSin(X,Y) + \sigma * IndSem(X,Y) + \omega * IndStr(X,Y)$$

where:

$$\theta + \sigma + \omega = 1$$

In particular:

$$\theta = \frac{IndSin(X,Y)}{IndSin(X,Y) + IndSem(X,Y) + IndStr(X,Y)}$$

$$\sigma = \frac{IndSem(X,Y)}{IndSin(X,Y) + IndSem(X,Y) + IndStr(X,Y)}$$

$$\omega = \frac{IndStr(X,Y)}{IndSin(X,Y) + IndSem(X,Y) + IndStr(X,Y)}$$

and

$$IndSin(X,Y) = 0.5 * (\alpha * ED + \beta * TF) + 0.5 * (\max(AC, IM, AB))$$
$$\alpha = ED/(ED + TF) \text{ e } \beta = TF/(ED + TF)$$

$$IndSem(X,Y) = 0.5 * (\gamma * SS + \delta * GR) + 0.5 * (\max(SI, DE, LA))$$
$$\gamma = SS/(SS + GR) \text{ e } \delta = GR/(SS + GR)$$
$$IndStr(X,Y) = 0.5 * (ATT) + 0.5 * (\max(ISI, ISP))$$

After this first step, the mapping among the various nodes that are in the N ontologies is obtained. The second step is the mapping among the relations. So the following index is introduced:

$$\boldsymbol{IndRel(x,y) = \min(Mapping(A,C), Mapping(B,D), RO(x,y))}$$

where x and y are the comparing relations while A and B are their domains and C and D are their co-domains. In particular,

$$\boldsymbol{RO(R_1,R_2) = \sqrt{CM(d(R_1), d(R_2) * CM(r(R_1), r(R_2)}}$$

and

$$\boldsymbol{CM(C_1,C_2) = \frac{|UC(C_1,H_1) \cap UC(C_2,H_2)|}{|UC(C_1,H_1) \cup UC(C_2,H_2)|}}$$

In this case H_1 is the taxonomy related to the concept C_1 while H_2 is the taxonomy related to the concept C_2. The function UC (Upward Cotopy) is so defined:

$$UC(C_i, H) = \{C_j \in C | H(C_i, C_j)\}$$

The last step is the mapping among the attributes. This task is accomplished by the introduction of this index:

$$IndAtt(x,y) = \max\left(IndSin(x,y), \min\left(Mapping(A,C), equal\left(type_{range_x}, type_{range_y}\right)\right)\right)$$

After this phase the mapping process among the ontologies is obtained.

3. Experiment Setup and Results

In order to test the performance of the MAMA approach an experimental setup on standard datasets has been developed. In particular the experimental approach adopted was the same one developed in the SEALS project. The SEALS Project has developed a reference infrastructure known as the SEALS Platform to facilitate the formal evaluation of semantic technologies. This allows both large-scale evaluations to be run as well as ad-hoc evaluations by individuals or organizations. The SEALS evaluation setup aims at evaluating the competence of matching systems with respect to different evaluation criteria. The evaluation will focus on demonstrating the feasibility and benefits of automating matching evaluation. In this evaluation a limited set of criteria has been considered:

- Conformance: standard precision and recall, restricted semantic precision and recall, coherence

The evaluation setup contains three different scenarios, where the tools are evaluated according to common datasets and criteria:

- Scenario 1: Test data: Benchmark. Criteria: conformance with expected results
- Scenario 2: Test data: Anatomy. Criteria: conformance with expected results
- Scenario 3: Test data: Conference. Criteria: conformance with expected results and alignment coherence

The datasets were selected based on the existence of reliable reference alignments and experiences with using the datasets in evaluations:

- Systematic benchmark: the goal of this benchmark series is to identify the areas in which each matching algorithm is strong or weak. The test is based on one particular ontology dedicated to a very narrow domain of bibliography and a number of alternative ontologies on the same domain for which alignments are provided.

- Conference: collection of conference organization ontologies. This effort was expected to materialize in alignments as well as in interesting individual correspondences ('nuggets'), aggregated statistical observations and/or implicit design patterns.
- Anatomy: the anatomy real world case is about matching the Adult Mouse Anatomy (2744 classes) and the NCI Thesaurus (3304 classes) describing the human anatomy.

So in order to evaluate the performance of the proposed approach the following indexes, suggested by the SEALS project, have been adopted:

Precision = #Correct_Mappings/(#Correct_Mappings + #Wrong_Mappings)
Recall = #Correct_Mappings / (#Correct_Mappings+#Missed_Mappings)

$F_{Measure} = [(b^2+1)*Precision*Recall]/(b^2*Precision+Recall)$

In the $F_{Measure}$ evaluation index the parameter b has been set to 1 in order give the same importance to the precision and recall parameters as suggested . The results obtained by the use of the MAMA approach has been compared with the same ones obtained by other methodologies developed in the literature. The experimental results are shown in Figure 1.

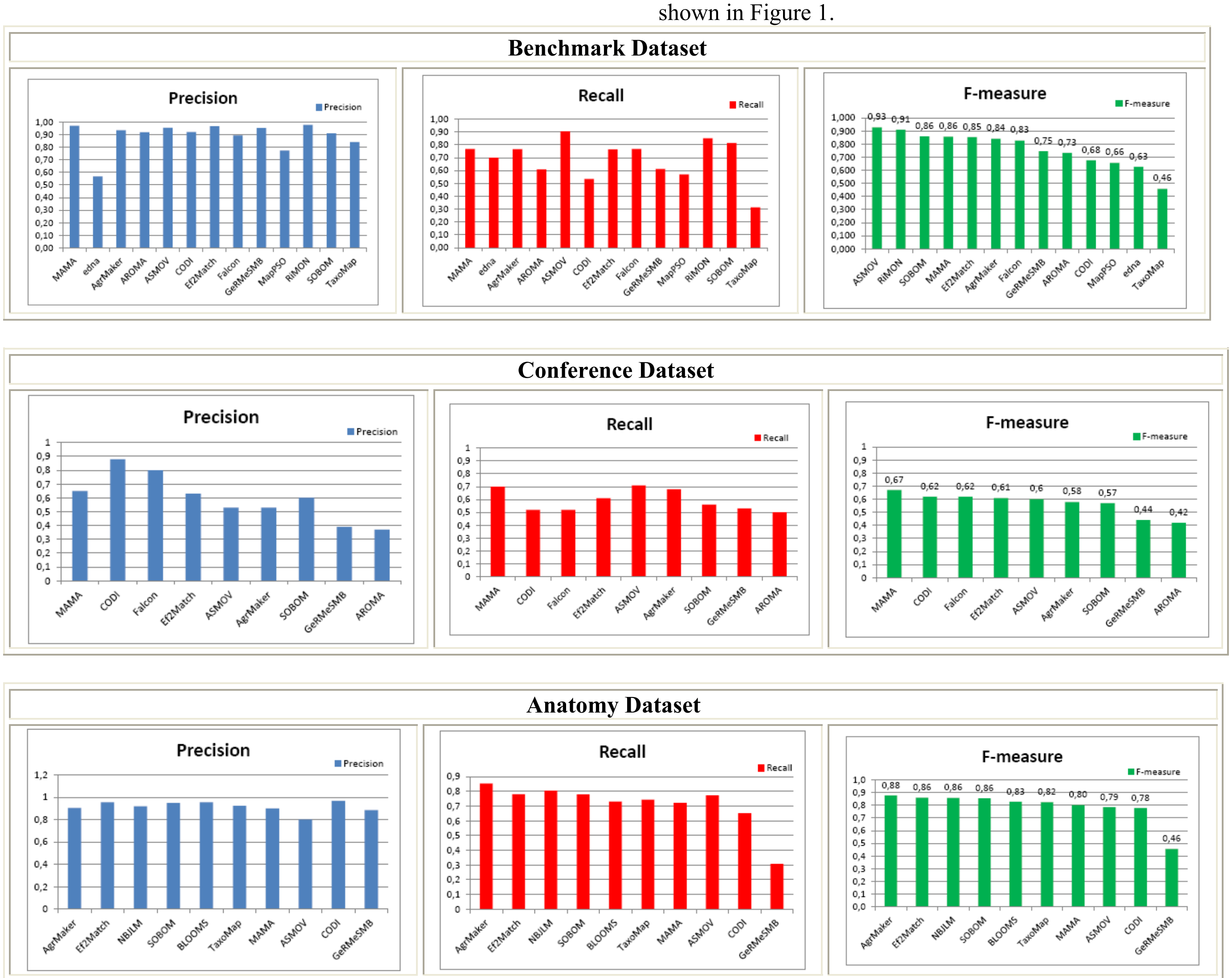


Figure 1. Experimental results of the MAMA approach.

In order to measure the overall performance of the system the average value of $F_{Measure}$ parameter was evaluated (Figure 2).

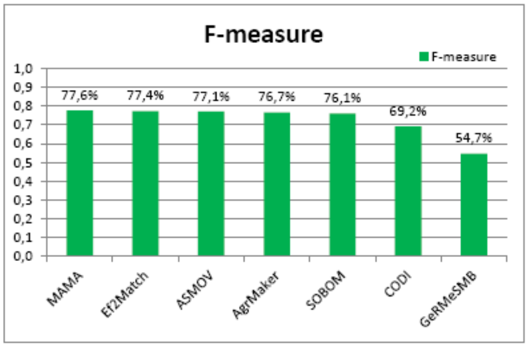


Figure 2. The average F-measure.

The MAMA approach shows good results for each dataset, although it does not always achieve the best result. On the other hand, it is interesting to underline that the MAMA approach shows the best value of average $F_{Measure}$. In fact, some approaches work well only for some datasets under certain conditions, while the MAMA approach is general purpose. As depicted in Figure 4 only six approaches, along with the MAMA approach, can be applied for each dataset and condition. The MAMA approach, as previously said, handles the ontology mapping problem in a very general way and therefore is able to adapt itself to the various cases.

4. Component Reuse based upon Ontology Mapping

Ontology mapping, which is an important part of ontology integration, can promote sharing and communication among different ontologies. The incorporation of ontology into software engineering can improve the reuse of software assets effectively [18]. In recent years, it has become less likely to develop complete new software systems from scratch. It becomes very important to develop software by adapting or combining existing reusable components [17]. We observe that requirement specification can provide a data source for ontology model and also the vital link for the combination of software engineering and ontology [19]. With this insight, a software component reuse approach based on ontology mapping is formulated in Figure 3.

The main idea is to process customer requirements and reusable components using ontology mapping techniques, and then construct the mapping between ontology nodes and reusable components. This approach can promote the reuse of software components. We can construct the target ontology model by analyzing requirement specification and then calculate the similarity between target ontology and source ontology using ontology mapping techniques. Therefore, we can identify the matched source ontology nodes and the corresponding sets of reusable components and then construct the mapping between the target ontology nodes and the reusable components. We assume that the mapping between source ontology nodes and reusable components has been realized, so every source ontology node has matched several reusable components. After the ontology mapping, the target ontology nodes also have matched the reusable components through the "bridge" of the source ontology nodes.

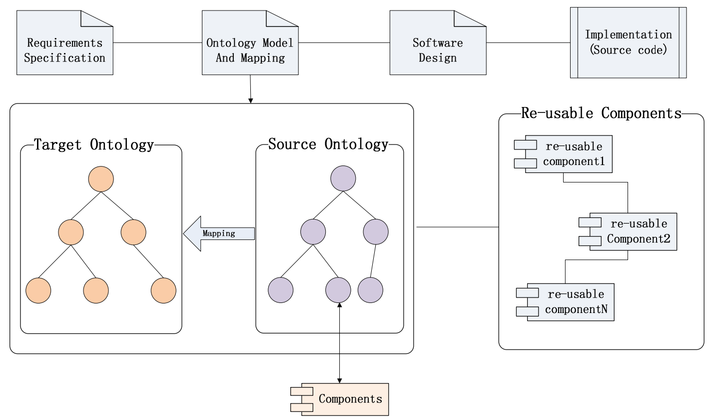


Figure 3. Software component reuse approach based on ontology mapping.

CROM Algorithm: Through the above analysis, we find that the mapping between ontology nodes and reusable components is the key to realize the CROM (Component Reuse through Ontology Mapping) algorithm. We can construct the mapping by using requirement engineering. With requirement engineering, we will decompose requirement specification into several fragments and every fragment of requirement contains several functional points. Each functional point contains the input and output which are designed to match the requirement. The requirement and the corresponding input/output are the basis for the design and implementation of software components. Each node of ontology model is related to each fragment of requirement specification, and each fragment is related to several functional points so that each node of ontology model is also related to several functional points. This unique approach to construct the mapping between ontology node and reusable components by the functional point is the heart of the CROM algorithm. In general, the mapping operation is illustrated in Figure 4.

Ontology nodes and reusable components have N:N relationship. Every ontology node may correspond to several reusable components, and every reusable component may correspond to several ontology nodes.

Even though function description of customer requirement is the same, the attributes are often different, so it is difficult for each software component to completely meet different requirements of different ontology node. We need to calculate the matching degree between software component and ontology node. We consider the matching degree between a concept of the ontology node C and a reusable software component S to be a number P (C, S) between 0 and 1, with 0 representing unmatched and 1 representing completely matched. The formula is: P (C, S) = f (S) / f (C), with f (S) representing the matched functional points number of the current reusable software component, and f (C) representing the total functional points number of the current ontology node.

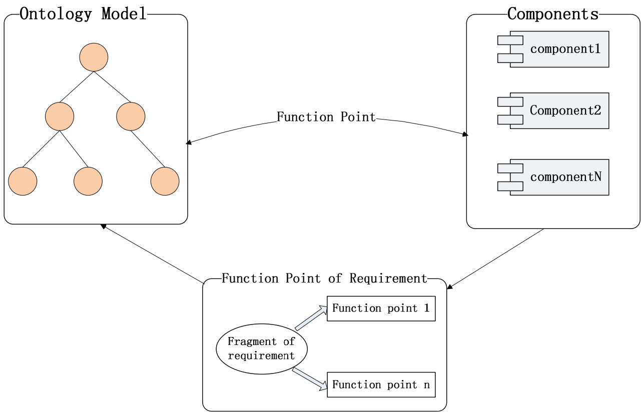


Figure 4. The mapping between ontology node and reusable components.

Experimental Design: The experiment is divided into two parts. The first part is to construct the application domain ontology model and the mapping between the application domain ontology nodes and the reusable software components. The data source is the application domain requirement specification and several sets of reusable software components.

Firstly, we need to construct the *application domain ontology* model. We adopt the ontological concept to divide the application domain requirement specification into different application fragments. For example, the *general equipment application* may contain four fragments: equipment, purchase, storage and organization. Every fragment of the requirement contains the whole functional points and input/output and then we will construct the ontology on the application domain by processing different fragments describing the application domain and by mining them into common vocabularies as domain-specific concepts. So every application domain ontology node has several corresponding functional points and input/output. For example, equipment maintenance node may contain new, updating and delete three functional points.

Secondly, we construct the mapping between application domain ontology nodes and reusable software components based on the functional points. We will match the functional points and input/output between every node and the reusable components by traversing the application domain ontology and then calculating P (C, S). If the number P(C, S) is greater than 0, the reusable component satisfies matching conditions. Because the reusable components are limited, we cannot guarantee all of the application domain ontology nodes have corresponding components that can be matched. Those nodes will be matched later during the software development and then we can put those components into the reusable components libraries.

The second part is to realize the reuse of the software components based on the ontology mapping technology and the result of the first part. The data source is the *sub-application domain* requirement specification. For example *engine equipment* is a sub-application domain of general equipment.

Firstly, we need to construct the sub-application domain ontology model. We can accomplish this step according to the first step of the first part.

Secondly, we need to realize the mapping between sub-application domain ontology and application domain ontology. We can calculate the mapping index according to the MAMA approach. The final result shows that 45 nodes are equal, 9 nodes are similar and 3 nodes are not similar. According to the software components reuse approach based on the ontology mapping, it is highly likely that the reusable components that match the 45 application domain ontology nodes also match the 45 sub-application domain ontology nodes one by one.

Finally, we will check the mapping result between the sub-application domain ontology nodes and the reusable software components by the functional points. We can calculate the number P(C, S) between sub-application domain ontology nodes and the reusable components according to the second step of the first part, and then calculate r1, defined as the ratio of reused components over total number of components, to measure the rate of component reuse for the sub-application domain. For those components that are not directly reusable because of different attributes, but with high P(C, S) value, we can adjust the P(C, S) value according to results of the previous step, which was calculated by using MAMA to map the attribute concepts. Then we can calculate r2, the revised rate of component reuse for the sub-application domain. To compare r1 and r2, the results are shown in Figure 5.

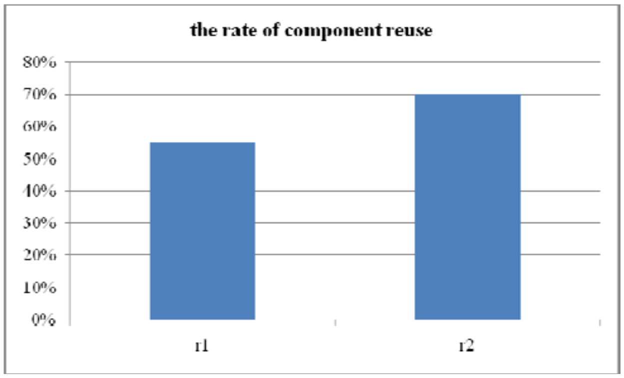


Figure 5. Improvement in component reuse rate.

Through this experiment, we conclude that we can increase the percentage of the components reuse by matching attribute names through ontology mapping. In other words, we can systematically change the name of variables in a program so that it can be reused in another sub-application domain.

5. Conclusion

This paper presented a novel approach for ontology mapping. It uses many existing and/or improved indexes and a novel way to combine them. Experimental results demonstrate the effectiveness of the approach approach. Future research topics include the continuous improvement of the various indexes and the definition of a merging approach.

As a practical application of this methodology, the software component reuse approach based on ontology mapping emphasizes the application of the ontology mapping technology and the reuse of software components. In software engineering, ontology and ontology mapping technology can be very useful to promote the sharing and reuse of the domain knowledge. Through the mapping between the ontology nodes and reusable software components, we can achieve the reuse of the software assets from the ontology model to the software components. Further research topics include the refinement of mapping techniques between the ontology nodes and software components.

References

[1] Thomas C. Jepsen, "Just What Is An Ontology, Anyway?," It Professional, Vol. 11, No. 5, Pp. 22-27, Sep./Oct. 2009, Doi:10.1109/Mitp.2009.105

[2] M. Shamsfard, "The State Of The Art In Ontology Learning : A Framework For Comparison," Computer Engineering, 2007, Pp. 1-28

[3] L. Zhou, "Ontology Learning: State Of The Art And Open Issues," Information Technology And Management, Vol. 8, Mar. 2007, Pp. 241-252.

[4] H. Bao, H. Liu, J. Yu, And H. Xu, "An Ontology-Based Semantic Integration For Digital," Architecture, 2005, Pp. 626 – 631

[5] G. Li, Z. Luo, And J. Shao, "Multi-Mapping Based Ontology Merging System Design," Framework, 2010, Pp. 5-11.

[6] M. Mao, Y. Peng, And M. Spring, "Ontology Mapping: As A Binary Classification Problem," 2008 Fourth International Conference On Semantics, Knowledge And Grid, Dec. 2008, Pp. 20-25.

[7] C. Zheng, Y.-Ping Shen, And L.I.N. Mei, "Ontology Mapping Based On Structures And Instances," Machine Learning, 2010, Pp. 11-14.

[8] L. Guan-Yu And L. Shu-Peng, "Formal Concept Analysis Based Ontology Merging Method 2 3," Science And Technology, 2010, Pp. 279-282.

[9] S. Li, H. Fan, Y. Wang, And L. Hong, "A Model And Approach For Heterogeneous Ontology Automatic Merging," 2009 International Joint Conference On Computational Sciences And Optimization, Apr. 2009, Pp. 214-217.

[10] H. Xia, Z. Li, H. Wang, And Y. Gu, "A Lightweight Method Of Web Service Ontology Merging Based On Concept Lattice," The 2nd Ieee Asia-Pacific Service Computing Conference (Apscc 2007), Dec. 2007, Pp. 304-311.

[11] Studer R, Benjamins Vr, Fensel D, "Knowledge Engineering: Principles And Methods", Ieee Transactions On Data And Knowledge Engineering, 25(1-2): 161- 199, 1998

[12] Wu Ya-Juan Lang Ji-Sheng Shang Fu-Hua, "A Similarity-Based Approach For Ontology Mapping", Proceedings Of 2009 4th International Conference On Computer Science & Education

[13] Gruber, T.R, Translation Approach To Portable Ontology Specification, Knowledge Acquisition, Vol. 5, Pp. 199-220, 1993

[14] Cheng-Yong, Huang-He, Qiu-Lirong, Shi-Zhongzhi. Similarity-Based Dynamic Multi-Dimension Concept Mapping Algorithm. Mini- Micro Systems. 2006.6.

[15] Nie Zhao-Hui, Wang Yin–Ling ,Research On Attribute Mapping Between Similarity Ontology. Computer Simulation. 2006.3

[16] Wordnet: http://wordnet.princeton.edu/

[17] Motoshi Saeki, "Ontology-Based Software Development Techniques", ERCIM News No.58, 2004.7

[18] Waralak V. Siricharoen. Ontologies and object models in object oriented software engineering. International Journal of Computer Science,33(1):1–6, 2007

[19] Haruhiko Kaiya, Motoshi Saeki, Ontology Based Requirements Analysis: Lightweight Semantic Processing Approach, Pp.1-8, 2005

Improving Text Retrieval Accuracy Using a Graph of Terms

Fabio Clarizia, Francesco Colace, Luca Greco, Massimo De Santo, Paolo Napoletano
Department of Electronics Engineering and Computer Engineering
University of Salerno, Fisciano, SA, 84084

***Abstract*—It has been demonstrated that a way to increase the number of relevant documents returned by an informational query performed on a Web repository is to expand the original query with additional knowledge, for instance coded through other topic-related terms.**

In this paper we propose a new technique to build automatically, through the probabilistic topic model and given a small set of documents on a topic, the expansion of a query based on a mixed *Graph of Terms* ($m\mathcal{GT}$) representation composed of two levels: the *conceptual level*, a set of interconnected terms representing concepts (undirected edges), and the *word level* composed of the cloud of interconnected words specifying a concept (directed edges). A $m\mathcal{GT}$ can be automatically learnt from a small set of documents through two learning stages and thanks to the probabilistic topic model. We have evaluated the performance through a comparison between our searching methodology and a classic one which considers the query expansion formed of only the list of concepts and words composing the graph and so where relations have not been considered. The results obtained show that our system, independently of the topic, is able to retrieve more relevant web pages.

***Index Terms*—Text retrieval, query formulation, term extraction, probabilistic topic model, relevance feedback.**

I. Introduction

The idea of taking advantage of additional knowledge, represented through a query expansion technique, to retrieve relevant web pages in informational querying, has been largely treated in the literature, where manual, interactive and automatic techniques have been proposed [1][2]. All existing techniques have been demonstrated to be more or less effective, with the interactive and hybrid ones being the best, and in turn to provide better results in terms of relevance.

The idea behind these techniques is that it may be sufficient to represent "the meaning" of what the user has in mind when performing a web search, or in other terms "the main concept" (or a set of concepts) of the preferred topic in which the user is interested, and so extending the original query with such a-priori knowledge. Following this way it is likely that web pages containing valuable, detailed information on that topic, will be ranked higher. Now a question arises: *how can I learn "the main concept" of a preferred topic?*

In the case of interactive environment, several methods answer this question. Some solutions rely on the interaction between the user and a "Collection Dependent Knowledge Structures", that is mostly represented by structured information stored in a database (mainly used in the case of e-commerce platform). Here "the main concept" can be better specified by choosing the appropriate metadata. Another set of solutions rely on the interaction with a "Collection Independent Knowledge Structures", that is mostly represented by domain vocabulary or glossary (mostly used in controlled domain-dependent environments). The last set of solutions rely on the interaction with the "Search Results" returned by the system after the initial query. In this case the system presents to the user "the main concept" coded through a list of terms based on their occurrences in an identified set of documents. Then the user feeds back his or her choice of terms. The document set on which this analysis is based may either be simply a set retrieved in the usual fashion (and chosen or accepted by the user as a suitable set for this purpose), or it may consist of documents individually selected as relevant by the user. Due to the fact that they best meet user requirements, such documents carry implicitly "the meaning" and so "the main concept".

We argue that, independently of how the list of terms has been extracted, the performance of the last methods can be improved if we use, as *vector of features*, a more complex representation instead of a simple list of words. Technically speaking, when you input an informational query, containing such additional knowledge, and search in a web repository, you are performing a sort of binary supervised text categorization task where the *vector of features* can be considered an explicit *profile* (or prototypical document) of the category [3].

In this proposal we argue that the *vector of features*, that we call next mixed *Graph of Terms*, can be automatically extracted from a set of documents following a kind of *term clustering* technique weighted by the probabilistic topic model. The graph learning procedure is composed of two stages and leads us to a two level representation. Firstly, we group terms with a high degree of pairwise semantic relatedness so obtaining several groups, each of them represented by a cloud of *words* and their respective centroids that we call *concepts*. In this way we obtain the lowest level, namely the *words level*. Later, we compute the second level, namely the *conceptual level*, by inferring semantic relatedness between centroids, and so *concepts*.

Such a structure, even if the size of the repository from which it is learnt is small, has demonstrated document clustering properties. In fact, this technique when employed for the classification of the ModApte split of the Reuters-21578 dataset where 1% of the training set for each category has been used, has demonstrated quite a good accuracy, evaluated through the F_1 measure[1]. The fact that the size of the repository, needed to build a suitable mixed *Graph of Terms* for filtering relevant information, can be small suggests that our approach can be employed to a real interactive query expansion process. In fact, it has been found that only 100 documents can be hand-labeled in 90 minutes and in this case the accuracy of classifiers (amongst which we find Support Vector Machine based methods), learnt from this reduced training set, could be around 30%. Consequently, most users of a practical system do not want to do labeling tasks for a long time only to obtain a better level of accuracy, they obviously prefer algorithms that have high accuracy, but do not require a large amount of manual labeling tasks [4][5].

The evaluation of the method has been conducted on a web repository collected at University of X by crawling a huge number of web pages from the website thomasnet.com. The results obtained, independently of the context, show that our system is able to retrieve more relevant web pages.

II. Graph definition

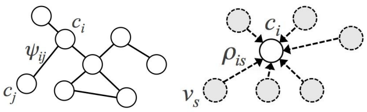
A mixed *Graph of Terms* ($m\mathcal{GT}$) is a hierarchical structure composed of two levels of information (see Fig. 2). One is the conceptual level that is represented by a set of interconnected terms representing concepts, namely a *Graph of Concepts* (see Fig. 1(a)), while the lowest level is composed by the cloud of words which makes a term be a concept (see Fig. 1(b)). More formally, let us define a *Graph of Concepts* as a triple $\mathcal{G}_\mathcal{C} = \langle N, E, C \rangle$ where N is a finite set of nodes, E is a set of edges weighted by ψ_{ij} on N, such that $\langle N, E \rangle$ is an a-directed graph (see Fig. 1(a)), and C is a finite set of concepts, such that for any node $n_i \in N$ there is one and only one concept $c_i \in C$. The weight ψ_{ij} can be considered as the degree of semantic correlation between two concepts c_i *is-related*$_{\psi_{ij}}$*-to* c_j and it can be considered as a probability: $\psi_{ij} = P(c_i, c_j)$. The probability of $\mathcal{G}_\mathcal{C}$ given a parameter τ can be written as the joint probability between all the concepts. By following the theory on the factorisation of undirected graph, we can consider such a joint probability as a product of functions where each of this can be considered as the weight ψ_{ij}. Then we have

$$P(\mathcal{G}_\mathcal{C}|\tau) = P(c_1, \cdots, c_H|\tau) = \frac{1}{Z} \prod_{(i,j)\in E_\tau} \psi_{ij}, \quad (1)$$

where H is the number of concepts, $Z = \sum_{\mathcal{G}_\mathcal{C}} \prod_{(i,j)\in E_\tau} \psi_{ij}$ is a normalisation constant and the parameter τ can be used to modulate the number of edges of the graph.

Each concept c_i can be defined as a rooted graph of words v_s and a set of links weighted by ρ_{is} (see Fig. 1(b)).

[1]The authors have already demonstrated such accuracy by performing a comparison between the proposed method and the SVM. Due to the blind review such a work can not be cited.



(a) (b)

Fig. 1. 1(a) Theoretical representation of a Graph of Concepts. 1(b) Graphical representation of a Concept.

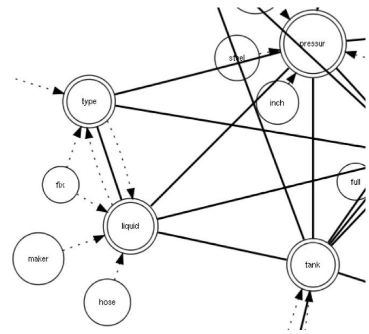


Fig. 2. Part of the mixed *Graph of Terms* learnt by a set of documents on the topic *Storage Tanks*. The double circle stands for a concept while the single stands for a words.

The weight ρ_{is} can measure how far a word is related to a concept, or in other words how much we need such a word to specify that concept. We can consider such a weight as a probability: $\rho_{is} = P(c_i|v_s)$. The probability of the concept given a parameter μ, which we call c_i, is defined as the factorisation of ρ_{is}

$$P(c_i|\{v_1, \cdots, v_{V_\mu}\}) = \frac{1}{Z_C} \prod_{s\in S_\mu} \rho_{is}, \quad (2)$$

where $Z_C = \sum_C \prod_{s\in S_\mu} \rho_{is}$ is a normalisation constant, V_μ is the number of words defining the concept, such a number depending on the parameter μ.

A. Graph learning procedure

A $m\mathcal{GT}$ is defined by the probability $P(\mathcal{G}_\mathcal{C}|\tau)$, which defines a graph of connected H concepts and the number of edges depends on τ, and by H probabilities of the concepts, $P(c_i|\{v_1, \cdots, v_{V_{\mu_i}}\})$, where the number of edges depends on μ_i. Once each ψ_{ij} and ρ_{is} is known (*Relations Learning*), to determine the final graph we need to compute the appropriate set of parameters $\mathbf{\Lambda} = (H, \tau, \mu_1, \cdots, \mu_H)$ (*Parameters Learning*). Note that a concept is lexically identified by a word of a vocabulary, and it can be represented by a set of connected words (see Fig. 1(b)). Formally, a *word* is an item of a vocabulary indexed by $\{1, \cdots, V\}$, each word using unit-basis vectors that have a single component equal to one and all other components equal to zero. Thus, the vth word in the vocabulary is represented by a V-vector w such that $w^p = 1$ and $w^q = 0$ for $p \neq q$. A document is a sequence of L words denoted by $\mathbf{w} = (w_1, w_2, \cdots, w_L)$, where w_n is the nth word in the sequence. A corpus is a collection of M documents denoted by $\mathcal{D} = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_M\}$.

Conceptual Level		
Concept i	Concept j	Relation Factor (ψ_{ij})
tank	roof	4,0
tank	water	3,37246
tank	storag	3,33194
...	...	...
liquid	type	3,43828
liquid	pressur	3,07028
...	...	...

Word Level		
Concept i	Word s	Relation Factor (ρ_{is})
tank	larg	2,0
tank	construct	1,6
...	...	...
liquid	maker	1,11673
liquid	fix	1
...	...	...

TABLE I
AN EXAMPLE OF A $m\mathcal{GT}$ FOR THE TOPIC *Storage Tank* EXTRACTED FROM A SUBSET OF DOCUMENTS.

1) Relations Learning: Due to the fact that each concept is lexically represented by a word, then we have that $\rho_{is} = P(c_i|v_s) = P(v_i|v_s)$, where the concept c_i is lexically identified by v_i, and $\psi_{ij} = P(c_i, c_j) = P(v_i, v_j)$ where the concept c_i and c_j are lexically identified by v_i and v_j respectively. As a result, all the relations of the $m\mathcal{GT}$ can be represented by $P(v_i, v_j)\ \forall i, j \in V$ which can be considered as a *word association problem*[2] and so it can be solved through a smoothed version of the generative model introduced in [6] called latent Dirichlet allocation, which makes use of Gibbs sampling [7].

2) Parameters Learning : Once each ψ_{ij} and ρ_{is} is known, we have to find a value for the parameter H, which establishes the number of concepts, a value for τ and finally values for μ_i ,$\forall i \in [1, \cdots, H]$. As a consequence, we have $H + 2$ parameters which modulate the shape of the graph. If we let the parameters assuming different values, we can observe different graph $m\mathcal{GT}_t$ for each set of parameters, $\mathbf{\Lambda}_t = (H, \tau, \mu_1, \cdots, \mu_H)_t$ extracted from the same set of documents, where t is representative of different parameter values. In this case we need a criterion to choose which graph is the best for our scope, e.g., we could be interested in those $m\mathcal{GT}$s which better represent each document of the feeding repository. A way of saying that a $m\mathcal{GT}$, given the parameters, is the best possible for that set of documents is to demonstrate that it produces the maximum score attainable for each of the documents when the same graph is used as a knowledge base for querying in a set containing just those documents which have fed the $m\mathcal{GT}$ builder. For this purpose let us suppose of indexing the corpus $\mathcal{D}$ following the *term frequency-inverse document frequency* (*tf-idf*) model, and let $m\mathcal{GT}_t$ be the t-th possible graph built from that repository with the set of parameters $\mathbf{\Lambda}_t$. Let us suppose of using a text-based search engine, in this case Lucene (http://lucene.apache.org), which has the ability to assign a rank or an order to documents $\mathbf{w}$ that match a query $\mathbf{q}$ following the *cosine similarity* model $\mathcal{S}(\mathbf{q}, \mathbf{w})$.

Lucene embeds very efficient searching and scoring algorithms based on both the *tf-idf* model and the Boolean model. In order to perform the $m\mathcal{GT}$-based search, we have customised Lucene's querying mechanism. Since our graphs are represented as pairs of related words, see Table I, where the relationship strength is described by a real value (namely ψ_{ij} and ρ_{is}, the *Relation factors*), including such a graph in the searching process would basically result in performing a query like $((tank$ AND $roof)^{4.0})$ OR $((tank$ AND $larg)^{2.0})$..., that means searching the pair of words *tank* AND *roof* with a boost factor of 4.0 OR the pair of words *tank* AND *larg* with a boost factor of 2.0 and so on. Table I shows an example of a graphical representation for the topic *Storage Tank*, where ψ_{ij} and ρ_{is} are named Relation factors and later it will be clear why such values do not belong to $[0, 1]$. The default boost value inserted by the Lucene Boolean model is 1, and it is presumably assigned to the words that compose the original query. A way to make the contribution from the terms of the graph be more important is to set a boost higher than 1. Moreover, due to the fact that we consider relations between concepts more important than relations between concepts and words (this is a consequence of the hierarchical structure of the graph), therefore we have shifted the original set of ψ_{ij} and ρ_{is} from $[0, 1]$ to the sets $[3, 4]$ and $[1, 2]$ respectively. The score function, based on vector cosine similarity, used to perform the query is the following $\mathcal{S}(\mathbf{q}, \mathbf{w}) = \mathcal{C}(\mathbf{q}) \cdot \mathcal{N}_\mathbf{q}(\mathbf{q}) \cdot \sum_{q \in \mathbf{q}} (tf_{q,\mathbf{w}} \cdot idf_q^2 \cdot \mathcal{B}_q \cdot \mathcal{N}(q, \mathbf{q}))$ where $tf_{q,\mathbf{w}}$ correlates to the term's frequency, defined as the number of times term q appears in the currently scored document $\mathbf{w}$. Documents that have more occurrences of a given term receive a higher score. The default computation for idf_q is, $idf_q = 1 + \log\left(\frac{M}{df_q}\right)$, with M being the total number of documents and df_q the total number of documents in which the term q appears. This means that rarer terms give a higher contribution to the total score. $\mathcal{C}(\mathbf{q})$ is a score factor based on how many of the query's terms are found in the specified document. Typically, a document that contains more of the query's terms will receive a higher score than another document with fewer query terms. $\mathcal{N}_\mathbf{q}(q, \mathbf{q})$ and $\mathcal{N}_\mathbf{q}(\mathbf{q})$ are normalising factors used to make scores between queries comparable, therefore we will not explain them. $\mathcal{B}_q$ is a search boost factor of term q in the query $\mathbf{q}$ as specified in the query text (see query syntax above). Due to the boost factor we can assign a greater weight to a term of the query that are more relevant for that topic. As a consequence, each $m\mathcal{GT}$ can be represented by the Lucene boolean syntax, which alternatively corresponds to two vectors, one representing all the pairs $\mathbf{q}_t = \{q_1, \cdots, q_U\}_t$, and one representing each relation factor, that is the Lucene boost, $\mathcal{B}_{\mathbf{q}_t} = \{\mathcal{B}_{q_1}, \cdots, \mathcal{B}_{q_U}\}_t$. By performing a Lucene search query that uses the current graph $m\mathcal{GT}$ on the same repository $\mathcal{D}$, we obtain a score for each document $\mathbf{w}_i$ and then we have $\mathbf{S}_t = \{\mathcal{S}(\mathbf{q}_t, \mathbf{w}_1), \cdots, \mathcal{S}(\mathbf{q}_t, \mathbf{w}_M)\}_t$, where each of them depends on the set $\mathbf{\Lambda}_t$. To compute the best value of $\mathbf{\Lambda}$ we can maximise the score value for each documents, which means that we are looking for the graph which best describes each document of the

[2]The authors have reported the mathematical formulation that brings from the LDA to $P(v_i, v_j)$ in a paper that can not be cited due to the blind review.

repository from which it has been learnt. It should be noted that such an optimisation maximises at same time all M elements of $\mathbf{S}_t$. Alternatively, in order to reduce the number of the objectives to being optimised, we can contemporary maximise the mean value of the scores and minimise their standard deviation, which turns a multi-objectives problem into a two-objectives one. Additionally, we can reformulate the latter problem by means of a linear combination of its objectives, thus obtaining a single objective function, i.e., *Fitness* ($\mathcal{F}$), which depends on $\mathbf{\Lambda}_t$, $\mathcal{F}(\mathbf{\Lambda}_t) = E_m[\mathcal{S}(\mathbf{q}_t, \mathbf{w}_m)] - \sigma_m[\mathcal{S}(\mathbf{q}_t, \mathbf{w}_m)]$, where E_m is the mean value of all element of $\mathbf{S}_t$ and σ_m being the standard deviation. By summing up, the parameters learning procedure is represented as follows, $\mathbf{\Lambda}^* = \underset{t}{\operatorname{argmax}}\{\mathcal{F}(\mathbf{\Lambda}_t)\}$. The fitness function depends on $H + 2$ parameters, hence the space of possible solutions could grow exponentially. Due to the fact that we would not have small or too big graph (medium size of a graph is more handle during the querying process), we suppose that the number H of concepts can vary from a minimum of 5 to a maximum of 20, and considering that it is an integer number we conclude that the number of possible values for H is 15. Considering that ψ_{ij} and ρ_{is} are probabilities, and so real value such as $\tau \in [0, 1]$ and each $\mu_i \in [0, 1]$. It means that if we use a step of 1% to explore the entire set $[0, 1]$, then we have 100 possible values for τ and 100 for each μ_i, which makes $100 \times 100 \times H \times 15$ possible values of Λ, that is $750,000$ for $H = 5$ and $3,000,000$ for $H = 20$. To limit such a space we can reduce the numbers of parameters, for instance we can consider $\mu_i = \mu$, $\forall i \in [1, \cdots, H]$ and so obtaining $150,000$, independently from H, possible values of Λ. Furthermore we have reduced the remaining space of solution by applying a clustering methods, that is the *K-means* algorithm, to all ψ_{ij} and ρ_{is} values. Following this approach and choosing for instance 10 classes of values for τ and μ, we obtain that the space of possible Λ is $10 \times 10 \times 15$, that is $1,500$. As a consequence, the optimum solution can be exactly obtained after the exploration of the entire space of solutions. Furthermore, it is quite important to make clear that the mixed *Graph of Terms* can not be considered as a co-occurrence matrix, that have already demonstrated limitations in query expansion technique [8][1][2]. In fact, the core of the graph is the probability $P(v_i, v_j)$, which we compute through the probabilistic topic model and particularly thanks to the word association problem. In the topic model, the word association can be considered as a problem of prediction: given that a cue is presented, which new words might occur next in that context? It means that the model does not take into account the fact that two words occur in the same document, but that they occur in the same document when a specific topic is assigned to the document itself [7], in fact $P(v_i|v_j)$ is the result of a sum over all the topics.

III. Experimental evaluation

The evaluation of the method has been conducted on a web repository collected at University of X by crawling 154,243 web pages for a total of 3.0 GB by using the website ThomasNet (http://www.thomasnet.com) as index of URLs, the reference language being English[3]. ThomasNet, known as the "big green books" and "Thomas Registry", is a multi-volume directory of industrial product information covering 650,000 distributors, manufacturers and service companies within 67,000-plus industrial categories. We have downloaded webpages from the company websites related to 50 categories of products, randomly chosen from the ThomasNet directory. Note that even if the presence or absence of categories in the repository depends on the random choices made during the crawling stage, it could happen that webpages from some business companies cover categories that are different from those randomly chosen. This means that the repository is not to be considered as representative of a low number of categories but as a reasonable collection of hundreds of categories. Once the repository has been collected, a human has explored it performing key words based queries and so roughly assessing the distribution of the categories in the repository following this criterion: a topic is considered to be well represented if the repository contains about 300 webpages dealing with that topic. At the end of this process we have obtained 9 categories: 1. *Lubricant*, 2. *Pump*, 3. *Adhesive*, 4. *Generator*, 5. *Transformers*, 6. *Inverter*, 7. *Valve*, 8. *LAN Cables*, 9. *Storage Tanks*. For the performance evaluations of the proposed technique we have used an open source text-based web search engine, Lucene from Apache project, and so we have indexed the repository by using the *term frequency-inverse document frequency* (*tf-idf*) model. This engine assigns a rank or an order to documents $\mathbf{w}$ that match a query $\mathbf{q}$ following the *cosine similarity* model $\mathcal{S}(\mathbf{q}, \mathbf{w})$, as reported above.

Topic	Query	$\sharp$ of terms	$\sharp$ of pairs
1	Lubricant	54	69
2	Pump	63	70
3	Adhesive	45	67
4	Generator	58	68
5	Transformers	67	82
6	Inverter	62	84
7	Valve	47	66
8	LAN Cables	69	85
9	Storage Tanks	51	66
Average Size		**57**	**73**

TABLE II
Number of terms and pairs for each $m\mathcal{GT}$.

Note that the $m\mathcal{GT}$ is different from a simple list of key words because of the presence of two features: the relations between terms and the hierarchical differentiation between simple words and concepts. To demonstrate the discriminative property of such features we have to prove that the results obtained performing the proposed approach are significantly better than the results obtained by performing the same queries composed of the simple list of words extracted from the $m\mathcal{GT}$. Summarizing we have compared two different query expansion searching methodologies:

1) based on the mixed *Graph of Terms* (we refer to the

[3]The repository will be public on our website to allow further investigations from other researchers.

Topic		eR	eAP	eR pr	eP5	eP10	eP15	eP20	eP30	eP100
1	$m\mathcal{GT}$	64	0.59	0.70	1.00	0.78	0.71	0.74	0.57	0.55
	$\mathcal{WL}$	64	0.33	0.40	0.75	0.67	0.64	0.74	0.66	0.35
2	$m\mathcal{GT}$	76	0.56	0.60	1.00	1.00	0.86	0.74	0.69	0.63
	$\mathcal{WL}$	76	0.25	0.40	0.75	0.67	0.71	0.63	0.55	0.37
3	$m\mathcal{GT}$	75	0.74	0.72	1.00	1.00	1.00	1.00	0.79	0.67
	$\mathcal{WL}$	75	0.37	0.44	0.50	0.78	0.86	0.89	0.62	0.44
4	$m\mathcal{GT}$	73	0.50	0.59	1.00	0.67	0.79	0.84	0.86	0.49
	$\mathcal{WL}$	73	0.68	0.66	0.75	0.89	0.93	0.95	0.83	0.62
5	$m\mathcal{GT}$	49	0.48	0.47	1.00	0.89	0.93	0.84	0.55	0.36
	$\mathcal{WL}$	49	0.30	0.43	1.00	0.56	0.36	0.37	0.38	0.31
6	$m\mathcal{GT}$	38	0.59	0.60	0.75	0.78	0.79	0.84	0.72	0.33
	$\mathcal{WL}$	38	0.67	0.66	0.75	0.89	0.93	0.89	0.72	0.35
7	$m\mathcal{GT}$	99	0.62	0.76	1.00	0.89	0.79	0.84	0.83	0.76
	$\mathcal{WL}$	99	0.40	0.57	1.00	0.67	0.64	0.63	0.62	0.57
8	$m\mathcal{GT}$	28	0.32	0.32	0.50	0.56	0.43	0.32	0.35	0.24
	$\mathcal{WL}$	28	0.47	0.40	1.00	0.56	0.57	0.47	0.38	0.23
9	$m\mathcal{GT}$	45	0.74	0.67	1.00	1.00	0.93	0.89	0.79	0.43
	$\mathcal{WL}$	45	0.15	0.16	0.75	0.56	0.43	0.37	0.24	0.16

TABLE III
INDICES OF PERFORMANCE ON DIFFERENT TOPICS.

Topic/Algorithm		Rel	CG	DCG	IDCG	nDCG
1	$m\mathcal{GT}$	55	80	25.54	30.03	0.85
	$\mathcal{WL}$	35	42	15.50	17.42	0.89
2	$m\mathcal{GT}$	62	81	24.26	28.58	0.85
	$\mathcal{WL}$	37	55	17.05	23.69	0.72
3	$m\mathcal{GT}$	67	76	18.57	24.30	0.76
	$\mathcal{WL}$	44	50	12.05	18.43	0.65
4	$m\mathcal{GT}$	48	63	19.33	24.27	0.80
	$\mathcal{WL}$	61	74	21.35	25.50	0.84
5	$m\mathcal{GT}$	36	60	23.71	26.18	0.90
	$\mathcal{WL}$	31	44	16.33	20.07	0.81
6	$m\mathcal{GT}$	33	39	10.70	16.37	0.65
	$\mathcal{WL}$	35	41	11.07	16.75	0.66
7	$m\mathcal{GT}$	76	98	25.41	32.21	0.79
	$\mathcal{WL}$	57	85	23.75	31.62	0.75
8	$m\mathcal{GT}$	24	32	11.82	15.83	0.75
	$\mathcal{WL}$	27	35	12.37	16.46	0.75
9	$m\mathcal{GT}$	43	60	20.98	24.34	0.86
	$\mathcal{WL}$	16	20	8.78	11.23	0.78

TABLE V
CUMULATIVE GAIN (CG), DISCOUNTED CUMULATIVE GAIN (DCG), NORMALIZED DISCOUNTED CUMULATIVE GAIN (NDCG)

run	eMAP	eRprec	eP5	eP10	eP15	eP20	eP30	eP100
$m\mathcal{GT}$	0.57	0.60	0.92	0.84	0.80	0.78	0.69	0.50
$\mathcal{WL}$	0.40	0.46	0.81	0.69	0.68	0.66	0.56	0.38

TABLE IV
AVERAGE VALUES OF PERFORMANCE

results obtained with this technology as $m\mathcal{GT}$);

2) based on only the words from the $m\mathcal{GT}$ without relations (named Words List *WL*).

For each scenario, each person has identified a set of documents from which we extracted the $m\mathcal{GT}$ for each topic, obtaining 9 small repositories each composed of 3 documents, which corresponds to about 1% of the relevant set of documents for each topic (those are considered as training sets). The structure of each $m\mathcal{GT}$ obtained for each topic, that is the number of terms and the number of pairs, is reported in table II. The average size of the list of terms and the list of pairs is 57 and 73 respectively. These structures have been obtained thanks to the learning procedure and best value of concepts has been chosen between all possible $H \in [5, 15]$. We have calculated the average number of concepts for each structure and we have obtained $H = 13$. It means that each concept has been specified by $57/13 \approx 4$ words belonging to the *word level*.

For each context, we have asked different humans (and so 9 in total) to assign graded judgments of relevance to the first 100 pages obtained by querying both search engines for each topic. Due to the fact that the number of evaluations for each topic, and so the number of topics itself, is small, humans have evaluated, in contrast to the Minimum Test Collection method [9], all the results obtained. The assessment is based on three levels of relevance: *high relevant*, *relevant* and *not relevant*. To get a measure in some way ”objective” of the results proposed by the two search engines, we have prepared an *xml* based schema, to code the user intentions, and to avoid cases of ambiguity. For each topic we have defined a schema as in the following example:

```
<topic number=”3” type=”faceted” >
  <query> adhesive</query>
    <description>
      I am looking for information on adhesive.
    </description>
      <subtopic number=”1” type=”inf” >
        I am looking for web pages containing
        datasheets of several adhesive types
      </subtopic>
      <subtopic number=”2” type=”inf” >
        I am looking for descriptions of adhesives
        as products
      </subtopic>
      <subtopic number=”3” type=”inf” >
        I am looking for a list of adhesive categories
        and their description
      </subtopic>
      <subtopic number=”4” type=”inf” >
        I am looking for information on adhesive
        usages
      </subtopic>
</topic>
```

This approach allows us to involve a small number of evaluators, as suggested by *TREC* [4], that defines the structure of the xml schema employed in these experiments. Measures of *Precision* and *Recall*, by using standard indicators provided by [10][11][12][13], have been calculated,

$$eAP = \frac{1}{ER} \sum_{i=1}^{k} \frac{x_i}{i} + \sum_{j>i} \frac{x_i x_j}{j} \tag{3}$$

$$ePrec@k = eP@k = \frac{1}{k} \sum_{i=1}^{k} x_i \tag{4}$$

$$ERprec = \frac{1}{ER} \sum_{i=1}^{ER} x_i, \; ER = \sum_{i=1}^{n} x_i \tag{5}$$

eAP indicates the average precision on a topic, x_i and x_j are Boolean indicators of relevance, k is the cardinality of the considered result set (k=100), ER is a subset of relevant documents [5]. The factor $ERprec$ is the precision at the level ER, that is when the $Precision \simeq Recall$ (or in other terms at the break-even point). The measure $eMAP$ is the average of all eAP over topics. The measure $eP@k$ is the precision at level k (for instance $eP5$ is the precision calculated by taking the top 5 results). In Tables III we find these measures for each topic while in table IV these measures across topics are shown. The overall behavior of

[4]The Text Retrieval Conference (TREC).

[5]The value of ER depends on the intersection of the result of both search engines

the $m\mathcal{GT}$ method is better than the $\mathcal{WL}$, especially for the topic 2, 3 and 7. In fact in these cases the proposed method have listed 62, 67 and 76 relevant or high relevant documents in the top 100, that is about 68% (see also the column Rel of the table V). However, in the case of topics 4, 6 and 8 the number of relevant documents is comparable between systems, with the percentage of relevant documents retrieved being about 30%, that is less than half of the worst value obtained for the topic 2. This suggests that the systems are comparable only if the total number of relevant documents returned by both systems less than 50%. This probably happens due to the fact that the documents feeding the graph builder have not covered, in terms of topics, all the examples present in the repository. Notwithstanding this, the most important fact is that, when the graph is added to the initial query, the web search engine shows better performances than the case of the simple word list. It should be also noticed that also the word list has been provided by our method.

In Table V further standard measures of performance are shown which take into account the quality of the results related to the position in which they are presented. We have mainly considered the *Cumulative Gain* (CG) and the *Discounted Cumulative Gain* (DCG) and, to make comparable the results obtained by both the systems, we have also computed the *normalized Discontinued Cumulative Gain* ($nDCG$). This last measure is obtained thanks to the *Ideal DCG*, which is computed by placing relevant results on the top, that is $nDCG_x = \frac{DCG_x}{IDCG_x}$. We hve

$$CG_x = \sum_{i=1}^{n} rel_i,\ DCG_x = \sum_{i=1}^{n} \frac{2^{rel_i} - 1}{\log_2(1+i)}. \quad (6)$$

In these measures we have considered that $rel = 2$ when the document is *High Relevant*, $rel = 1$ when the document is *Relevant* and $rel = 0$ when the document is *Not Relevant*. As we can see, the results on the topic 4, 6 and 8 are the worst cases, while topics 2, 3, 5, 7, 9 are the best, as confirmed by previous discussions on table III. In the case of topic 1, even if the result is good, we obtained a value of $nDCG$ for the $\mathcal{WL}$ method that is higher.

IV. Conclusions

In this work we have demonstrated that a mixed *Graph of Terms* based on a hierarchical representation is capable of retrieving more relevant web pages in informational querying, even if the size of the feeding corpus is small.

The fact that the size of the repository, needed to build a suitable $m\mathcal{GT}$ for filtering relevant information, can be small, suggests that our approach can be employed in a kind of interactive query expansion process, where the user can initially perform a query composed of key terms, that he knows about the topic, and later can select relevant documents from the result set and so feed the $m\mathcal{GT}$ builder. At this point, the system can add the knowledge extracted from those documents suggested by the user, and the query can be performed again.

References

[1] E. N. Efthimiadis, "Query expansion," in *Annual Review of Information Systems and Technology*, M. E. Williams, Ed., 1996, pp. 121–187.

[2] P. R. Christopher D. Manning and H. Schtze, *Introduction to Information Retrieval*. Cambridge University, 2008.

[3] F. Sebastiani, "Machine learning in automated text categorization," *ACM Comput. Surv.*, vol. 34, pp. 1–47, March 2002. [Online]. Available: http://doi.acm.org/10.1145/505282.505283

[4] A. McCallum, K. Nigam, J. Rennie, and K. Seymore, "A machine learning approach to building domain-specific search engines," in *Proceedings of the 16th international joint conference on Artificial intelligence - Volume 2*. Morgan Kaufmann, 1999, pp. 662–667. [Online]. Available: http://portal.acm.org/citation.cfm?id=1624312.1624313

[5] Y. Ko and J. Seo, "Text classification from unlabeled documents with bootstrapping and feature projection techniques," *Inf. Process. Manage.*, vol. 45, pp. 70–83, January 2009. [Online]. Available: http://portal.acm.org/citation.cfm?id=1460927.1461016

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, no. 993–1022, 2003.

[7] T. L. Griffiths, M. Steyvers, and J. B. Tenenbaum, "Topics in semantic representation," *Psychological Review*, vol. 114, no. 2, pp. 211–244, 2007.

[8] H. J. Peat and P. Willett, "The limitations of term co-occurrence data for query expansion in document retrieval systems," *Journal of the American Society for Information Science*, vol. 42, pp. 378–383, January 1991.

[9] B. Carterette, J. Allan, and R. Sitaraman, "Minimal test collections for retrieval evaluation," in *29th International ACM SIGIR Conference on Research and development in information retrieval*, 2008.

[10] K. Jarvelin and J. Kekalainen, "Cumulated gain-based evaluation of ir techniques," *ACM Transactions on Information Systems*, vol. 20(4), no. 422–446, 2002.

[11] B. Croft, D. Metzler, and T. Strohman, "Addison wesley," *Search Engines: Information Retrieval in Practice*, no. 422–446, 2009.

[12] J. Allan, J. A. Aslam, B. Carterette, V. Pavlu, and E. Kanoulas, "Million query track 2008 overview," in *In 17th Text REtrieval Conference (TREC 2008) Proceedings*, 2008.

[13] C. L. A. Clarke, N. Craswell, and I. Soboroff, "Overview of the trec 2009 web track," in *In 18th Text REtrieval Conference (TREC 2009) Proceedings*, 2009.

Handhold Object Detection and Event Analysis Using Visual Interaction Clues

Jun-Wei Hsieh, Chia-Lung Lin
Dept. of C. S. and Engineering, NTOU
No.2, Beining Rd., Keelung 202, Taiwan
shieh@ntou.edu.tw

Jiun-Chen Cheng, Pin Wu, and Duan-Yu Chen
Dep. of Electrical Engineering, YZU
135 Yuan-Tung Road, Chung-Li 320, Taiwan

Abstract

This paper addresses the multiplicity problem of video events by dividing an event context to its intra-path context and inter-path context. Then, the novel event analysis method is applied for analyzing various interaction events (e.g. smoking, eating, and phoning) happening between hands and faces. This problem is very challenging since there is no prior knowledge (like shape, color, size, and texture) about the hand-hold objects. To address this problem, a novel ratio histogram is first proposed for finding important color bins for locating desired objects through a re-projection technique. Then, a code book method is then used for object tracking and feature extraction. Then, each action is defined as a temporally ordered set of repetitive symbols. In real cases, an event will have various representations under different lighting and weather conditions. To reflect the multiplicity of an event, we divide an event to two parts, i.e., the intra-path context and inter-path context. The first one models the dynamic properties of an event within a path and the second one is to measure the multiplicity of an event between paths. The intra-path context uses Markov chains to capture the repetitions of action primitives. The multiplicity of an event is modeled into the inter-path context using a weighted edit distance. A Bayesian inference scheme is then used to find the best set of event parameters directly from videos and classify events to different clusters. Experimental results show this scheme is effective to detect and analyze different daily events (e.g. smoking, eating, and phoning) even though variant colors and sizes on objects and dressing appearances are handled.

1. Introduction

Human action analysis [1]-[7] is an important task in various application domains like video surveillance, video retrieval, human-computer interaction systems, and so on. Characterization of human action is equivalent to dealing with a sequence of video frames that contain both spatial and temporal information. The challenge in human action analysis is how to properly characterize spatial-temporal information and then facilitate subsequent comparison/recognition tasks. To build the action primitives for event analysis, the event representation scheme can be divided to two categories, i.e., the model-based and feature-based ones. For the model-based scheme, some syntactic primitives are extracted for event representation. For example, Fengjun and Nevatia [1] used a posture-based scheme for converting frames to different action nets. Then, these action nets are classified using the Viterbi algorithm. For the feature-based scheme, some points of interest are detected and tracked for event classification. For example, Fathi and Mori [2] tracked corner features for obtaining their motion vectors and then classified these vectors to different event types. In [6], Laptev *et al*. detected space-time interest points (STIP) and obtained their motion flows to extract several key frames for event representation. Compared with the model-based method, the feature-based method usually does not require the background being still and thus can be applied to videos with larger changes in background. However, the success of this method strongly depends on a large set of well tracking points for building the connections between frames.

In addition to event features, another key problem in event analysis is how to model the temporal and spatial dynamics of events. For example, in [4], Laxton *et al*. used dynamic Bayesian network to model the interactions between objects. Then, the Viterbi-like algorithm is adopted for seeking the best path for event interpretation. In[3], Jojic *et al*. embedded a template matching technique in a transformed hidden Markov model (THMM) for event analysis. A serious problem related to HMMs involves how to specify and learn the HMM model structure. It is more challenging to analyze the events happening between two objects since the object interactions are too complicated for modeling. Filipovych and Ribeiro [5] used a probabilistic graphical model to recognize the primitive actor-object interactions like "grasping" or "touching" a fork (or a spoon).

This paper addresses the problem to analyze the events (e.g. smoking, eating, and phoning) caused by objects hold by hands and interacting between hands and faces. The detection task is challenging since there is no prior knowledge about the object's shape, texture and color. In addition, its visual features will change under different lighting and weather conditions. The major contribution of this paper is to present a novel histogram-based framework for detecting each hand-hold object from videos no matter what shape and color it has. In addition, a novel event system is proposed for analyzing various interaction events (like smoking, phoning, and drinking) happening between hands and faces directly from videos. Fig. 1 shows the flowchart of this system. First of all, the Adaboost scheme [9] is adopted for detecting various faces from videos. Then, a novel ratio histogram is proposed for comparing the color differences of faces during the interaction process. After that, significant color bins for the objects of interest can be extracted from the histogram. Then, each desired hand-hold object can be detected and located using a re-projection technique. After that, three visual features including the object size, the distance between the object and the mouth, and the smog density are extracted for event representation. Then, each event can be represented as a temporally ordered set of repetitive symbols (or a path). In real cases, the same event can be represented by different paths with different beginning states, repetitions, and ending states. To deal with the multiplicity of an event, a Bayesian framework is then proposed for analyzing a path not only from its inter-path context but also its intra-path context. The intra-path context uses Markov chains to capture the repetitions of action primitives and the transition probabilities between two action primitives. The multiplicity of an event is modeled into the inter-path context using a weighted edit distance. When the multiplicity is considered, different events can be more accurately analyzed even though there are different time scaling changes, initial and ending states, and conversion errors between them. Experimental results have proved the superiority of the proposed method in event analysis.

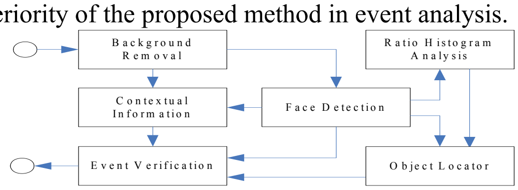


Fig. 1: Flowchart of the proposed system for analyzing events happening between faces and hands.

2. Hand-hold Object Detection

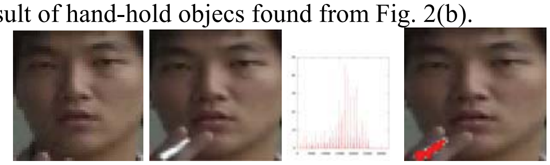
In what follows, a novel approach is first proposed for detecting various hand-hold objects for event analysis.

2.1. Detection of Objects on Face

To analyze the events caused by the interactions between faces and hands, the face region is an important cue for event analysis. The Adaboot scheme [9]is first used to detect all possible faces from videos. After that, the hold-hold object should be first detected. However, its visual features like size, shape, color, orientation, and texture are unknown. Thus, in what follows, a novel ratio histogram is proposed for detecting various hand-hold objects from a face. Let F_t denote the tth frame of a face F with the color histogram $H_t(i)$. Then, the rth order of the forward ratio histogram is defined as

$$RatioH_t^{r+}(i) = \frac{H_t(i)}{1+H_{t-r}(i)}. \qquad (1)$$

Fig. 2(c) shows the ratio histogram between (a) and (b). The missed colors between (a) and (b) can be identified by finding the bins whose responses are larger in $RatioH_t^{r+}(i)$. Let $T^H_{Ratio_t^r}$ be the average value of all bins in the ratio histogram. For a color bin k, if $RatioH_t^{r+}(k) > 1.5\,T^H_{Ratio_t^r}$, the kth color will be an important color for highlighting the object region. Then, with a re-projection scheme, the hand-hold object can be identified. (d) is the detection result of hand-hold objecs found from Fig. 2(b).



(a) (b) (c) (d)

Fig. 2: Object identification using color re-projection.

2.2. Object Tracking by Code Book

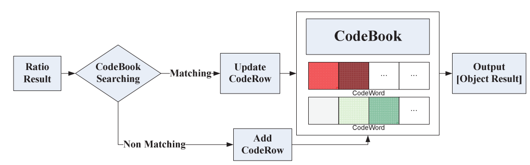


Fig. 3: Object tracking using a codebook technique.

(a) Object size (b) Distance (c) Smog condition

Fig. 4: Object size, the distance to the mouth, smog density are used for event representation.

Once the hand-hold object is detected, its positions should be tracked. To track it, the codebook scheme is modified for recording its visual properties including size, color, and position. The flowchart is shown in Fig. 3. When an object appears or disappear in the face, it will be checked whether it has been recorded in the codebook list. If yes, its current visual property will be recorded into the codebook list. If not, a new record will be created in the codebook. Like Fig. 4, the object size and the object distance to the mouth will be used to code each video frame to different symbols. In addition, two bounding boxes beside the face upon the shoulder are automatically set to calculate the smog density as the third feature for event representation. The smog density can be extracted using a subtraction technique [8].

3. Interaction Events between Hands and Face

To analyze the interactive behaviors between faces and hands, three features including object size, object distance to the mouth, and smog density are used to convert each frame to different action primitives. Details of the frame coding and event representation are discussed as here.

3.1. Frame Coding

Since three features are used, this paper uses three bins to convert each frame to a symbol m_i. Since the importance of each feature is different in event analysis, different weights are added to each bin. Let w_{b_k} denote the weight for the kth bit and $\oplus$ be the *X*-OR operator. Then, the distance between m_i and m_j can be measured as follows:

$$\beta[m_i, m_j] = \sum_{k=0}^{K} w_{b_k}(b_k^{m_i} \oplus b_k^{m_j}), \qquad (2)$$

where b_k^m is the kth bit of the symbol m. Fig. 5 shows different weights for symbols comparison. The values of w_{b_2}, w_{b_2}, and w_{b_0} are to set to 3, 2, 1, respectively. The bit weight w_{b_2} of the object size is highest, since it is more robust to identify cigarettes than other features. The bit weight w_{b_0} of smog density is lowest since it is easily disturbed by noise.

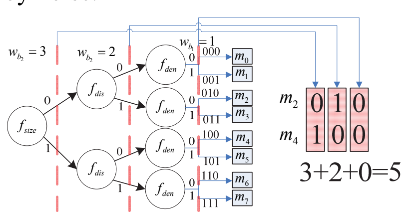


Fig. 5: Different weights for symbol comparisons.

3.2. Event Representation

Let M denote the set of action primitives used for converting an event to a string. Given an event observation O, we can use M to convert O into a series of symbols, i.e., $O = \{a_1, a_2, ..., a_n\}$, where $a_t \in M$. In real cases, an event will include many repeated actions. Thus, there are many repeated symbols a_i in O. To represent O more compactly, a new symbol $\mathbf{s}_i$ is used to denote the repeated condition where the ith action in O is a_i and repeated r_i times; that is, $\mathbf{s}_i = (a_i, r_i)$ and $a_i \in M$. Thus, O can be more compactly represented as:

$$O = \{(a_1, r_1), (a_2, r_2), ..., (a_n, r_n)\} = (\mathbf{a}, \mathbf{r}), \qquad (3)$$

where $\mathbf{a} = (a_1, a_2, ..., a_n)$ and $\mathbf{r} = (r_1, r_2, ..., r_n)$.

4. Event Analysis Using Markov Chains

To recognize O, its best event E_{opt} is modeled as a Bayesian MAP(maximum a posterior) optimization:

$$E_{opt} = \arg\max_{E_i \in \mathbf{E}} P(E_i \mid O), \qquad (4)$$

where $\mathbf{E} = \{E_1, E_2, ..., E_{|\mathbf{E}|}\}$ denotes an event configuration. Using Bayesian rules, the right term can be decomposed to

$$P(E_i \mid O) \propto P(E_i)P(O \mid E_i), \qquad (5)$$

where $P(O \mid E_i)$ is a joint likelihood and $P(E_i)$ is a prior. Due to noise and lighting conditions, each event type E_i will own different string representations with different scaling, coding errors, and beginning symbols. To reflect its multiplicity, this paper divide its context to two parts, i.e., the intra-path context O_{intra} and inter-path context O_{inter}. The first one models the dynamic properties of an event within a path and the second one is to measure the multiplicity of an event between its different path representations. Thus, the term $P(O \mid E_i)$ is decomposed to two components:

$$P(O \mid E_i) = P(O_{intra} \mid E_i)P(O_{inter} \mid E_i). \qquad (6)$$

4.1. Intra-path Likelihood

This paper uses a state automata to capture the intra-path context of an event. Fig. 6 shows this state automata for event representation. Each state S includes an action primitive a and the repetition length r. C is a node which places spatial constraints on the primitive a_t. Then, the intra-path likelihood of E_i under O can be decomposed into the form:

$$P(O_{intra} \mid E_i) = P(\mathbf{r} \mid \mathbf{a}, E_i)P(\mathbf{a} \mid E_i), \qquad (7)$$

where $P(\mathbf{r} \mid \mathbf{a}, E_i)$ is the repetition likelihood model of $\mathbf{r}$ when E_i and $\mathbf{a}$ are given. $P(\mathbf{a} \mid E_i)$ is the likelihood model between $\mathbf{a}$ and E_i. $P(O_{intra} \mid E_i)$ follows a Markovian model given by

$$P(O_{intra} \mid E_i) = \left[\prod_{t=1}^{n} p_{rep}(r_t \mid a_t, E_i)\right] \times \left[p(a_1, E_i)\prod_{t=2}^{n} p_{tran}(a_t \mid a_{t-1}, E_i)\right], \qquad (8)$$

where $p_{rep}(r_t \mid a_t, E_i)$ denotes the probability of the action primitive a_t which repeats r_t times under E_i, $p_{tran}(a_t \mid a_{t-1}, E_i)$ is the probability of a_{t-1} transiting to a_t for E_i, and $p(a_t, E_i)$ is the prior probability of a_t for E_i. Actually, Eq.(8) considers only the dynamic properties of action primitives a_t within E_i. In real cases, different paths with different lengths from different initial states and ending conditions can lead to the same event E_i. If only one path is used to model E_i, $P(O_{intra} \mid E_i)$ will be lower and lead to the failure of event classification if O is another resembling path for representing E_i. When considering the multiplicity of E_i, not only the intra-path context but also the inter-path context should be considered for analyzing E_i more accurately.

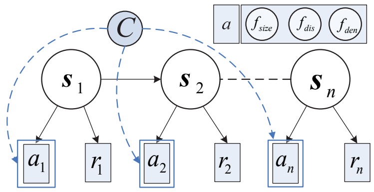


Fig. 6: State automata for event representation.

4.2. Inter-path Likelihood


Fig. 7: A compact representation for a smoking event, *i.e.*, $\{(m_5,5),(m_6,1),(m_4,11),(m_5,2),(m_4,1),(m_5,2),(m_7,8),(m_7,17)\}$.

To reflect the multiplicity of an event, this paper first uses the edit distance to measure the distance between two strings S_p and S_q. The edit distance between S_p and S_q is defined as

$$Ed\_Dist(S_p,S_q)=\xi_{S_p,S_q}(|S_p|,|S_q|). \qquad (9)$$

$\xi_{S_p,S_q}[m,n]$ can be recursively calculated as follows:

$$\begin{aligned}\xi_{S_p,S_q}[m,n]=\min(&\xi_{S_p,S_q}[m-1,n],\xi_{S_p,S_q}[m,n-1],\\ &\xi_{S_p,S_q}[m-1,n-1]+\beta[m-1,n-1])\end{aligned}, \qquad (10)$$

where $\beta[m,n]$ is defined in Eq.(2). Then, for an event E_i, its multiplicity is modeled in its inter-path context through a layer structure. To construct this layer structure, a set V_{E_i} of training event videos is first collected. V_{E_i} forms an event space to capture different syntactic meanings of E_i. For the *j*th video $v_{i,j}$ in V_{E_i}, we will use Eq.(3) to convert it to a path representation $\phi_{i,j}$. Like Fig. 7, we can represent a smoking event as a path like

$\{(m_5,5),(m_6,1),(m_4,11),(m_5,2),(m_4,1),(m_5,2),(m_7,8),(m_7,17)\}$.

Let ϕ_{E_i} denote the collection of all paths $\phi_{i,j}$, *i.e.*, $\phi_{E_i}=\bigcup\phi_{i,j}$. From ϕ_{E_i}, different path hypotheses will be generated and form a layer structure from which the inter-path context of E_i can be extracted. In what follows, the layer structure is discussed.

For each event path $\phi_{i,j}$ in ϕ_{E_i}, its state s_t will associate with an action primitive a_t and the repetition length r_t. Let $\boldsymbol{\gamma}_{i,j}$ denote the union of all repetition lengths r_t, i.e., $\boldsymbol{\gamma}_{i,j}=\bigcup r_t$. After sorting, all elements in $\boldsymbol{\gamma}_{i,j}$ are arranged in ascendant order. Like the example shown in Fig. 7, $\boldsymbol{\gamma}_{i,j}=\{1,2,5,8,11,17\}$. In real cases, the action primitive associated with a shorter repetition length often plays an less important role in event representation. Thus, with the repetition length r_t as a parameter, different path hypotheses with different layers can be generated for representing E_i more accurately. Let $\boldsymbol{\gamma}^k_{i,j}$ denote the *k*th element in $\boldsymbol{\gamma}_{i,j}$. Based on $\boldsymbol{\gamma}^k_{i,j}$, the *k*th layer representation of $\phi_{i,j}$ is formed:

$$\phi^k_{i,j}=\left\{(a_t,r_t)\mid r_t\geq\boldsymbol{\gamma}^k_{i,j},(a_t,r_t)\in\phi_{i,j}\right\}. \qquad (11)$$

The layer is higher, more important syntactic meanings of $\phi_{i,j}$ will be extracted and more noisy symbols in $\phi_{i,j}$ will be filtered out. Let Φ_{E_i} denote the hypothesis set collecting all of the generated path hypotheses $\phi^k_{i,j}$ from each $\phi_{i,j}$ in ϕ_{E_i}, *i.e.*, $\Phi_{E_i}=\{\phi^k_{i,j}\mid\phi^k_{i,j}\in\phi_{i,j}$ and $\phi_{i,j}\in\phi_{E_i}\}$. When constructing Φ_{E_i} from V_{E_i}, the frequency $\alpha_{E_i,l}$ of each hypothesis $\hbar_{E_i,l}$ appearing in Φ_{E_i} is also recorded. After normalization, $\alpha_{E_i,l}$ will become a weight to record the importance of $\hbar_{E_i,l}$ to represent E_i. Given an observation O, a layer structure Φ_O of path hypotheses $\hbar_{o,k}$ can be also generated using Eq.(11). The distance between $\hbar_{o,k}$ and E_i can be calculated by

$$dist(\hbar_{o,k},E_i)=\min_{\hbar_{E_i,l}\in\Phi_{E_i}} EditDist(\hbar_{o,k},\hbar_{E_i,l}), \qquad (12)$$

where $EditDist(\hbar_{o,k},\hbar_{E_i,l})$ is the edit distance between $\hbar_{o,k}$ and $\hbar_{E_i,l}$. Based on Eq.(12), the distance between O and E_i using the inter-path context is defined as

$$dis(O,E_i)=\sum_{\hbar_{o,k}\in\Phi_O}\alpha_{E_i,\lambda}dis(\hbar_{o,k},\Phi_{E_i}), \qquad (13)$$

where $\alpha_{E_i,\lambda}$ is the weight of the λ-th hypothesis $\hbar_{E_i,\lambda}$ in Φ_{E_i} satisfying

$$\hbar_{E_i,\lambda}=\arg\min_{\hbar_{E_i,j}\in\Phi_{E_i}} EditDist(\hbar_{o,k},\hbar_{E_i,j}). \qquad (14)$$

Then, based on Eq.(13), the inter-path likelihood between O and E_i is defined as follows:

$$P(O_{inter}\mid E_i)=\exp(-\frac{dis(O,E_i)}{\sigma^2_{E_i}}), \qquad (15)$$

where σ_{E_i} is the variance of $dis(O,E_i)$.

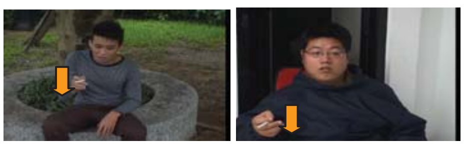
4.3. Transition Probability between Primitives

To get the transition probability $p_{tran}(m_k \mid m_j, E_i)$, a set V_{E_i} of training videos is first collected. For each video v_l in V_{E_i} , it will be decomposed to a series of symbols $m_{l,t}$. Then, $p_{tran}(m_k \mid m_j, E_i)$ can be converted as

$$p_{tran}(m_k \mid m_j, E_i) = \frac{1}{N_f} \sum_{l=1}^{|V_{E_i}|} \sum_{t=0}^{|v_l|-1} \delta_{j,k}^{l,t} , \qquad (16)$$

where $N_f = \sum_{l=1}^{|V_{E_i}|} |v_l| - 1$ and $\delta_{i,j}^{l,t}$ is a transition flag which is 1 if $m_{l,t} = m_j$ and $m_{l,t+1} = m_k$; otherwise, 0. After training, the transition probability table $p_{tran}(m_k \mid m_j, E_i)$ for E_i can be well obtained.

4.4. Repetition Probability Learning



(a) 9 repetitions (b) 27 repetitions

Fig. 8: Different repetitions of an action primitives observed from different persons when a smoking event was performed.

In addition to transition probability, the repetition of an action primitive is also important for event analysis. In real cases, different persons will perform the same action at different speeds. Like Fig. 8, when a smoking event was analyzed, different persons performed the same action primitive with different repetitions. This paper uses the Gaussian mixture model (GMM) to build the repetition probability model of action primitives for event analysis.

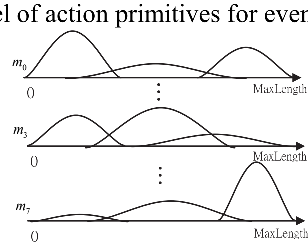


Fig. 9: A GMM-based scheme used to model the repetition probability of each action primitive.

Let $p_{rep}(r \mid m_j, E_i)$ denote the probability of the action primitive m_j which repeats r times for the event E_i . Like Fig. 9, the repetition property of each action primitive m_j is modeled with a set of Gaussian models. The GMMs-based method using K Gaussian distributions to model $p_{rep}(r \mid m_j, E_i)$ of E_i has the form

$$p_{rep}(r \mid m_j, E_i) = \sum_{k=1}^{K} \beta_{j,k}^{E_i} \exp(-[\frac{r - \mu_{j,k}^{E_i}}{\sigma_{j,k}^{E_i}}]^2) , \qquad (17)$$

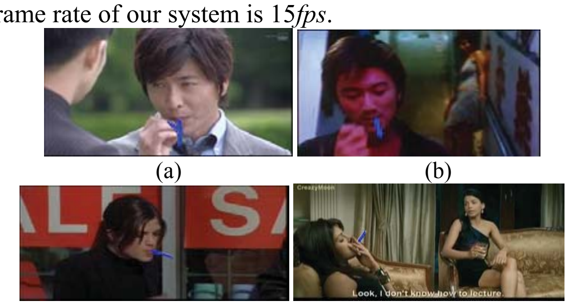
where $\beta_{j,k}^{E_i}$ is the weight of the kth model satisfying $\sum_{k=1}^{K} \beta_{j,k}^{E_i} = 1$ for E_i . Then, a vector $\theta_j^{E_i}$ containing $3K$ elements for E_i can be formed, *i.e.*, $\theta_j^{E} = (\beta_{j,1}^{E_i}, \mu_{j,1}^{E_i}, \sigma_{j,1}^{E_i}, ..., \beta_{j,K}^{E_i}, \mu_{j,K}^{E_i}, \sigma_{j,K}^{E_i})$. K is set to five in this paper. With the expectation maximization algorithm, the estimate of $\theta_j^{E_i}$ can be iteratively obtained. Based on Eq.(17), we can obtain the repetition probability $p_{rep}(r \mid m_j, E_i)$ and use it in Eq.(8).

4.5. Event Classification

This paper proposes a two-stage scheme to analyze various events happening between a face and hands. At the waiting stage, the system determines whether there is an object on the detected face. If yes, a series of action primitives will be collected. If the times of action primitive transitions is larger than n , an event observation O is then driven. At the analyzing stage, we use its intra-path and inter path contexts to calculate $P(O_{intra} \mid E_i)$ and $P(O_{inter} \mid E_i)$ (see Eq.(6)). Then, its corresponding event $E_{optimal}$ can be determined from Eq.(4).

5. Experimental Results

To analyze the performance of our proposed approach, three event types were collected in this paper, i.e., smoking, and phoning. Each type was performed by twenty actors five times under different lighting conditions; thus there were 100 sequences collected for each event. We should emphasize our proposed scheme can be easily extended to analyze other kinds of events directly from videos. The frame rate of our system is 15*fps*.



(a) (b)

(c) (d)

Fig. 10: Cigarette detection from four movies. (a) and (b): Front views. (c) and (d): Sided view.

Fig. 10 shows the detection results of cigarettes from four movies. The cigarettes even with different orientations were still well detected. Another difficult case is to detect objects on faces from a sided view. Fig. 10(c) and (d) shows the results of cigarette detection when a sided view was adopted. Each case was successfully detected even though these cigarettes were with different orientations and observed from different views. Fig. 11 shows the results of can and phone detection. Table 1 lists the accuracy analyses of hand-hold object detection. The average accuracy of cigarette detection is 90.7 %. Since the can size is larger than the cigarette size, its accuracy is better than cigarette detection. In real cases, the phones sometimes will be occluded by hands. The occlusion condition often leads to the failure of phoning detection. Thus, its accuracy is lower, i.e., 86.97%.

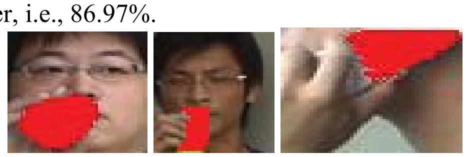


Fig. 11: Results of can and phone detection.

Table 1: Accuracy analyses of hand-hold object detection.

Objects	Frames	True Pos.	False Neg.	False Pos.	Miss rate	False Alarm	Accuracy
Cigarette	3012	2732	248	32	8.23%	1.06%	90.7 %
Drink	9128	8941	39	148	0.42%	1.62%	97.9 %
Phone	2811	2445	235	131	8.36%	4.66%	86.97%

Fig. 12: Results of smoking event detection for videos.

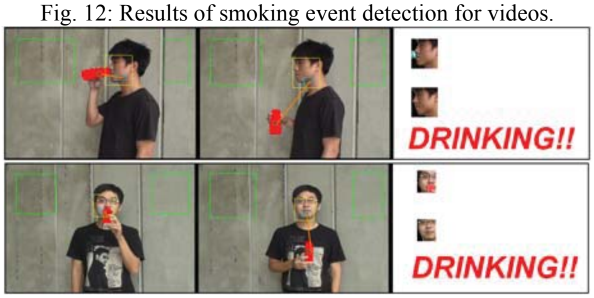


Fig. 13: Results of drinking event detection when different face orientations were observed.

Table 2: Accuracy analysis of event detection.

Types	Smoking	Drinking	Phone	Miss
Smoking	89%	3%	1%	7%
Drinking	4%	92%	1%	3%
Phoning	2%	3%	86%	9%

Fig. 12 shows the cases of smoking event analysis. Fig. 13 shows the cases of drinking event analysis under different view orientations. Table 2 shows the accuracy analyses among the three events. For the phoning event, like phone detection, sine the phone is easily occluded by hands, its accuracy is the lowest. All the above experiments have proved that the proposed method is a robust, accurate, and powerful tool for analyzing interaction events between hands and faces.

6. Conclusions

This paper presents a novel approach for analyzing the events between faces and hand-hold objects. The contributions of this paper are summarized as follows:

(a) A novel histogram-based scheme was proposed for detecting hand-hold objects. Thus, different abnormal events happening between faces and hands can be further analyzed.

(b) A coding technique was proposed for converting each frame to a series of symbols. Then, a path-based method can be proposed for representing different events happening on faces.

(c) A novel event representation method was proposed for capturing the multiplicity of events. Thus, even though an event owns different visual variations, it still can be well represented.

(d) A HHM method was proposed for capturing the dynamics of an event. Thus, even though an event has different time-scaling problem, it still can be well analyzed.

The average accuracy of our proposed system is 89%.

7. References

[1] L. Fengjun and R. Nevatia, “Single View Human Action Recognition using Key Pose Matching and Viterbi Path Searching,” IEEE Conference on CVPR, pp. 1-8, Jun. 2007.

[2] A. Fathi and G. Mori, “Action recognition by learning mid-level motion features,” IEEE Conference on CVPR, pp. 1-8, 2008.

[3] N. Jojic, et al., “Transformed hidden Markov models: estimating mixture models and inferring spatial transformations in video sequences,” In Proc. Conf. CVPR, vol. 2, pp. 26-33, 2000.

[4] B. Laxton, et al., “Leveraging temporal, contextual and ordering constraints for recognizing complex activities in video,” IEEE Conference on CVPR, pp. 1-8, Jun. 2007.

[5] R. Filipovych and E. Ribeiro, “Recognizing Primitive Interactions by Exploring Actor-Object States,” IEEE Conference on CVPR, pp. 1-7 Jun. 2008.

[6] I. Laptev and P. Perez, “Retrieving actions in movies,” International Conference on Computer Vision, Oct. 2007.

[7] S. Yifan, et al., “Propagation Networks for Recognition of Partially Ordered Sequential Action,” IEEE Conference on CVPR, pp. 862-869, July 2004.

[8] K. Kim, et al., “Real-time Foreground-Background Segmentation using Codebook Model,” Real-time Imaging, vol. 11, no. 3, pp. 172-185, June 2005.

[9] P. Viola and M. Jones, “Robust Real-time Object Detection,” International Journal of Computer Vision, vol.57, no.2, pp.137-154, May 2004.

Soft Contact-lens Sensor for Monitoring Tear Sugar as Novel Wearable Device of Body Sensor Network

M.K. Chu, K. Mitsubayashi
Dept. of Advanced Sci. & Tech. for Biomedical Sensors,
Tokyo Medical and Dental University
Tokyo, Japan

K. Miyajima, T. Arakawa, H. Kudo, K. Mitsubayashi
Dept. of Biomedical Devices and Instrumentation,
Tokyo Medical and Dental University
Tokyo, Japan
m.bdi@tmd.ac.jp

***Abstract*— A soft contact-lens amperometric glucose sensor as novel wearable device of body sensor network was fabricated and tested. Also, the sensor was utilized to tear glucose monitoring. The sensor was constructed by immobilizing GOD onto a flexible oxygen electrode, which was fabricated using "Soft-MEMS" techniques onto a functional polymer membrane. In purpose of bioinstrumentation, adhesive agents were not used for constructing the flexible biosensor. Linear relationship between glucose concentration and output current was obtained in a range of 0.039 – 0.537 mmol/l. Current dependences on pH and temperature were also evaluated. The current was largest at pH 7.0 and the current increased when temperature increased. This indicates that the output current depends on enzyme activity. Based on the basic characteristics investigation, the glucose sensor was applied to measurement of glucose in tear fluids on an eye site of a Japan white rabbit. The change of tear glucose level induced by oral-administration of glucose was monitored as a current change of the sensor attached on the eye site. In this investigation, the tear glucose level varied from 0.2 mmol/l to 0.5 mmol/l. Although there was a delay of several tens of minutes towards blood sugar level, it is considered to be possible that non-invasive continuous glucose monitoring can be realized using the flexible biosensor.**

Keywords-wearable sensor; body sensor network; tear sugar; diabetes control

I. INTRODUCTION

A continuous and non-invasive bio/chemical monitoring with a body network system is required for daily human health care, especially chronic and/or lifestyle diseases (Fig. 1). Rapid increasing of diabetes mellitus is now global problem and development of a safe and convenient blood sugar level monitoring technology is strongly required. While "finger pricking" is a relatively painless and commonly used method for blood glucose monitoring today, there are several points to be improved. One of the most serious issues is that this method forces diabetic patients to frequent check with invasive self-care kits [1]. Although many measurement methods of have been investigated [2-4], the inconvenience of blood sugar monitoring is still not fundamentally improved. For this reason, development non-invasive and continuous method of blood sugar monitoring, which is available in daily life within a reasonable cost, is strongly requested. Generally, continuous glucose level monitoring do not measure blood glucose directly, but rely instead on measurement of the glucose levels in other biological fluids. Particularly, correlation with glucose level in interstitial fluid of subcutaneous tissue to blood glucose level is often used [5]. Relationships between other biochemical substances of body fluids (tears, airway mucus, sweat and saliva) and personal physical conditions are also reported [6,7] and expected to be used for continuous bioinstrumentations. We paid attention to the relationship between the tear glucose level and the blood sugar level which is previously reported using discrete monitoring method such as capillary electrophoresis [8].

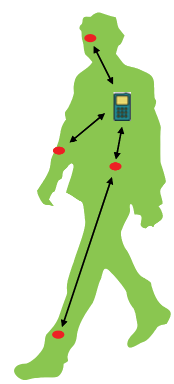


Figure 1. Wearable Biosensors with Body Sensor Network for Human Bio/Chemical Monitoring.

In this study, we have developed a flexible amperometric biosensor towards potential use within eyes for continuous tear glucose monitoring. The biosensor was fabricated using "Soft-MEMS" techniques [9]. The sensor has an 84 μm thick laminar structure, which contains gas-permeable membrane, non-permeable membrane, electrolytes, electrodes and glucose immobilized membrane. The electrodes were formed using microfabrication process. Also, the biosensor was utilized to continuous glucose monitoring in tear fluids on rabbit's eye site. The advantage of the flexible sensor is that the sensor has thinner and flexible structure suitable for biomedical instrumentation because they do not cause discomfort. The sensor consists of reed-shaped functional polymers and film electrodes.

In this paper, details of the design, fabrication and evaluation of the flexible biosensor are presented. Also, the result of continuous and non-invasive tear glucose monitoring using the flexible biosensor is reported.

II. Experimental Section

A. Design and fabrication of a flexible biosensor

A schematic structure of the flexible biosensor is illustrated in Fig. 2. The glucose sensor has a 3mm x 50mm x 84µm laminar structure, which consists of an enzyme immobilized membrane and film-like oxygen electrode (Pt working electrode and Ag/AgCl reference/counter electrode). Owing to continuous tear glucose measurement in mind, whole structural members were constructed with flexible polymers. Adhesive agents were not used for constructing the wearable glucose sensor. The sensor was constructed by immobilizing the enzyme membrane onto the sensitive area of the oxygen electrode.

The film-type oxygen electrodes has four layers; (i) a flexible gas-permeable membrane (polypropylene, thickness: 25 µm), (ii) a 200 nm thick Pt electrode and a 300 nm thick Ag/AgCl electrode, (iii) a membrane filter with dimensions of 1 mm x 1.5 mm (IsoporeTM TKTP04700, Millipore Corp., USA) containing electrolytic solution (0.1 mol/l KCl 198-03545, Wako Pure Chemical Industries, Ltd., Japan) and (iv) a non-permeable membrane (Ionomer, film thickness: 50µm).

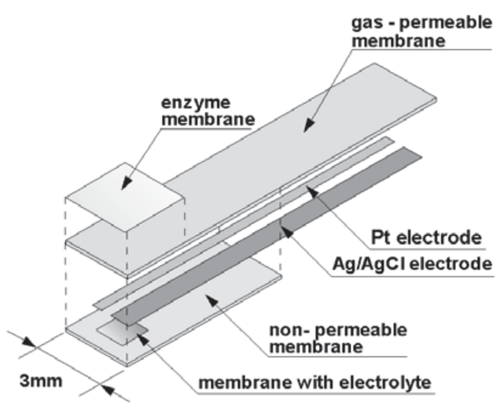


Figure 2. Structure of a wearable sensor for tear glucose monitoring.

The electrode reaction is given as following equations:

Cathode (Pt) : $O_2+2H_2O+4e^-\quad 4OH^-$

Anode (Ag/AgCl) : $4Ag+4Cl^-\quad 4AgCl+4e^-$

The oxygen electrode was fabricated using MEMS (Micro Electro Mechanical Systems) techniques. At first, the gas-permeable membrane was fixed on a dummy silicon wafer using polyimide tape. Positive photoresist (Shipley S1818, Rohm and Haas Electronic Materials Co., USA) was then spin-coated. In order to prevent the gas-permeable membrane from thermal damage, the positive resist was cured at 60 °C. The photoresist was patterned into the shape of electrodes using UV exposure system (MA-10, Mikasa CO., LTD., Japan). Then, a 200 nm thick Pt was sputtered (CFS-4ES-231, SHIBAURA ENGINEERING WORKS CO., LTD., Japan) onto the patterned surface with a background pressure of 3.0 x 10-3 Pa and the sputtering pressure of 3.0 x 10-1 Pa. The pattern of the Pt working electrode was covered by positive photoresist (RP-2, San-hayato Co., Japan). And a 300nm Ag was deposited directly on Pt. The gas-permeable membrane was then released from the dummy Si wafer and the Pt and Ag electrodes were formed using lift-off process. The Ag/AgCl electrode was formed by controlled-potential electrolysis. A constant voltage of 110 mV was applied to the Ag film in a 0.1 mmol/l HCl solution.

Then, the flexible oxygen electrode was constructed in a sandwich configuration with a membrane filter (IsoporeTM) containing electrolyte solution between the gas-permeable membrane with Pt and Ag/AgCl electrode and the non-permeable sheet. All the edges of the layer-built cell with 0.1 mmol/l KCl were fastened by heat-seal system (SURE Sealer, NL-201P, Ishizaki Electricity Manufacturing Co. Ltd., Japan). The temperature was 150 °C. Thus, the flexible Clark-type oxygen electrode with a bag-like electrolyte cell was obtained.

The wearable glucose sensor was fabricated by immobilizing glucose oxidase (GOD: EC1.1.3.4, G-7141, Sigma Chemical Co., USA) onto the sensing region of the flexible oxygen electrode. GOD was immobilized using water-soluble photosensitive resin (AWP: Azide-unit pendant Water-soluble Photopolymer, Toyo Gosei Kogyo Co., Ltd., Japan) as shown in the upper part of Fig. 1. In order to improve the contact between the gas-permeable membrane and the enzyme immobilized membrane, an aminopropylsilane monolayer was prepared on the surface of the gas-permeable membrane.The sensing region of the oxygen electrode was coated with 1-3 aminopropylsilane monolayer. Then, it was rinsed and dried for 30 min at room temperature. 15 mg of GOD was mixed into 500 µl phosphate buffer solution (PBS: pH 7.0, 20 mmol/l). The 40µl GOD and PBS mixture was mixed with a 40µl AWP. The GOD/PBS and AWP mixture was applied to the monolayer formed on the oxygen sensor as an enzyme membrane. AWP was cured under 5 °C. After that, the enzyme membrane was cured using UV light, thus obtained the flexible biosensor for tear glucose measurement as shown in Fig 3.

Figure 3. Photograph of the flexible glucose sensor.

B. Evaluation of the flexible biosensor

The sensor was calibrated using a batch measurement system using a 50 ml measuring cell filled with PBS (pH 7.4, 20 mmol/l, containing glucose solution). Oxygen consumption induced by enzyme reaction of the enzyme immobilized membrane was measured by two-electrode electrochemical method. The sensor was connected to the measurement system and the reduction potential of -550 mV versus Ag/AgCl reference/counter electrode was applied to the Pt working electrode using PC-controlled potentiostat (HAB-151, HOKUTO DENKO Co., Japan). The output currents were recorded by PC, which was connected to the potentiostat via A/D converter (ADC-16, pico Technology Ltd., UK), within the glucose concentration ranging from 0.025 – 1.475 mmol/l.

The current dependences on various pH and temperatures at defined glucose concentration (1mmol/l) were also investigated. The constant potential of - 550 mV versus Ag/AgCl electrode was applied to the Pt electrode in solutions of various pH (6.0 – 12.0) and various temperature (15 – 60 °C). The current change induced by varying operating condition was recorded using PC. In the experiment, mixtures of disodium hydrogen phosphate (196-02835, Wako Pure Chemical Industries, LTD., Japan) and potassium dihydrogenphosphate (169-04245, Wako Pure Chemical Industries, LTD., Japan), which pH was controlled using pH meter (D-25, Horiba Korea LTD., Korea) to measure pH dependence.

C. Tear glucose measurement using the flexible biosensor

Using the flexible glucose sensor, tear glucose monitoring on eye site of a rabbit (Japan white rabbit, sex: female, age: 18 month, weight: 2kg) was carried out. The rabbit was placed into a cylindrical fixation device. In order to reduce physical load, the animal experiment was carried out without use of an anesthetic. The experimental method of glucose monitoring on eye site is shown in Fig. 4. The sensing region of the sensor (glucose immobilized membrane) was attached on the pupil and the other end of the sensor was fixed using fixing tape. The sensor was operated using the PC-controlled potentiostat and the output current was continuously monitored using PC, which was also used for evaluation of basic characteristics.

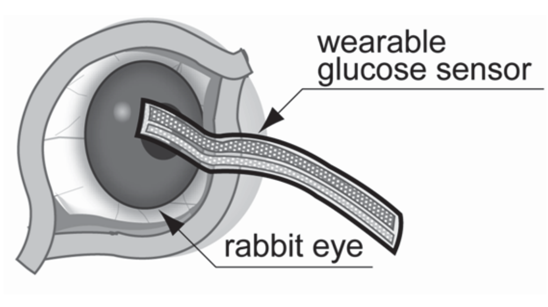


Figure 4. Measurement method of glucose concentrations in tear fluids using the flexible glucose sensor.

The monitoring test on eye site was carried out within two steps. At first, the sensor output during steady state was monitored and the stability of the output current was investigated. After that, the change of tear glucose level induced by oral glucose load was measured using the glucose sensor. The sensor was attached on the rabbit's eye site as well as the preceding test. Glucose solution was then orally-administrated to the rabbit. Quantity of the glucose solution was determined to 2g (1g of glucose per 1kg of weight). Blood glucose level was also measured using a commercially produced monitoring kit (MEDISAFE, TERUMO Co., Japan) as a controlled study. Blood sample was taken from the rabbit's ear.

III. Results and Discussion

A. Characteristics of the flexible biosensor

The calibration plot of the flexible biosensor is shown in Fig. 4. In this figure, the linear relationship between glucose concentration and output current was confirmed in a range of 0.039 – 0.537 mmol/l, deduced by regression analysis, as shown by following equation;

$$\text{output current } (\mu A) = -0.06 + 1.822\ [\text{glucose (mmol/l)}]$$

The calibration range of the glucose sensor covers the glucose concentration of normal tear fluid (0.14 mmol/l) [10] and the result indicates that the flexible biosensor has an appropriate measurement range for tear glucose measurements in diabetic patients. The typical response curve of the sensor is also shown in the inset figure of Fig. 5.

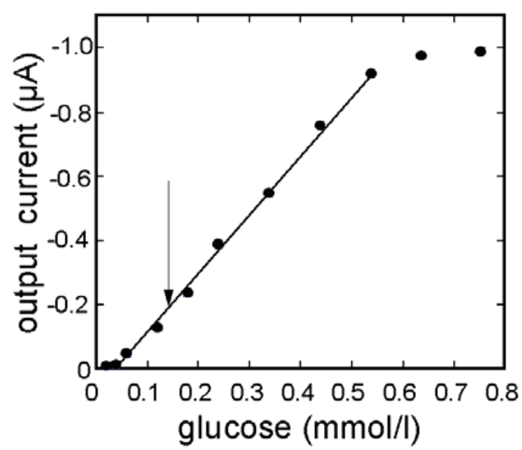


Figure 5. Calibration curve of the wearable glucose sensor.

The current change induced by various pH showed a peek at pH 7.0, which is the optimal pH of the enzyme activity. In the pH ranging from pH 5.0 to 7.0, the current increases as pH increases. The current rapidly decreased during pH 7.0 – 7.5. In case of pH > 7.5, the output current showed a slow decrease. This is because the activity of immobilized enzyme degrades significantly in alkaline pH. The current increases as the temperature increase less than 45 °C. This indicates that the reaction rate is the dominant factor in this temperature range. On the other hand, the current decreases over 45 °C. In this temperature range, thermal deactivation of immobilized enzyme determines the output current rather than reaction rate.

B. *Tear glucose monitoring using the flexible biosensor*

Prior to continuous tear glucose measurement, the sensor was attached on the rabbit eye as mentioned before and tear glucose level of the rabbit eye was monitored repeatedly. The average value of tear glucose level was 0.124 mmol/l. Those glucose concentrations in tear fluids were estimated from the calibration plots. Those values were fixed by pH and temperature dependence.

The estimated tear glucose level and blood sugar level during the oral administration of glucose to the Japan white rabbit were compared. We confirmed the stability of the sensor output for 10 minutes and then oral glucose tolerant test was carried out. The initial estimated glucose level in tear fluids was 0.2 mmol/l and increased up to 0.5 mmol/l by the experimentation. As the figure indicates, the glucose level in tear fluids increases by 3-fold while blood glucose increases by 2-fold. This might be considered that the output current includes an effect of interferences. Blood glucose level increased immediately when glucose load was applied to the rabbit. On the other hand, tear glucose level increased within a delay of 10 - 20 minutes in compare with blood glucose level. The initial current change started 10 minutes after glucose administration. The tear glucose level significantly increased from 40 to 60 minutes after oral administration. Although the effect of interferences and a delay of tens of minutes, the output current, which has relationship with the tear glucose level, tracked the blood glucose level with a considerable correlation during this investigation. Significant variation of the output current was not observed except when the sensor displacement was occurred by blinking during 5 minutes. During the investigation, the sensor output was not degraded or drifted. Thus, continuous tear glucose monitoring was successfully demonstrated. We have to improve a novel wearable sensor like a contact lens for long term analysis of tear glucose level because the strip-style flexible sensor is easy to come off from the rabbit eye.

The flexible biosensor is not only useful for continuous measurement at the ophthalmic site and the skin surface, but also useful for the chemical analysis in the biological fluids secreted from the internal organs, tissues, etc.

IV. CONCLUSIONS

A flexible and wearable amperometric glucose sensor was developed and utilized to tear glucose monitoring. The sensor was constructed by immobilizing GOD onto a flexible oxygen electrode, which was fabricated using MEMS techniques onto a functional polymer membrane. Linear relationship between glucose concentration and output current was obtained in a range of 0.039 – 0.537 mmol/l.

Owing to unique features of flexibility, biocompatibility and thinner structure, continuous tear glucose monitoring was carried out. The sensor was sufficiently stable and sensitive when it is attached on rabbit eye. Thus, change of tear glucose level induced by oral administration of glucose was measured using the biosensor. Blood sugar level was also measured by a commercially produced monitoring kit as a controlled study. As a result, the tear glucose level was increased from 0.2 to 0.5 mmol/l. Although the change of tear glucose level delayed in tens of minutes from that of blood sugar level, it is considered to be possible that non-invasive continuous glucose monitoring can be realized using the glucose sensor.

ACKNOWLEDGMENT

This work is partly supported by “Next-Generation Actuators Leading Breakthroughs”, MEXT Special Funds for Education and Research “Advanced Research Program in Sensing Biology”, JSPS (Japan Society for the Promotion of Science) Grants-in-Aid for Science Research System and Nakatani Foundation of Electronic Measuring Technology Advancement.

REFERENCES

[1] Badugu, R., Lakowicz, J.R., Geddes, C.D., 2003. A glucose sensing contact lens: A non-invasive technique for continuous physiological glucose monitoring, Journal of Fluorescence 13, 371-37.

[2] D'Auria, S., Dicesare, N., Gryczynski, Z., Gryczynski, I., Rossi, M., Lakowicz, J. R., 2000. A thermophilic apoglucose dehydrogenase asa nonconsuming glucose sensor. Biochem. Biophys. Res. Commun.274, 727–73

[3] Tolosa, L., Gryczynski, I., Eichorn, L. R., Dattelbaum, J. D., Castellano, F. N., Rao, G., Lakowicz, J. R., 1999. Glucose sensors for low cost lifetime-based sensing using a genetically engineeredprotein. Anal. Biochem. 267, 114–120.

[4] Heise, H. M., Marbach, R., Koschinsky, T. H., Gries, F. A., 1994. Non-invasive blood glucose sensors based on near-infrared spec-troscopy. Ann. Occup. Hyg. 18, 439–447

[5] Wilson G. S. and Gifford R., 2005. Biosensors for real-time in vivo measurements, Biosensors and Bioelectronics 20, 2388-2403.

[6] Mitsubayashi, K., Suzuki, M., Tamiya, E., Karube, I., 1994. Analysis of metabolites in sweat as a measure of physical condition. Anal. Chim. Acta 289, 27-34.

[7] Daum, K. M., Hill, R. M., 1982. Human tear glucose. Invest. Ophthalmol. Vis. Sci. 22, 509-515.

[8] Sen, D.K., Sarin, G.S., 1980. Tear glucose levels in normal people and in diabetic patients. Br. J. Ophthalmol. 64, 693-699.

[9] Kudo H., Sawada T., Kazawa E., Yoshida H., Iwasaki Y. and Mitsubayashi K., 2006. A flexible and wearable glucose sensor based on functional polymers with Soft-MEMS techniques. Biosens. Bioelectronics 22, 558-562

[10] Jin, Z., Chen, R., Colon, L.A., 1997. Determination of Glucose in Submicroliter Samples by CE-LIF Using Precolumn or On-Column Enzymatic Reactions. Anal. Chem. 69, 1326-1331.

Palamede: a Multi-Press Open Journal System

P. Bellini, P. Nesi and G. Pantaleo

Distributed Systems and Internet Technology Lab, DISIT Lab, http://www.disit.dsi.unifi.it
Department of Systems and Informatics,
University of Florence – Firenze,
Italy
pbellini@dsi.unifi.it, nesi@dsi.unifi.it, pantaleo@dsi.unifi.it

***Abstract*— The strong movement of Open Access journal is convincing many institutions to adopt an infrastructure for collecting and distributing open access journals. For this purpose, there exists popular open source software for electronic journal publishing and editorial management. Most of them are WEB based multi-language platforms capable of hosting several journals in a single press or editorial board environment. With the recent wave of Open Access and the aim of reaching critical mass, many institutions are thinking to adopt common solutions to aggregate the open access journals of multiple institutions/presses on the same infrastructure. This paper presents requirements and analysis for the development of multi-press platform, and its related implementation and validation performed expanding OJS (Open Journal System) as a multi-press and multi-journal platform. It involved a deep reengineering of the originally distributed OJS architecture. The proposed solution refers to Palamede project and that has produced the experimental portal (http://palamede.fupress.com), and which it has been validated by a test experimentation of three Italian University Presses: the Firenze University Press (FUP), the University of Parma and the Forum Editrice Press of the University of Udine.**

***Keywords* - e-journal, open access journals, online publishing system, multi-press open journal, Open Journal System (OJS), peer-review editorial process.**

I. Introduction

The strong movement of Open Access journal is convincing many institutions to adopt an infrastructure for collecting and distributing open access journals. For this purpose, there exists popular open source software for electronic journal publishing and editorial management.

There are many online-developed tools and frameworks designed for building open access digital repositories, working as digital assets and content management systems, and offering services for online content publishing. Among them:

- OPEN Journal System (OJS) developed as part of the research program of the Public Knowledge Project (PKP), started in the end of the 1990s at the University of British Columbia [1].
- DSpace project [3], developed by joint efforts of MIT and HP Labs in Cambridge University, is worthy to be mentioned for digital preservation activity.
- Fedora Repository [4] into the Duraspace [5] organization.
- EPrints [6] and the SHERPA [7] projects aim at addressing OAI-compliant and institutionally-based repositories to expand benefits of academic open access and dissemination.
- EScholarship of University of California [8] focused on academic dissemination and scholarly publishing services.
- Daedalus of the University of Glasgow [9] mainly focused on academic dissemination and scholarly publishing services.

Among the above presented solutions, OJS is an Open Source software (available under the GNU public license [2]), developed with the purpose of creating a WEB based solution, for multi-users and multi-language platform, which is able to support the whole editorial process, from authors' submissions to online publishing, indexing and archiving, including the peer-review editorial process. OJS has become one of the most widespread electronic journal publishing system within the web community, counting on over 8300 installations in the world, as reported on web site [1]. Further and more detailed information about the last updated OJS system version can be retrieved in [10]. Technical documentation is available at [11] and [12].

The aim of this paper is to present the work performed in Palamede project that is focused on realizing a multipress open journal management systems to host multiple presses and put in common a set of services. This approach should improve the flexibility and the quality of electronic journals publishing system, as well as to expand access to research.

The Palamede project involved *FRD - Fondazione Rinascimento Digitale* foundation [13] as financial partner, the *DISIT Lab* of the Department of Systems and Informatics of University of Florence [14] as technical partner, and editorial groups belonging to three Italian Universities: *Firenze University Press* [15] (University of Florence), *Università di Parma* (University of Parma [16]) and *Forum Editrice Universitaria Udinese* [17] (University of Udine). The starting activity of the project was focused into gathering and collecting requirements, suggestions from the University Presses. The main demanding aspect was found to be a commonly aimed goal of a unique aggregator of collected resources, which could be able, though, to preserve the individual and functional identity of each press.

This paper presents requirements and analysis for the development of multi-press platform, and its related implementation and validation performed expanding. For this purpose, the team identified as starting point the above mentioned OJS solution, that, as the others, is only suitable to cope with only one press editor or institution.

The article is structure as follows. In Section II, the requirements are reported. Section III is devoted to the architecture and analysis. Section IV presents some design aspects and the validation performed. Conclusions are drawn in Section V.

II. REQUIREMENTS

In this section, the novel contribution to the presently distributed OJS implementation to allow it coping with multiple presses is presented. The presented solution aims at improving OJS in order to have more flexibility in the direction of providing support for managing multiple press and multi-journal front-end.

The achievement of such a goal is believed to provide many advantages for both editors and users. Among them:

- resources are collected under a unique aggregator, which can grant uniformity of behavior in document indexing and metadata harvesting (following the OAI-PMH standard [18], or other OAI compliant services, e.g. OAIster). Thus reducing the costs for content management for the institutions and provides advantages to the users that may access to multiple press journals on the same web portal;
- such a system offers facilities in committing thematic-based researches upon single journals, specific sets of journals or presses, or upon the whole set of published journals; by this way, the opportunity of cross-thematic approaches research is given;
- users are requested to register only once in the portal; then they have the possibility to subscribe all the presses and journals they are interested in. Otherwise, users should registered to all the different presses or journal sites they are interested to. This is also an evident advantage for the institutions managing the single presses since they may have a cross advertising among the users that are registered on the other presses.

OJS, currently released at version 2.3.4, presents the following main features [9]. The OJS:

- is a multi-language and multi-user platform, installed and controlled locally;
- offers an automated peer-review editorial process for submitted articles, with e-mail notification and commenting ability;
- manages submitted content, for each published journal, can be comprehensively managed and indexed by the editorial staff.

Moreover, OJS offers several other tools, plugins and utilities: a text string search engine; reading tools for published content; a payment module to accept journal fees and a complete online help support.

In OJS, a hierarchy of user roles is used to define editorial roles within the hosted journals. The administrator can perform basic activities as: manage the site graphical style and layout (which is fully configurable through the use of CSS style sheets), create new journals and enroll one or more journal managers.

On the other hand, the journal manager acts in a way similar to the web site manager for his own journal: he owns the permissions to handle the journal graphical layout; furthermore, he can create editorial groups and enroll users as one (or more) of the following roles: editor, section editor, layout editor, copyeditor, author, subscription manager, proofreader and reader. All of these member types have specific tasks in the editorial process; a detailed description of these functionalities is referred to [9].

Finally, the journal manager is in charge of defining the guidelines for the peer-review process, the journal archiving methods, the editorial policies and the methods of subscriptions and access.

A comprehensive overview of the workflow chart of the extended multi-press OJS is depicted in Figure 1, *in which are also presented our additions.*

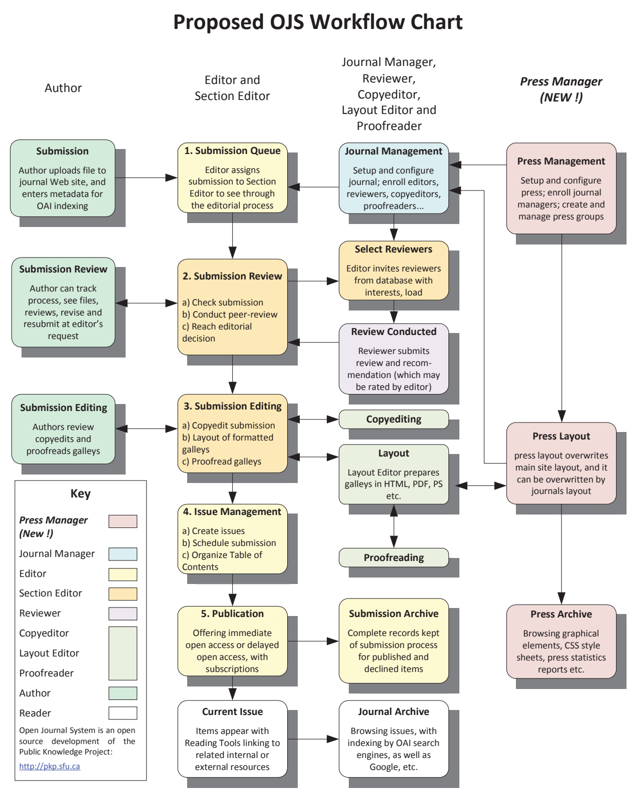


Figure 1. Functional architecture of extended OJS editorial workflow.

The PHP object-oriented architecture of OJS contains different type of classes which model, amongst other objects, both the editorial entities (journals, articles, issues etc...) and the roles of registered users (journal manager, editor, author, reader etc…). The operation of creating a multi-press OJS involved a deep and structural reengineering of the original OJS architecture. All the characteristics and features, concerning the single journal management, have been maintained.

The main efforts of our work has been spent in performing a deeper analysis on the present mono-press version of the OJS to identify the minimum changes to satisfy the new requirements without degenerating the present OJS structured and concepts. The aim has been oriented to implement the new classes which are needed to create the new press entity (the most important ones are the Press class and the Press Manager class) and the appropriate modifications to the source code, where needed, in order to be correctly interfaced with the new objects. A more detailed overview of the novel architecture of the multi-press OJS system is presented in next section.

New Classes included in multi-press OJS	
Modules	**Classes**
PressAbout	`PressAboutHandler`
PressManager	`PressEmailHandler`
	`PressFilesHandler`
	`PressManagerFunctionsHandler`
	`PressGroupHandler`
	`PressLanguagesHandler`
	`PressManagerHandler`
	`PressManagerJournalHandler`
	`PressPeopleHandler`
	`PressPluginHandler`
	`PressPluginManagementHandler`
	`PressManagerSettingsHandler`
	`PressSetupHandler`
	`PressStatisticsHandler`
PressSearch	`PressSearchHandler`
PressUser	`PressUserHandler`
Admin_Form	`PressSettingsForm`
File	`PressFileManager`
Press	`Press`
	`PressDAO`
	`PressSettingsDAO`
	`PressStatisticsDAO`
PressManager_Form	`PressGroupForm`
	`PressLanguageSettingsForm`
	`PressSettingsForm`
	`PressUserManagementForm`
PressManager_Form_Setup	`PressSetupForm`
	`PressSetupStep1Form`
	`PressSetupStep2Form`
	`PressSetupStep3Form`
Template	`TemplatePressManager`

TABLE I. COMPREHENSIVE LIST OF NEW ADDED CLASSES IN THE EXTENDED MULTI-PRESS OJS SYSTEM. NEW CREATED MODULE NAMES ARE DISPLAYED, IN THE LEFT COLUMN, IN ITALIC STYLE.

III. SYSTEM ARCHITECTURE AND ANALYSIS

The native architecture of OJS is designed in a way such that the user interface, the data storage and management and the control structure are kept separated. This configuration recalls the Model-View-Controller (MVC) structure. Let us to illustrate the main details of the three sub-structures:

- **User interface** is set up through the use of Smarty Templates, which assemble HTML pages to be displayed to the users.
- **Control structure** is organized into three types of classes:
  - ***Page Classes***, which receive users requests by the web browser, recall the suitable classes for requests processing, and finally call up the appropriate Smarty template to generate the response;
  - ***Action Classes***, which are in charge of performing the processing of user requests;
  - ***Model Classes***, which implement the objects representing the system various entities, e.g., journals, articles, issues, users and roles.
- **Data storage and management** is carried out by ***Data Access Objects*** (DAOs), which fulfill all interaction queries with the SQL database (MySQL and PostgreSQL databases are currently supported).

A inner back-end type of classes, the **Support Class**, provides core and other miscellaneous functionalities and utilities.

A. Design of the multipress OJS version

The design model of native OJS architecture has been maintained in the new multi-press OJS. The previous brief architectural description is useful to better identify and understand the new implementations and contributions described in this section.

Let us to analyze the modifications supplied, according to the formerly recalled MVC structure:

- **Control Structure** – The new created classes mainly belong to Page and Mode Class types. Actually, the most common sorts of tasks performed by Action Classes are related to specific user roles, e.g., sending e-mails and notifications, handling uploaded files etc.. A complete list of new built-up Page and Model Classes is presented in Table 1.
- **Data Storage and Management** – The database design have been modified to store information and data for the press as a new editorial entity. Consequently, two tables were added: `presses` and `pressSettings`, as well as the corresponding DAOs: `PressDAO` and `PressSettingsDAO` classes. In other tables, describing press related objects, a press identifier field has been added as a row. The relational structure of the new multi-press database is depicted in Figure 2.
- **User Interface** – The implementation of new classes in the Control Structure required the creation of new Smarty templates, in order to be correctly integrated with the user interface. A list of the templates added in the extended OJS system is shown in Table 2.

In addition to the implementation of new classes and templates, it is relevant to stress that most of the other source files composing the native OJS framework have been modified, in order to be appropriately interfaced within the novel multi-press architecture.

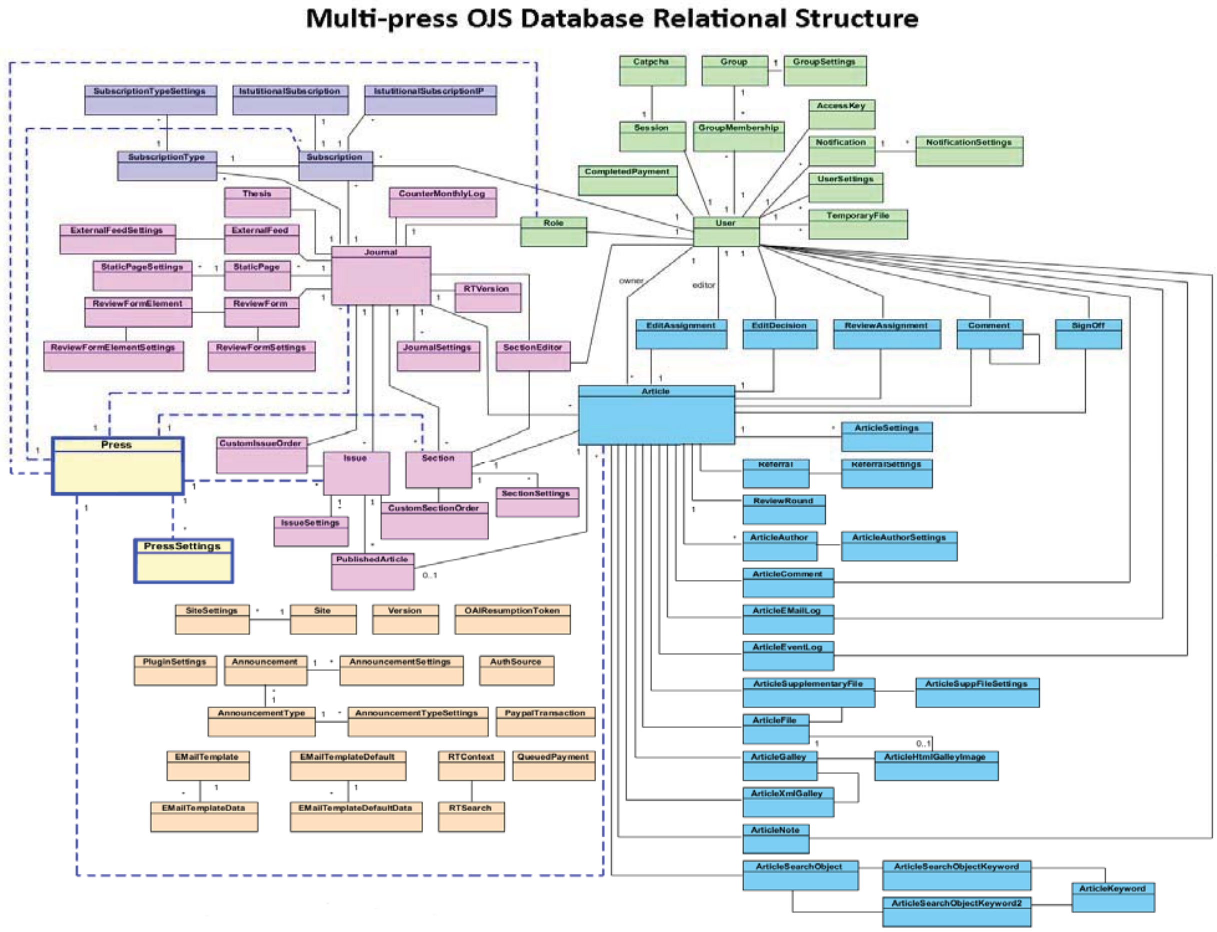


Figure 2. Representation of the relational structure of the multi-press OJS Database. The new added tables (Press and PressSettings) and corresponding new built relations are in evidence.

In the following, a summary of the main additions, improvements and revisions which have been applied to the original OJS architecture in the reengineering procedure, is listed:

- Creation of new Classes, Smarty templates and Database objects concerning the Press editorial entity.
- Creation of Press manager Role, and definitions of its main functionalities:
  - Creation / Enrollment of Journal managers;
  - Creation Editing / Deletion of new journals;
  - Graphical layout management of the press (which may be replaced by different style elements at journal level);
  - Creation of user groups related to Press functionalities.
- Modification of Site Administrator role:
  - Creation / Editing / Deletion of new presses;
  - Creation / Enrollment of Press managers.
- Implementation of statistics at press level, monitoring number of registered users, number of published issues and articles.
- Creation of new Smarty templates for Press functionalities.
- Addition of localization xml files for new created user interface entries.
- Modification of the OJS source code in order to correctly interface the original OJS features within the new multi-press architecture.

The Press Manager role has been designed assembling specific functionalities of both the Site Administrator and the Journal Manager: from the former he inherited the skill of customizing the general graphical layout of his own press, as well as creating, modifying and deleting new journals; from the latter the management of press registered users, user groups and statistic plugins, as well as the control of file browsing functionality within the press.

IV. Applications and Results

The extended multi-press OJS framework has been used to realize an experimental web portal, which represents the main technical activity of the *Palamede* research project.

The outcome of this experimental stage of the project is available online [19]. Some illustrations of the new features of the web platform are presented in Figure 3. The application of multi-press OJS in the Palamede portal has been validated through an experimental activity carried out by the three University Presses taking part to the project (see Section II.A). The aims of this experimentation have been the creation of three public presses and four journals, as summarized in Table 3. Press managers have edited their own press layout and, for each Journal, at least one issue has been published.

The whole peer-review editorial review has been tested by new registered journal managers, editors, section editors, authors, reviewers and proofreaders. All users playing those roles have been qualified users already using the OJS and other open journal systems. Here follows a final numerical balance of fulfilled editorial actions in the experimental activity of *Palamede* portal:

- **4 Presses** have been created: three published by the involved University Presses and one created by the DISIT Lab for testing purposes, which has not been published;
- **5 Journals** have been created: four published by the University Presses (see Table 3) and one by DISIT for tests;
- A total of **4 issues**, among all the published journals, have been produced.

V. Conclusions And Future Work

The goal of reengineering the original OJS framework and creating a new multi-press platform has been successfully accomplished. This result has been positively received within the University Presses and the partners taking part to the *Palamede Project* experimentation, and moreover within the PKP community: actually the presented work has been announced in the PKP development forum [20], and it received positive comments, with the perspective of evaluating (and possibly redistributing) the source code of a final, full functional version. Regarding this topic, some guide lines for short-term future work have been already planned. They mostly concern search methods improvement, since the aggregation of several presses, dealing with different research areas and topics, requires a more efficient search engine.

Future implementations can be grouped into the following main activities:

- semantic indexing for search engine optimization and fuzzy queries;
- production of statistics for queries;
- techniques for monitoring number of accesses, downloads etc.;
- alternative payment methods for subscription and journal fees (actually, OJS natively supports PayPal only).

Table II. Browser-like table listing all the new Smarty templates (.tpl files) included in the extended multi-press OJS, grouped by system installation folders. New created folder names are displayed in italic style.

New Smarty Templates included in multi-press OJS (Browsed by OJS installation folder)		
<OJS Install Folder>\Templates		
About	indexPress.tpl	
	press.tpl	
	pressContact.tpl	
	pressGroups.tpl	
	pressHistory.tpl	
	pressStatistics.tpl	
Admin	presses.tpl	
	pressJournals.tpl	
Index	press.tpl	
PressManager		
	PressEmails	pressEmails.tpl
	PressFiles	index.tpl
	PressGroups	pressGroupForm.tpl
		pressGroups.tpl
		pressMemerships.tpl
		selectPressUser.tpl
	PressPeople	pressEnrollment.tpl
		pressEnrollSync.tpl
		pressUserProfile.tpl
		pressUserProfileForm.tpl
		searchPressUsers.tpl
		selectMergePressUser.tpl
	PressPlugins	managePressPlugins.tpl
		pressPlugins.tpl
	PressSetup	index.tpl
		pressSettingsSaved.tpl
		pressSetupHeader.tpl
		step1.tpl
		step2.tpl
		step3.tpl
	PressStatistics	index.tpl
		pressReportGenerator.tpl
		pressStatistics.tpl
	index.tpl	
	pressLanguageSettings.tpl	
PressUser	index.tpl	

Furthermore, the following longer-term future developments have been considered:

- acquisition of instruments in order to integrate the actual OJS framework with models for automatic management of different content types (including media) and metadata. A useful tool to achieve this purpose is represented by the AXMEDIS Content Processing Solutions (AXCP) [21]. AXMEDIS [22] is a large project developed by the DISIT Lab in its recent past activity, in collaboration with several important international partners;
- multi-platform distribution: PCs, mobile devices, PDAs, iPhone, iPad, Android etc.;
- solutions for social networking and distance learning.

TABLE III. SUMMARY OF PUBLISHED PRESSES AND JOURNALS IN THE EXPERIMENTAL ACTIVITY OF THE *PALAMEDE* PORTAL.

Presses	Journals
Università di Parma	• Papyrotheke
FORUM Editrice Universitaria Udinese	• Popolazione e Storia • Plurilinguismo. Contatti di Lingue e Culture
Firenze University Press	• Annali del Dipartimento di Filosofia

ACKNOWLEDGMENTS

The authors would thank the partners of the Palamede project: Florence University Press, Fondazione Rinanscimento Digitale and the universities involved in the trial and validation phases. A particular thanks to Prof. G. Mari, Dott. P. Cotoneschi, Prof. A. Tammaro, Dott. M. Rufino, Dott. M. Lunghi, Dott. R. Petrini.

REFERENCES

[1] PKP – Public Knowledge Project. [Online]. Available: http://pkp.sfu.ca/
[2] GNU – General Public License. [Online]. Available: http://www.gnu.org/licenses/gpl.html
[3] Dspace software. [Online]. Available: http://www.dspace.org/
[4] Fedora Commons – Flexible Extensible Digital Object Repository Architecture. [Online]. Available: http://www.fedora-commons.org/
[5] Duraspace Project. [Online]. Available: http://www.duraspace.org/
[6] EPrints – Open Access and Institutional Repositories. [Online]. Available: http://www.eprints.org/
[7] SHERPA Project. [Online]. Available: http://www.sherpa.ac.uk/
[8] EScholarship – University of California. [Online]. Available: http://escholarship.org/
[9] Daedalus – a JISC FAIR Project. [Online]. Available: http://www.lib.gla.ac.uk/daedalus/
[10] J. Willinsky, K. Stranack, A. Smecher, and J. MacGregor, “Open Journal Systems: a complete guide to online publishing”, 2nd ed. for OJS 2.3.3, Simon Fraser University Library, Sep. 2010. [Online]. Available: http://pkp.sfu.ca/ojs/docs/userguide/2.3.3/userguide.pdf
[11] A. Smecher, “OJS technical reference: Version 2.1”, 4th ed., Simon Fraser University Library, Jun. 2010. [Online]. Available: http://pkp.sfu.ca/ojs/docs/technicalreference/2.1/technicalreference.pdf
[12] “OJS API reference v2.2.1”. [Online]. Available: http://pkp.sfu.ca/ojs/doxygen/html/index.html
[13] FRD – Fondazione Rinascimento Digitale. [Online]. Available: http://www.rinascimento-digitale.it/cont_E.phtml
[14] DISIT Lab – Department of Systems and Informatics, University of Florence. [Online]. Available: http://www.disit.dsi.unifi.it/
[15] FUP – Firenze University Press. [Online]. Available: http://www.fupress.com/eng/
[16] Università di Parma. [Online]. Available: http://www.unipr.it/
[17] Forum Editrice Universitaria Udinese. [Online]. Available: http://www.forumeditrice.it/
[18] OAI-PMH: Open Archive Initiative Protocol for Metadata Harvesting. [Online]. Available: http://www.openarchives.org/pmh/
[19] Palamede Research Project. [Online]. Available: http://palamede.fupress.com/
[20] PKP Forum, OJS Development. [Online]. Available: http://pkp.sfu.ca/support/forum/viewtopic.php?f=9&t=6815
[21] AXMEDIS Content Processing Solutions. [Online]. Available: http://www.axmedis.org/com/index.php?option=com_content&task=view&id=94&Itemid=34
[22] AXMEDIS Project – Automatic Production of Cross Media Content for Multi-Channel Distribution. [Online]. Available: http://www.axmedis.org/

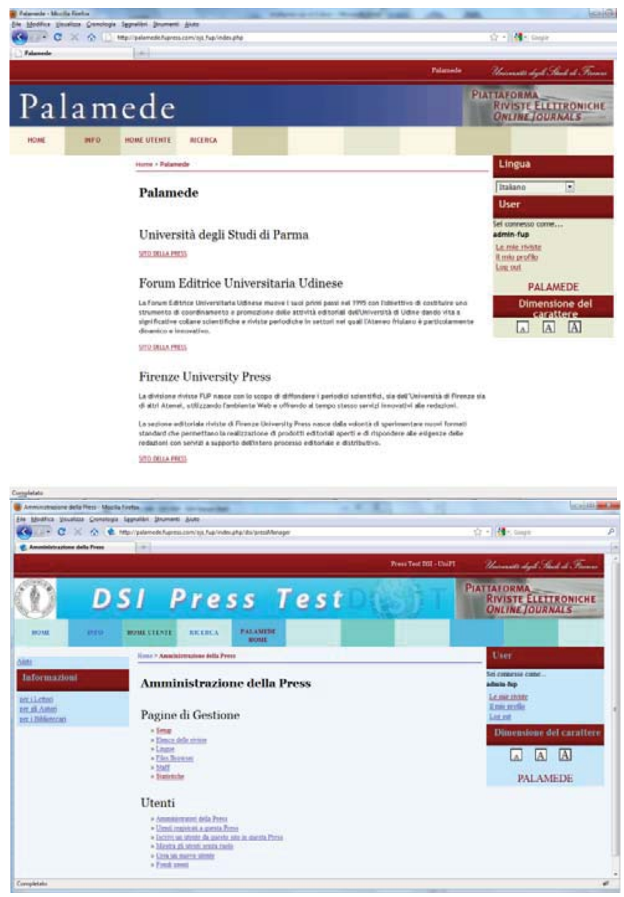


Figure 4. Home page of the *Palamede* portal. Example of Press Management page and customized graphical layout: *DISIT-DSI Test Press*.

A Reference Context Model for Development of Security Systems

Joachim Hansson, Rego Granlund, Niklas Hallberg, Fredrik Lantz, Erland Jungert
FOI (Swedish Defence Research Agency)
Box 1165, S-581 11 Linköping, Sweden
{joachim.hansson, rego.granlund, nikha, flantz, jungert}@foi.se

Abstract: The security systems domain is complex and involves a number of aspects of concern. For instance, it is often unclear what initial aspects and prerequisites that are needed to be explored in the development of security systems. To overcome this, systematic and methodological approaches to document a common view of the context are needed, to enhance the development of security systems, i.e. to make them more fit for use. The objective of this paper is a reference context model that supports context analysis, with the aim to determine the borders that divide the context from the system to be developed and to describe vital prospects and restrictions of the context. The reference context model consists of a series of views that include information needed to determine the preconditions required for the following steps in the systems development that is to explore stakeholders' needs, determine the requirements, and development of the systems architecture. The development of the reference context model requires determination of contextual objects and their relations. Required is also participation from stakeholders as their knowledge and experience from the domain are essential to ensure the quality of the model. The different views, which are extendable, forms the context model and are further discussed in this paper.

Keywords: Context model, security systems, systems development,

1. INTRODUCTION

The development of technology intensive security systems has been subject to increasing attention during the last decade. Even thought there are a number of success stories, several development efforts have failed. Many of the failures in systems developments have their origin in the early phases of the development [1]; in many cases due to purely technology driven development. Development of security systems is complex and there is a need for approaches that takes its starting point in the stakeholders' needs and in additional prerequisites of the context of use. Hence, there are needs for development of approaches that supports the exploration of the context, in which the developed systems should operate, into account. This includes identification and establishment of a shared understanding of preconditions required for development of security systems.

In this work, security systems refer to systems applied to the protection of, stationary as well as mobile, objects that need protection from various threats and antagonistic attacks. Such security systems require capabilities for surveillance and incident management [2]. Generally, security systems have to support (1) surveillance of external events, (2) decision-making, (3) task execution, and (4) communication of tasks and information [3]. Security systems also need to be able to operate effectively and not disturbing normal activates.

Thus, to support the accomplishment of improved understanding of security systems a *reference context model* (RCM) is proposed. RCM constitutes a structure supporting context analysis as a basis for the development of security systems. The outcome of the context analysis is a context model aimed at being used to guide the subsequent needs assessment process [4]. The context analysis and its relations to RCM and the systems development process is illustrated in Figure 1. Modeling is an important mechanism to be able to understand complex phenomena [5]. Hence, in systems development it is a useful technique to explore the context in focus, determine requirements, and communicating with the stakeholders [6] [7] [8]. The RCM prescribes a number of views, i.e., interest of concerns of the model [9]. The characteristic of the RCM is that it guides the context analysis in obtaining project specific context models. Hence, the purpose of the context analysis is to extend the reference model so that it becomes project specific, which is a necessity for adaptation to the application in focus.

The outline of this paper is as follows. Section 2 presents the objective of the work. In section 3 the seven different views of the reference context model are described and discussed. In section 4 related works is discussed and finally in section 5 the conclusions of the work are presented.

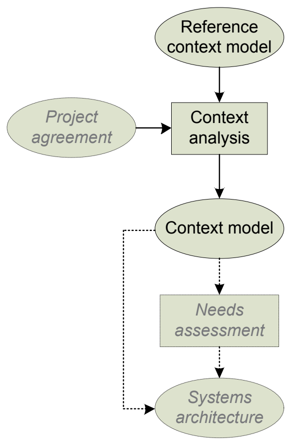


Figure 1. The context analysis and its relation to the needs assessment and system architecture development.

2. OBJECTIVES

The objective of this paper is to present a support to *context analysis,* in form of a *reference context model* (RCM) that includes stakeholders, governing documents, co-existing systems, environments and relationships between them. Further, an initially important task in the context analysis is to clearly establish an understanding and documentation of the borders that divide the context from the system to be developed. Thus the RCM becomes an important input for subsequent needs assessments.

The main motivation for the reference context model is that it supports the development of project specific context models. A context model is important since it promotes consensus between developers and customers concerning the limits of the work in focus and it enhance the needs assessments. Thus an important factor in the context analysis phase is to involve the customers in a dialogue to eventually reach consensus in what concerns the content of the context model.

3. THE REFERENCE CONTEXT MODEL

In this section the views that constitute the RCM are described. This includes seven different views types including some additional considerations. These views are concerned with:

- Objects of concern to the problem domain including their hierarchical structure and definitions.
- The stakeholders, categorized with respect to whether they have a direct or an indirect impact on the design of the security system.
- The stakeholders' impact areas (SIA) and the definitions of those areas.
- The prioritized relations between the stakeholders and the impact areas.
- The design basis threat (DBT).
- Governing documents.
- Identified interviewees.

3.1 Model structure

The model structure includes the objects of relevance to the domain of security systems. Thus, a primary work has been to first identify these objects and second to prioritize them with respect to the application in focus. On top level, the hierarchical object structure includes seven object classes (Figure 2). The hierarchical structure is similar to an ontology and will in many applications contain up to a couple of hundred different objects.

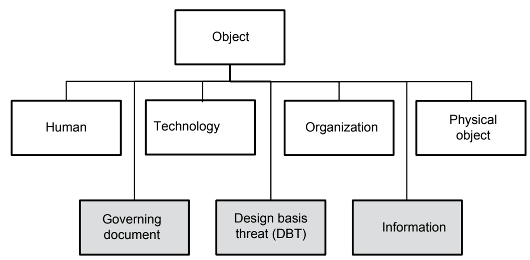


Figure 2. The top level of the object structure.

In Figure 2, the gray shaded objects concern objects that have been given a high priority while the white rectangles are unprioritised. Further down in the structure the members of the, so far, unprioritised branches are given relevant priorities. The former concerns *DBT* that is further discussed in section 3.5 and *governing documents* in section 3.5. *Information* will not be discussed further as it to a high degree is application dependent and generally concerns information that will be communicated between engaged parts. Examples of information are priory data and data gathered from various sources such as sensors. Thus information is basically something that must be determined during the needs assessment and capability determination processes.

The *Human branch* concerns stakeholders that are discussed subsequently. *Technology* contains systems and functions currently existing, which may or may not be subject to replacement by newer systems. In Figure 3 some selected objects from this branch of the structure are presented.

The *Technology* branch contains *Physical protection* which, including its subsequent parts, all have been given high priority whereas *Operation system* has been given a low priority since it has a very low interest when considering security aspects. *Intelligence* has been given a medium priority here, although this may vary from application to application.

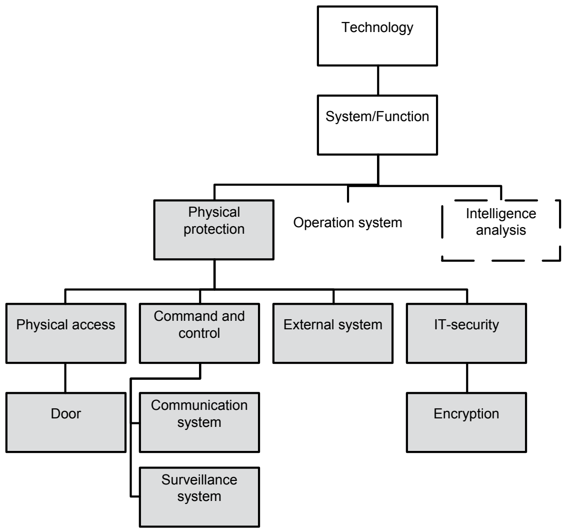


Figure 3. The *Technology* branch of the object structure.

The *Organization* branch may involve a large number of organizations of which some may be of supportive character that may be engaged in various incidents. Examples of such organizations are the police department, the fire department and medical care providers. In many other applications will organizations like counties, communities and other more specialized governmental organizations also have an impact on the final security system.

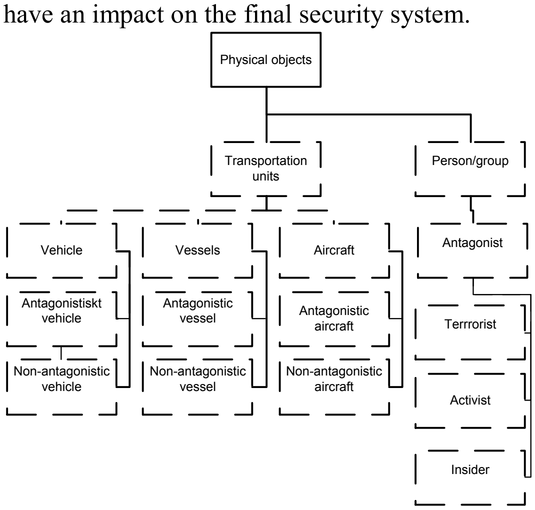


Figure 4. The *Physical* object branch of the object structure where all objects have been given a medium high priority.

Furthermore, there is also the *Physical object* branch, which may grow to a more substantial size depending on the application in question. This branch may include various antagonistic types such as *terrorist*, *activist* and *insider*. A further sub branch may contain object types like *vehicle* and *aircraft*, which may be controlled either by antagonists or co-operative agents or other organizations. Parts of this rather large branch can be seen in Figure 4. Clearly, the transportation units can be extended much further including, for instance, *truck* and *car* but of course this depends on the users´ requirements. Also, in this illustration all the objects have been given a medium priority but that may not always be the case.

In addition to the model structure a list of definition of all objects appearing in the structure has been assembled, basically to support the stakeholders and to avoid confusions in the discussions between systems developer and customers and end-users.

3.2 Stakeholders

An important activity in the context modeling is to identify those stakeholders that may or can have an impact on the final security system. Stakeholders can be identified from documents supplied by the customers or through various dialogues with the customers. In some cases, external documents concerned with the actual application area can and must be sorted out. It is, at any rate, important to identify a broad spectrum of stakeholders to get a comprehensive picture of the stakeholders and their situation to make sure that none of them or groups of them is neglected; this is of great importance for the determination of the stakeholder/SIA relations' matrix; described in section 3.4.

Identified stakeholders are eventually grouped into three different levels. Level one includes basically end-users of the system subject to development while level two includes stakeholders that will have a direct influence on the systems design. Level three, finally, includes stakeholders that just will have an indirect influence on the final system. The grading of the different stakeholders is carried out in a dialogue between the customers and the system designers.

Examples of stakeholders on the three different levels are (Level 1) end-users, e.g. operators, (Level 2) Security managers, IT-managers, representatives from crisis management organizations and (Level 3) Counties, Police etc. This is not a complete list but can nevertheless be seen as a small group of potential stakeholders.

3.3 Stakeholders' impact areas (SIA)

When all stakeholders are identified the next activity is to elicit the stakeholders with respect to those parts of the security system which may be of particular concern to each one of them. Thus, the elicitation aims to select a variety of stakeholders that (1) can have an impact on the security system, (2) will represent the three different levels of the stakeholders and (3) will cover all aspect of the security system of concern to the actual application. Elicitation of stakeholders from the different levels is important since the stakeholders on different levels have different needs [1]. To ensure that relevant stakeholders are elicited, because of the correct, reasons the security systems' domain need to be subdivided into different areas, so-called stakeholder's impact areas. In this work a number of some general and different stakeholder's impact areas are identified, for example:

- Protection against intruders,
- Surveillance system,
- The security system's functionality during incidents,
- Levels of preparedness,
- Routines for in and out passage (access management),
- Intelligence.

When identifying the different SIAs aspects concerning physical protection as described by, e.g. Garcia [10] must be considered. Of importance is also to gain a joint understanding between system designers and customers. Hence, subdividing the security systems domain into relevant impact areas and the description of each impact area are of importance. This part of the context modeling is done in dialogue between the system developer and the customers. Stakeholders' relations to the different impact areas are described further in the next section.

3.4 Stakeholder impact area relation matrix

Different stakeholders have interest and knowledge in different impact areas. Some stakeholders have interest and knowledge in just a single impact area while other stakeholders can have interest in most or sometimes all impact areas. Although these relationships are relatively simple, the identification of each stakeholder's relation to an impact area can be difficult to determine. Even the amount of relations will add complexity and lead to difficulties. An approach to deal with these difficulties is to create a relationship matrix. The matrix may connect any single stakeholder to any of the corresponding impact areas.

A first step to identify each stakeholder's relations to the impact areas is to request the customers to give a simple *yes* or *no* to the question whether a certain stakeholder has a relationship to be specified within the SIA. The answers can be used to determine whether a stakeholder posses a relational position and have knowledge to affect the corresponding impact area. Thus, the first step is to identify each stakeholder's impact area, the next step will be to prioritize each identified relation. Each relation, given the answer yes in the first step, will in step 2 be given a priority, 1-3, where 1 is the least significant and 3 is a relation with the highest significance. The outcome of step 2 is a complete matrix with priorities. The advantage of the matrix is that each stakeholder can now be summarized and compared against each other. Stakeholders with a high number will indicate their essential knowledge to the system. However, stakeholders with a low number will not directly be neglected. Stakeholders with a low number can have special knowledge in some impact areas and therefore still be of interest. The result of the matrix will give a good hint of the main stakeholders but will still demand some more analysis. A help in this analysis is to give each impact area its own priority, to indicate if there are some areas that are of higher or lower interest.

Table 1 illustrates a suitable example of the stakeholder impact area relation matrix. As mentioned, the stakeholders are grouped according to their group membership level. The stakeholders in red are considered to be of no interest to the application in focus. The remaining stakeholders are considered relevant for the future work although not everyone for every impact area. Thus, the gray impact cells in the matrix are of no or very little interest while the white cells have been given different priorities. For instance, the operator has been given a medium high priority (2) for the *Surveillance system* impact area.

At the top of the matrix the priority of each impact area can be found. Here, the priority of the *Surveillance system* impact area is 3. Furthermore, in this example only the security manager has been given high priority with respect to all the impact areas. A conclusion is that the matrix is clearly extendable both with respect to stakeholders and to impact areas.

Table 1: Example of the stakeholder impact area relation matrix.

	3	3	2	1	1	2
Stakeholder	**Protection against intruders**	**Surveillance system**	**Functionality during incident**	**Routines for in and out pass**	**Intelli-gence**	**Levels of prepared-ness**
Level 1 - Users						
Operator		2				
Backup operator		2				
Level 2 - Direct impact						
Crisis management						
Security manager	3	3	3	2	1	3
IT-manager						
Operations management		1				1
Fire department						
Police department					2	2
Level 3 -Indirect impact						
County						
Governmental security agency	1					
Medical organization						

3.5 Design basis threat

The objective of security systems is often linked to existing threat descriptions, which often are scenario-based. These descriptions are important in the design and they should be seen as dynamic, which means that they may and will change over time. The threats descriptions will continually be updated as a result of changes in the social climate and/or technological development, and the needs of the security system, including the operating organizations, should be updated to meet the new threats as well. This leads to the introduction of the *design basis threat (DBT)*, which normally is used as an estimation of a possible scenario.

The design basis threat is, according to IAEA [11], defined as “A fundamental principle of physical protection is that it should be based on the State’s current evaluation of the threat”. Thus DBT corresponds to a threat that should be used to determine the capabilities and capacities of the final security system. An illustration to this could be that if one expects intruders not to manage to climb over fences higher than 5 m then the designers of the system should build fences that are higher than 5 m to get a sufficient physical protection of the object. Another example is that if intruders are expected only at daytime the sensors could be equipped with just ordinary video cameras, which exclude the use of IR-sensors.

The DBT plays a central role in designing security systems. Especially when it comes to the determination of the degree to which the system should be able to respond to when facing a threat of any kind.

3.6 Governing documents

An important aspect of contextual analysis is to identify which documents that should be considered when designing security systems. Such documents can be divided into two classes, i.e. laws and statutes, documents describing existing procedures.

3.6.1 Laws and regulations

Examples of laws and statutes that must be followed are *national laws*, *government regulations* and *international agreements*. The national laws describe the requirements imposed on the operators that are responsible for the protection of the facilities and their personnel. Health and safety laws controlling the conditions under which staff can work within the security system. National regulations defined by the authorities clarify the laws and give recommendations on how laws should be followed. International agreements are defined both in unions like the EU and bilaterally between individual countries. These agreements are often clarified in the regulations defined by national authorities. Both national authorities and international organizations may inspect security installations with regard to existing laws, regulations and agreements.

When designing security systems experts in each area should be used to guarantee that laws and regulations are followed. It is therefore important to identify which laws and regulations that are affecting the system at an early state of the development process, which consequently can be done as part of the context modeling process.

3.6.2 Existing procedures

To get an understanding of how current and resembling security systems operate existing procedures in such systems can be studied. These procedures often describe how security systems should be operated in order to maintain reliable protection of the facilities. These procedures will most often give an understanding of the interactions that may take place between actors and organizations that manage and depend on the security systems. The existing procedures need to be studied when e.g. planning interviews of stakeholders so that the interviews can identify strengths and weaknesses of existing security systems, organizations and security cultures.

3.7 Interviewees

The next following view to be determined in the reference context modeling process is to lay the ground for the data collection, which basically is concerned with interviews of stakeholders to determine their needs [4]. Identification of interviewees can basically be resolved by means of the stakeholder impact area relation matrix.

To determine which stakeholders to interview and on which topics they should be interviewed is a serious question that need to be determined accordingly. However, as an outcome of the context analysis this can be determined in a fairly simple way by considering the content of the stakeholder/SIA matrix. From this matrix, the stakeholders with the highest priorities are selected to be interviewed about those impact areas for which they have been given the highest priority. This is a simple and relatively forward going activity, which also implies possible changes that can be subject to further discussions. For instance, if one decides to drop an impact area then this could mean that a certain stakeholder also may be dropped if she is just subject to be interviewed about that particular impact area.

Determination of interviewees can be made in two ways either by the designers alone or in a dialogue between the designers and the customers.

4. RELATED WORKS

Works related to the approach take here, i.e. to support the identification of the context for a given application area in the process of systems development, are spars. However, what have been found in the literature and which relates the work described here concerns basically the focus of the stakeholders during early stages of the systems development.

To focus on the viewpoints from the stakeholders at an early point is an elemental trend in system modeling and design. It is a strategy that has been commonly used in software systems design and information systems research. Researchers have been developing taxonomies and methods for modeling the different stakeholders view and demand on a system. A useful and classic model is the Onion model that can be used to indicate the hierarchical sphere of influence and boundaries [12]. It is important in the early phases of system analysis and system requirements to identify the stakeholders that should participate in the analyze process and what roles and involvement they should have in the participatory design process [13].

According to Yu et al. [14], models can be used to describe, understand and analyze a complex world and that the complexity of the social world presents formidable challenges for social modelling, which in some senses relates to the approach taken here; especially as it is concerned with the early approaches to systems development and requirement engineering. Furthermore [ibid], it is also argued that "to begin, we need to adopt a different starting point for understanding the world in which software and information systems are situated". A further work that also relates to the development of security systems is discussed by Elzinga et al. [15].

5. CONCLUSIONS

In this paper, aspects initially of importance to the development of security systems have been in focus. Among these aspects a particular concern is how to clearly establish an understanding and documentation of the borders that divided the context of the problem domain from the system to be developed. With this in mind, a reference context model has been developed and presented as a framework identifying the considerations that must subject to consideration during the systems development process. The application in focus here is security system, which is a comprehensive field, where stakeholders, governing documents, related systems, environments and relationship between them play fundamental roles for the final outcome of the systems development process. The reference context model described in this work addresses these aspects. The approach is also flexible and may thus be adapted to other application areas as well, e.g. military command and control systems. Especially such areas where a complex context is influencing the development process the need for a reliable context model is obvious.

A final observation of this work is also that customer participation in development of the RCM is of importance as otherwise the risk for development of a context model that includes information of no use and/or lack important information is obvious.

REFERENCES

[1] Kasser, J. E, A framework for understanding systems engineering, The Right Requirement Ltd., 2007, Cranfield, England.

[2] Jungert, E., Grönwall, C., Hallberg, N., Kylesten, B., Lantz, F., Eriksson, L. A Generic Architecture for Surveillance Systems, Proceedings of the 16th International Conference on Distributed Multimedia systems, Chicago, Il, October 14-16, 2010.

[3] Jungert, E., Hallberg, N., and Hunstad, A., A Service-based Command and Control Systems Architecture for Crisis Management, The Int. Journal of Emergency Management, Vol. 3, No. 2, 2006, pp 131-148.

[4] Hallberg N, Timpka T, and Eriksson H. The Medical Software Quality Deployment Method. Methods of Information in Medicine. Mar; 38(1):66-73.

[5] Jacobson, I., Ericsson, M., Jacobson, A., The object advantage: Business process reengineering with object technology, Addison-Wesley, Reading, MA, 1995.

[6] Hay, D.C., Requirements analysis: From business views to architecture, Prentice Hall, Upper Saddle River, New Jersey, 2003.

[7] Hull, E., Jackson, K., Dick, J., Requirements engineering, 2nd edition, Springer, New York, 2005.

[8] Carlock, P.G., Fenton, R.E., System of systems (SOS) enterprise systems engineering for information-intensive organizations, System Engineering 4(4), 2001, pp 242-261.

[9] IEEE 1471, Architecture Working Group, IEEE Recommended Practice for Architectural Description of Software-Intensive Systems (IEEE Std 1471-2000, IEEE, 2000).

[10] Garcia, M. L., The design and evaluation of physical protection systems, Elsevier, Burlington, MA, 2001.

[11] http://www-ns.iaea.org/security/dbt.asp?s=4

[12] Alexander, I. A Taxonomy of Stakeholders: Human Roles in System Development, In Journal of Technology and Human Interaction, 1, 1, 2005, pp 23-59.

[13] Vinka, P., Imadac, A. S., Zink, K. J. Defining stakeholder involvement in participatory design processes, In Applied Ergonomics, 39, 4, 2008, pp 519-526.

[14] Yu, E., Giorgini, P., Maiden, N., Mylopoulos, J. (eds) Social Modeling for Requirements Engineering, Cambridge, MA: MIT Press. 2011.

[15] Elzinga, P., Poelmans, J., Viaene, S., Dedene, G., Morsing, S., Terrorist threat assessment with formal concept analysis, 2010 IEEE International Conference on Intelligence and Security Informatics (ISI), Vancouver, BC, Canada, 23-26 May 2010, pp 77 – 82.

Crossing the Digital Divide: Design considerations for "all-inclusive" online language training

Don Maybin
Shonan Institute of Technology (Japan)

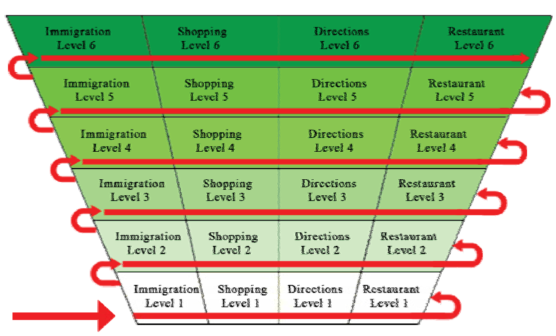
Abstract
The evolution from classroom to online of a language training website for developing basic communication skills in a foreign language with a very broad user base is briefly described. A description of the structure and teaching principles of the approach is followed by a discussion of the adaptation of the training system to an online application. Particular reference is made to the *digital divide*, which refers to the gap between those who have access to information technology and those who do not for various reasons, such as gender, income, age, education or physical location. Design measures taken to overcome such limitations in order to reach a wider audience are then described.

Introduction
This paper is a report on the adaptation of a classroom-based accelerated approach to language teaching, called *Action-Based Language Empowerment*, or ABLE[1], to an online training format, referred to as c-ABLE for *Computerized Action-Based Language Empowerment*. The online course is a short, intensive introduction to a target language, or L2, that enables learners to function autonomously in the L2, equipped with a knowledge of forms, communication strategies and a sense of self-efficacy that promote further learning and participation in the language[2].

Course Structure
In a traditional language training system, lessons are divided into thematic areas with one topic taught per lesson (e.g. Lesson 1: Introductions; Lesson 2: Asking Directions; etc.). This approach is reinforced by traditional formats, which, for convenience or lack of an alternative approach, are normally structured in units based on these themes. Such training materials are "vertical" in nature, self-contained and normally not repeated. For the majority of learners, unfortunately, this study arrangement is relatively ineffective. Unless material is constantly and consciously recycled or reviewed on the learner's initiative, previously studied language is forgotten as the course progresses[3].

The c-ABLE curriculum is divided into multiple levels with each functional area divided into compartments, or thematically-linked *sub-modules*. Users complete a set of *sub-modules*, for a specific function at one level then proceed horizontally to the sub-module set for the next functional area at the same level (Figure 1). Once all thematic modules at one level are completed, the user proceeds to modules at the next level, where material is recycled and expanded in a spiraling format. At the end of the course, all functional areas have been touched on repeatedly and recently. Language has been reiterated and reinforced, and ideally remains in the user's memory, ready for immediate application. This recycling format, with repetition of vocabulary, formulaic expressions and forms, incrementally enhances usage by users.

Figure 1: c-ABLE "horizontal" lesson arrangement and spiraling format



[1] During its development, ABLE has been used to teach some 350 learners in 16 target languages. A comprehensive list of courses and related workshops dating from 1979-2000 is given in Maybin (2002).

[2] For a description of an ABLE course used for training in Thai, see Maybin and Batten (2008)

[3] The *forgetting curve* indicates the decline of retention in memory over time. Memory of specific information is halved in a matter of days unless consciously reviewed (Ebbinghaus, 1885; Schacter, 2001)

The modular, spiraling format also results in a frequent change of topic. This frequency helps users maintain a high level of concentration throughout a study session. Learning is fast-paced and tightly organized with modules that keep users engaged, an advantage for those with a short attention span or low tolerance for long sessions focused on one theme[4].

c-ABLE uses a highly focused syllabus to cover 18 functional areas. The course introduces users to approximately 1,900 semantic units repeated at scheduled intervals in a variety of realistic situations. It consists of a large number of modular units each taking only a few minutes to complete. These modules are arranged in a spiraling format; users return repeatedly to typical situations with useful concepts and vocabulary, thus increasing confidence. Grammar, per se, is not featured; communication is. The emphasis is on native-like pronunciation and use of pragmatic moves that keep a conversation going, including conversation management strategies, such as asking for repetition.

Teaching Principles of c-ABLE

Apart from structure, the c-ABLE curriculum differs in many ways from typical pedagogical approaches, including:

1. Users select key functional goals from a menu of optional topic areas. The result is a tailored study experience which gives them a vested interest in successfully completing a course.

2. Although it is common practice in most pedagogical approaches to use the learners' L1 to soften the impact of the learning experience[5], in the case of c-ABLE such assistance is viewed as a crutch, detrimental to developing comfort with the L2. Directions for the user interface are given in the learner's L1; however, instructional materials are presented entirely in the target language.

3. The c-ABLE pedagogy focuses on the development of listening followed by speaking skills. Users are asked to devote their whole attention to activating aural memories of the sound-meaning (visual) combinations presented and told not to speak until they have reached the third level of their course. An initial silent period is required to allow for processing of target language phonemes, stress, tones, overall sentence intonation and other speech features unique to the spoken form of the L2[6]. The end result of c-ABLE is that most users sound more native-like when they do begin to talk.

4. There is no overt explanation of grammar. When key expressions are introduced, particularly those for carrying out specific communication strategies (see item 6 below), the language is presented as *formulaic*[7], introduced either as a complex unit or partially analyzed unit (e.g. "Please give me ___."). Users internalize these formulas through repeated exposure made comprehensible in clear, function-based contexts. *Subconscious proficiency* is the goal.

5. Core vocabulary (words and "chunks") are arranged in thematic groupings of three to six semantic units[8] based on relevance and frequency in the L2. Words are contextualized and presented in useful phrases in the latter stages of each sub-module to develop the user's ability to recognize and apply them in a *stream* of language.

6. There is an expectation that users will attempt to communicate in the L2. Key language for carrying out communication strategies is primarily derived from the *Control* model[9] and presented as formulaic expressions, including interrupting ("Excuse me."), asking someone to speak slower ("More slowly, please."), and clarifying meaning ("What does that mean?"). These expressions for conversation management are introduced through icons which allow users to access the necessary formulaic expression when there is a perceived need (i.e. contextualization), while simultaneously exposing the user repeatedly to a speech model for imitation, a form of "subliminal training" through inundation.

4 Johnstone and Percival's research (1976) indicates student attention patterns during lectures decline steadily based on a range of variables, including lecture theme and length of presentation.

5 As countenanced by Cook (2008, pp. 170-174).

6 Saville-Troika (1988) describes aspects of this silent period in the early stages of language development in children.

7 Referred to as "formulaic language" in Wray (2002).

8 Miller (1956) suggests human recall is more efficient when information is introduced in groupings of seven plus or minus two items (i.e. 5 to 9 items). In order to fall safely within this range, new information is introduced in groupings of six items in the ABLE curriculum.

9 Maybin and Bergschneider (1992)

The Digital Divide

The goal of c-ABLE is to be as all-inclusive as possible. In order to reach a broader audience, there is a need to address issues arising from the digital divide, defined simply as "…the troubling gap between those who use computers and the Internet, and those who do not"[10]. The reasons for limited access may include not only perceived minimal online skills, but also gender, income, race, geographical location, age or education, among others. This paper will focus on each of the above digital divide factors framed around design elements which have been or should be developed for c-ABLE to encourage greater access for those who have been discouraged by the technology.

Gender

It has been suggested that "one-half of the 'digital divide'… is fundamentally gender related."[11] With regard to online language training, tailoring is vital. c-ABLE has mechanisms that allow users of either sex to select functional areas relevant to their needs. A fundamental challenge has been to establish a system that allows for differences in male and female speech. The site's sign-up profile includes gender identification, which lines up the course with the user's sex. Instruction is given exclusively in the language form which matches the user's gender for three course levels before the user hears language of the opposite sex.

Income

Income has a major influence on the viability of e-learning. An individual with limited funds is less likely to own a computer or be able to afford payments for regular access to one (e.g. via an Internet café). "Prohibitive" fees can be small in some regions of the world. Three possible solutions to directly address this hurdle include:

- a free course that gives users a sense of the effectiveness of the training approach so they can make an informed decision as to whether to commit limited funds to a pay-for version, as well as develops rudimentary communication skills in a limited number of functional areas
- low fees to reduce the financial burden on those who have tested the free introduction course and wish to continue studying
- a smartphone application for those with limited online access.

Race

This paper interprets "race" in broad terms, using it to identify specific populations which speak "non-mainstream", possibly endangered language***s***, thus limiting their access to online study. In order to reach such populations, it is necessary to offer a user interface in the native language or, where funds are limited, to provide an "access language" that a majority of the population is familiar with (e.g. Spanish for Mayan speakers). To provide global access, each language entered in the online system ***is*** uploaded as a language to study and to study *from*, an interactive language "mesh". This entails the development of a specific platform, which allows for the uploading of multiple courses with the addition of each new language.

Location

Although major efforts are being made to allow global online access to all regions of the world[12], the reality is that even in more developed countries accessing a language training website on the Internet can prove problematic, particularly where Internet connections are unstable. Development of a smartphone application could potentially overcome this hurdle by providing downloadable files for offline study; however, given the interactive language "mesh" platform suggested above, an offline feature has its drawbacks since deletions or additions would require revisions be carried out one-by-one for each course. An alternative is to synchronize content for the courses so that each time a user "renews" his or her download for offline study, the course is automatically updated.

Age

The Internet offers endless opportunities for lifelong learning, yet the majority of sites concerned with language training focus on a youthful demographic overlooking the needs of older learners. Age is negatively associated with users Internet skills[13], while seniors intimidated by online technology are hardly encouraged by the complex user interface of many sites, which consider hidden menus, hovering and scrolling down as basic functions that everyone is familiar with. Common sense dictates that any interface should be user friendly, but where mature users are concerned special attention must be given to making the UI as obvious and intuitive as possible.

---

[10] Mehra et al., (2004, p. 782)

[11] Bimber (2000)

[12] For example, AMD's (Advanced Micro Devices) "50 X 15" project aims to provide Internet access to 50% of the world's population by 2015.

[13] Hargittai (2002)

Any site hoping to satisfy a broad demographic must develop *user empathy*. In the case of older users, this can be as simple as including high frequency phrases for the age group (e.g. "I'm retired."), visuals showing mature individuals as more than "grandparents", and including older voices in the sound data. To attract mature users, c-ABLE has incorporated all of the above, while beta testing has been carried out with seniors to ensure design assumptions targeted at this demographic are correct.

Education

Too often language study connotes academia. Language training can be elitist and, for those with limited formal education, an intimidating prospect restricted to "specialists". In the case of c-ABLE, a conscious attempt has been made to keep the pedagogy, as well as course content, as all-inclusive as possible. Traditional pedagogical tools, such as overt explanation of grammar or disassociated word lists, are eschewed. The focus is communicative and practical with emphasis placed on completing specific functions. The intention is that c-ABLE is attractive to users from a range of ages and academic backgrounds.

Website Status

The transition from classroom to online has not been without difficulties. Given the varied demands of an "all inclusive" website which is equally functional across a broad spectrum of languages and a wide user base, complications have and will continue to arise in the development of c-ABLE, particularly with regard to the multilingual nature of the site.

Sophisticated back-end tools have been developed for the translation and inputting of new languages; however, due to cultural and linguistic differences, each new language has required revisions, some extensive, to the tools and interactive language "mesh" platform itself. User feedback is of great importance in evaluating the site. For example, indications that many users would like access to written text pushed us to create a "language directory" to satisfy this "need to read"; however, the new feature may prove of limited value unless the user is comfortable with script written in Bulgarian, Chinese, Japanese or Thai.

On a positive note, a beta version of the website is now operational and being tested in a growing number of languages. Much of our success is due to the crucial support of an international group of native informants who work behind the scenes on this project. The site design is now reasonably stable, while processes for the preparation and uploading of new languages have become more streamlined. The classroom version has produced impressive results and it is hoped that c-ABLE will be equally as effective – and accessible.

References

Bimber, B. (2002). "Measuring the Gender Gap on the Internet", *Social Science Quarterly*: 81:3. 868-876

Cook, V. (2008). *Second language learning and language teaching* (4th ed) London:Hodder

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Original work published 1885)

Hargittai, E. (2002). "Second Level Digital Divide: Differences in People's Online Skills", *First Monday*. http://www.firstmonday.org/ issues/issue7_4/hargittai/

Johnstone, A. H.; Percival, F. (1976). "Attention breaks in lectures", *Education in Chemistry*: 13. 49-50

Maybin, D. (2002). "Practical Considerations in the Design of *Generic* Language Learning Software", *Memoirs of Shonan Institute of Technology*. 36:1. 87-97

Maybin, D.; Batten, P. (2008). "Action-based language empowerment in Thai", *JALT 2007 Conference Proceedings.* http://jalt-publications.org/archive/proceedings/2007 /E123.pdf

Maybin, D.; Bergschneider, L. (1992). "Control: An Independent Learning Model", *Cross Currents.* 19:2. 149-159

Mehra, Bharat; Merkel, Cecelia; Bishop, Ann P. (2004), "The Internet for empowerment of minority and marginalized users", *New Media and Society* **6** (6). 781–802

Miller, G. A. (1956). "The magical number seven, plus or minus one: Some limits on our capacity for processing information", *The Psychological Review*: 63. 81-97

Saville-Troike, M. (1988). "Private speech: Evidence for second language learning strategies during the silent period", *Journal of Child Language*: 15. 567-590

Tomasello, M. (2000). "First steps toward a usage based theory of language acquisition", *Cognitive Linguistics*; 11. 61-82

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press

Mapping Architectural Appearances, Affects, and Amodality

Amir Soltani
Department of Architecture
University of Cambridge
Cambridge, UK
as2135@cam.ac.uk

***Abstract*— Many hidden dimensions of architecture are perceived by mind but usually ignored during the design process; designers not often consider for instance, sensory and cognitive modalities as resourceful possibilities and solutions to creativity. In a *Cinematic Aided Design* research framework we can tap into the potentials of these hidden dimensions in the city through various filmic mapping techniques. Recently a different method of mapping and utilizing moving image as material for soft cinematic architecture is generating multidimensional perceptions between viewer and architectural surface appearance. This paper theorizes an innovative form of multimedia synthesis between image and sound, urban public art and architecture as projection mapping using contextual and/or non-narrative filmic contents. Today many cities create large-scale projector-based filmic displays that are ornamenting facades as performative urban screens. What kind of multimedia mapping is this moving image methodology?**

Keywords-Projection Mapping, Cinematic Aided Design, Urban Screen, Architecture

I. INTRODUCTION

Film and light projection on the surface of buildings reveal the materiality of architecture in a dynamic or even as Walter Benjamin noted “explosive” renditions; hence, film can be a medium for writing and translation of concealed dynamics in the built form and understanding the myths, sensations, physicality and actions of architecture [10]. This new spatial perception requires understanding of the organization in different spatial parameters being sensed and perceived. In this theoretical paper projection mapping is paralleled with amodal translation of architectural space with real-time emotive forms. For instance, we can apply Roland Barthes’ *Writing Degree Zero* as model to conceptualize ‘cinematic architecture’ creating a transparent free-form of (reading and writing) functionless, ephemeral architectural appearances (projections), which are dynamic and real-time. In spatial analysis of projection, amodality works closely with haptic and kinesthetic modes; Brian Massumi also connects amodality to synesthetic suggesting improved sensory perception. A multimedia form of images and sounds can map these sensory/cognitive fusions and enhance our interfacing with the cities and built forms.

II. FILMIC MAPPING: SPATIAL MONTAGE

In recent architectural works a sense of “spatial thickness” in appearance of architecture is provoked through explorations of many elements like transparency, reflection, gradation, and sensation of weightlessness [8]. A building can improve in sensual ways when designers are conscious of different gradations of transparency and juxtapositions, by creating altered thicknesses of spatial refinements and variations that have the sensation of movement and light in them. This kind of positive experience of space as Pallasmaa points out brings place and meaning more accessible and is a promising way of using interpretation of design. As a way of understanding the importance of sensory architecture we can use filmic parameters in a spatial design as temporal maps. In spatial analysis we use film as a methodology in research to divide a space into basic elements then build high-level concepts based on organisation of these elements. “By establishing a logic that controls the changes and the correlation of values” in spatialisation of digital moving image we create what Lev Manovich calls “spatial montage” [3]. Digital film helps us systematise spatial data and visualise them as digital information maps. Pallasmaa complains that architectural discourses are mainly engaged in “mapping the possible marginal territories of the art that responding to human existential questions” [3]. He is blaming many unrelated augmentation of imaginaries as slowly taking the emotional content of images and making them less meaningful, as just another commodity manufactured for another boring experience; calling to attention that architecture has become an endangered art form in this age.

Cartography is the process of converting reality into what we call maps and by way of using abstractions this gives us an image of our space, place and cities. However, projection mapping of multimedia transforms data maps into reality and information into physical space; so it’s a reverse process. New media artist Pablo Valbuena expresses the interior as well as the exterior of architecture by projections that transform architectural parts, urban furniture such as concrete benches, tiles into mapping surfaces and augmented sculptures. The spatial mapping projection on building’s facade and objects becomes an extension of the facade, by ‘wrapping’. Both wrapping and unwrapping concepts can be valuable as cinematic aided design methodologies sharing amodal characteristics that involves sensory and cognitive realms of deciphering space using film as hieroglyphic medium of exploring the city and its architectural appearances. Soft architectural enmeshment is a term I’d like to associate with projections that are ‘trans-morphing’ filmic skin into facades.

Figure 1. *555 KUBIK "How it would be, if a house was dreaming,"*

The cinematic effect of this mapping is as if building is possessed, or entangled in a dream. The precision of the mapping takes the viewer to an amodal state where it affects our perception. *555 Kubik* (2009) projections on Gallery of Contemporary Art for the client Kunsthalle in Hamburg was realised by German company Urban Screen is an example of trans-morphed ('act of temporal change') on the materiality of a building in direct relation to filmic space, time, as digital mapping algorithms. The square gridded building is dynamically 'wrapped and warped' by geometric data from structural analysis of the building into "newly interpreted freely conceptual and geometric approaches."

III. Projection Mapping as Degree Zero Architecture

Film gives us privileges in cognitive and sensory investigations of architecture in the urban environment. Can exploring film as spatial interface help to sensitize our vision towards the space? We propose that film can be a degree zero in writing about spatial, temporal phenomena that connects to sensory modalities and cognitive amodal perceptions; here degree zero means that point where designers can freely cross-synthesize, explore, deconstruct and let their imaginations flow. How does architecture reinvents its appearance in film and reality? Real-time mapping buildings through multimedia projection on facades are a make-belief state of *cinemagic* (Georges Méliès) and augmented reality from the viewer's (designer) vantage point.

The existentialistic qualities of film and architecture are in the freedom of expression of one medium to another which is hampered by architectural rules and regulations. Putting all these behind us we purely explore the meaning of architecture in mind as perception. This purity gives us multimodality in multimedia where architecture's plasticity is in its inherent dimensions. Today the connection between visible and tangible needs to be addressed and explained as separate dimensions, to have antithesis, an anti-architecture solution for testing the stylistic boundaries of design.

In *Writing Degree Zero* Barthes talks about a mode of neutrality of a zero element in linguistic term, a freedom from bourgeois style. Zero as a "style of absence which is almost an ideal absence of a style" [2]; he describes degree zero as "disengaging" [2], a rebellion. Through this 'degree zero' lens a designer wants to communicate a sort of true purist and freeform architecture; at degree zero creativity one will need to established connection with existential nature of architecture when it becomes simplified into pure sensory and cognitive objectives as a method of investigating as well as creatively conceptualizing the built form. As Barthes suggests a "genuine appearance of many-sidedness" [2] and freedom should be associated with this sort of design, temporarily disengaging architecture from its rules and regulations creating an imaginative architecture.

IV. Architectural Appearance and Affects in Flux

Cities have long used the multimedia effects of lights, shadows, sound and movements to extenuate the appearance of its buildings in site-specific installations. Powerful projection of searchlight beams into sky or buildings are still common in city downtowns. Dietrich Neumann explains the early architectural external lighting as an urban phenomenon that augmented building characters at nights. They were first used at the World Trade Fairs of Paris 1889 and Chicago 1893 to illuminate various exhibition buildings [4]. In 1899 searchlights where projected at angles from both sides of the East River, New York to create "what looked like ribs of a vaulted arch," [4] and possibly it was the first artistic installation to celebrate the return of the American fleet after the success of the Spanish Civil War [4]. Not until now it has been possible to clearly change the appearances of architecture so fluidly and believingly using projections. In 1912 The Seaman's Church Institute, New York produced a Titanic Memorial Lighthouse Tower. This was one of the first times light beams were projected into the sky as memorial similar to the New York's World Trade Centre's Tribute in Light (2002). Searchlights and lighting projections are the predecessor of projection mapping.

Today digital multimedia artists and architects can augment an existing structure with mapped projections despite its size. Architectural appearances can be often invoked by different social, cultural and political origins. Temporal changes to appearances of architecture even due to weatherization, decay of time, graffiti and vandalism may affect us; yet through projection mapping city appearances can change the bodies' experiences in performative and dynamic ways. The soft architectural phenomena as projection transforms the volumetric, texture, and weight characteristics of architecture into non-permanent alterations that we can evaluate and even translate individually or collectively. The spatial qualities of every particular architectural surface can be a new map that presents new appearances. This useful non-permanent alteration is a way of tapping into the hidden dimensions of architecture even after they are built, through filmic mapping. Virilio suggests that filmic techniques become an "open system" where anything is possible and it is no longer a matter of "depth of field or perspective" rather, that it is field of limitless perceptions and structures [10]; hence, film can be a medium for de-codifying architecture understanding its myth as well as its performative actions.

In film through montage we can create worlds that physically aren't possible, by simultaneously depicting multiple images into a single space superimposed to create "ontological montage," as Manovich explains similar to Rybczynski's Tango (1982) [3]. Ontological montage is to radically juxtapose different realities within a single space/image. Likewise projection mapping uses spatial montage in its programming of surface and space modulations however, the seams and boarders between different compositing layers are mapped and calibrated through the projection device prior to performance. These techniques in manipulation of cinematic features lead us to new diverse architectures in inexpensive and intuitive ways, creating perceptually amazing results. Its non-permanence gives it illusionary quality that is an inherent part of the fleeting moving images on the screen. During a projection performance, it can be reflexively experienced and evaluated through documentation and a range of experimentations using film. As part of cinematic aided design it expands the space for new simulation of architecture, the same way that in early 90s Jeffrey Shaw used filmic and cinematic visualizations. In a multimedia installations titled Place - a user's manual (1995) a way of augmenting the real space blurring the seams between real and virtual are questioned as the "physical arrangement in the virtual space is dynamically determined by the viewer's movements." Haptic senses of contact with the scale of projections create the perception of semi-immersive spaces. New media artists are currently doing interactive mapping versions creating interfaces and constructing phantasmic and imaginative cityscapes controlled by tracking the public audience. These are usually performed at nights when we have the black canvas of darkness and the light and shadow-play becomes a synesthetic medium for haptic and other modalities to synthesize new spatial appearances.

V. Amodal Mapping: Fusion of Senses

Amodal Perception and Completion is the spatial organization of objects in mind when partially seen or hidden by obstacles; thus, amodality can be considered in a research as semantic perception of architectural experience, and as representation of architectural regions as filmic surface planes, transversely mapping surface form and its meaning. Artists projecting augmented spaces as a real-time transmorphing act of change which takes place between skin of the film and the skin of a building, revealing amodal maps that otherwise are permanently hidden. For artists and designers amodal can link virtual to the physical; as a proposed cinematic toolkit for design in digital moving image systems amodal research can bridge the perception gap between deep space of filmic and physical spaces.

As Brian Massumi suggests amodal above all is a matter of philosophy which also lies between fusions with psychology of perception, sensory, and thoughts. In mainstream films the narrative is the most powerful mode of sensory control—if sound is stopped and viewed without audio the usual modal senses create interesting amodal perceptions i.e. the off-the-frame actions and mise en scène become more pronounced. "What lies transitionally between modes is amodal," a symbolic representation of sensory fusions [5]. Site-specific works of Raphael Lozano–Hemmer create amodal fusions between the physical and virtual through connections between remote locations and users mediating his installations by Internet. He called the series "Relational Architecture." Vectorial Elevation and Amodal Suspension are interactive installations transforming the city sky at night using multiple robotics searchlight beams that respond to text messages sent by people from a website. It was performed for the opening of the Yamaguchi Centre for Arts and Media in Japan in 2003. Massumi writes Amodal Suspension "requires us to reassess our notion of the analog and the digital, of language and code, meaning and force, human and non-human communication" and Lozano–Hemmer does this by aesthetically connecting "communication to its outside" [5]. In spatialising film we objectify it. In research we can combine perceptual with sensual, amodal with modal, and relate "fusion of senses" [6] with blending of senses (e.g. synesthetic, kinesthetic, and haptic).

VI. Soft Cinematic Architecture

Projecting film on 3D surfaces verses 2D screens requires a different spatial perception. The audience point of view and the screen in 3D world is a semi-immersive experience that allows our amodal tendencies to complete an imagined perception. A sort of soft data architecture is revealed on the surface of the film and now on the architecture as haptic imagery. Plasticity of architecture is returned to sensory grandeur between the texture of spatial screens where the real and fictional boundaries are blended, reshaped and their plasticity become evident in their visual forms. Pascal Schoning mentions Le Corbusier as having envisioned the spatiality which was "constantly changing in its appearance – a cinematic experience" [9]. He relates this materiality as energy released. Perhaps this is why Le Corbusier liked film as a way of visualizing architecture since it represented more closely his idea of plasticity in form. As Schoning also points out cinematic architecture's expressive language mediates between perception and projection, in other words mental and physical. It is in our minds that architecture synthesizes its plastic spatiality and in film architecture becomes a representative of its space as energy. "Cinematic architecture confronts the stable with the temporal" [9], it is a form of physical dialogue.

Learning digital multimedia concepts which involve programming complex architectural effects as Jean Nouvel says is of interest to today's architects; effects such as transparencies are common, it involves layers of light as materials in design. Ephemeral effects are ways of

Figure 2. Amodal Perception & Completion: perceived organisation of objects in space

programming buildings “differentially over time and play with temporary effects” as traditional architecture Nouvel says, plays with the idea of permanence [1]. Today’s designers want to take advantage of temporary and subtle effects that can explore the sensory and sensual aspects of architecture by testing its physics, its continuity in time space, as well as thinking of skin-like dynamics as in transparency, light as ways of complex programming different appearances. Cinematic architecture experiments with ideas like depth of field and other haptic senses of space that are drawn through our body. The multimedia projection embodiment of space is spatialising film by connecting evolutions between time and space, matter and light, sound and image, motion and emotion that continue evolving.

VII. Soft Ornamentation

Farshid Moussavi of Foreign Office Architects (FOA) says the modernist obsession with transparency was meant to make architecture more sincere, pure and visible whereas ‘ornamentation’ and decor according to Robert Venturi was a better choice of blending cultural expression with the urban sense. Through this expression buildings communicate with wider public, and cultural expressions [7]. Projecting and animating facades create sensual and temporal soft ornamentation that focuses public arts on architecture and art fusions. It is a positive way to involve people in city spaces.

Our relationship with reality is being desensitized in architecture and other contemporary cultures such as the arts [8]. The arts have been conceptualized but there emotional content and the embodied responses through senses are regularly being left out “like gradual emptying of images of their emotional content” [8]. He is suggesting the linking of art and architecture with “culture and a mental reality” [8] as a more desired way of understanding it, therefore what Pallasmaa and others are concerned that architecture is an endangered art form by referring to disappearance of the soft architecture that our minds could envision. It is losing its sensuality and embodied essence of place that gives us sense of place, progressively in a contemporary city we are disembodied and detached from the built form. How can projection mappings using digital multimedia improve this link?

Ornaments create rhythms that connect the surfaces to spaces to bodies and to senses. In 1940s Sir John Summerson’s essay Heavenly Mansions described ornaments as “surface modulations” similar to the surface of film. Today the modernist fear of ornamentation is slowly fading as sensory debates, and topological connections with cultures and societies make designers reconsider and resample a new digital ornamentation in the form of multimedia. Perhaps the most famous ornament related argument is Adolf Loos *Ornament and Crime* (1908) essay. He set the path for future modernist vision and mentality towards ornamentation as a degenerative, senseless, and superfluous time consuming feature. This began the conceptual and experimental approaches toward new architectural design standards that became the norm. One of the technologies that represent ornamentation is film, where it can represent a non-static surface of ornamental modulation as a performative surface. The contemporary definition or position of ornaments on buildings is of its topological capacity and the surface of built form has become a critical domain to investigate its surface performance and the role of new digital multimedia in architecture.

VIII. Conclusion

An architectural sensory translation of projection into expressive multimedia mapping has future dimensions in film, architecture, and the city. Ornamentation was disregarded and not considered appropriate traits of modern architecture due to its references to history and various other visual and sensory limitations; however, there is a comeback of ornamental sensibility for architectural exterior in the form of multimedia projection mapping that is not permanent yet greatly diverse and flexible. This kind of ornamental enmeshment created by projection is indulgence for the eye and equally for the mind, freeing the obligations and spatial restrictions with the architecture and it makes us neutral towards all possible risks. It is a sort of degree zero arrival at any type of architecture this helps us create radical and free-form architectural gestures that are expressive and flexible. One of its improvisational aspects is transparency which traverses through all modalities such as synesthetic and kinesthetic to amodal cognitive perceptions of objects in space and their trans-morphing potentials. An interface to the appearances of architecture is film as maps or topographic surfaces and as interface for perception of “architectonic form” [10].

We hypothesized with amodal perception as mapping ‘degree zero’ architectural affects through digital projection of emotive visuals. This form of cinematic aided design should be recognized as a research tool in environmental design. Amodal is both philosophic and psychological, both cognitive and sensory; it’s a way to deal with the sensory possibilities of architecture through spatialising vision and other modalities. Cities are filled with architecture and spaces that inhabit them, and people animate their cities in real-time spaces. Projection mapping creates soft cinematic architecture that is a new multimedia wave in visualizing the skin of the city and architecture.

References

[1] J. Baudrillard, J. Nouvel, “The Singular Objects of Architecture,” USA, Univ. of Minnesota Press, 2002.

[2] R. Barthes, “Writing Degree Zero,” New York, Hill and Wang, 1967.

[3] L. Manovich, “The Language of New Media,” USA, MIT Press, 2001.

[4] A. Marcus, D. Neumann, (Ed.), “Visualizing the City,” UK, Routledge 2007.

[5] B. Massumi, “Flash in Japan: Brian Massumi on Rafael Lozano-Hemmer’s Amodal Suspension” (Critical Essay), Art Forum, 2003.

[6] B. Massumi, “Parables for the Virtual: Movement, Affect, Sensation,” USA, Duke Univ. Press, 2002.

[7] F. Moussavi, (Ed.), “The Function of Ornament,” USA, Actar, 2006.

[8] J. Pallasmaa, “Eyes of the Skin: Architecture and the Senses,” UK, Wiley, 2005.

[9] P. Schoning, “Manifesto for a Cinematic Architecture,” London: AA Press, 2005.

[10] P. Virilio, “The Lost Dimension,” France, Semiotext(e), 1991.

Geomarketing Policies and Augmented Reality for Advertisement Delivery on Mobile Devices.

D. De Chiara, P. Di Giovanni, M. Sebillo, G.Tortora, G. Vitiello

Università degli Studi di Salerno
Salerno, Italy
{ddechiara, pdigiovanni, msebillo, tortora, gvitiello}@unisa.it

Abstract— **Modern mobile devices have become real personal computers, that increase the ability of building/extending existing applications by combining several technologies such as camera, GPS, 3D graphics and the Internet connection. Due to the permanent Internet accessing through mobile devices, advertising is a rapidly growing sector providing brands, agencies and marketers with the opportunity to connect with consumers beyond traditional and digital media and directly on their mobile phones. In particular, customized delivery of advertisements is recognized to be a promising approach to capture users' interests in certain domains. In this paper we propose to exploit the potential of the augmented reality and geo-marketing to create a complex system for targeted marketing.**

***Keywords-** Augmented reality, geo-marketing, advertising, mobile devices, geo-localization, web services*

I. Introduction

Nowadays, mobile computing is an essential part of everyone's daily routine. From the checking of business mails while on the go, to visiting social network sites while in the mall or at the airport, and carrying out various kinds of business and transactions, it plays a huge role to draw people together even when physically apart. This capability has transformed modern mobile devices into real personal computers, increasing the ability of building/extending existing applications through the combined use of several technologies such as camera, GPS, 3D graphics and the Internet connection. In particular, the availability of a permanent Internet accessing has created new business opportunities on which agencies and marketers have focused their attention.

Advertising is a rapidly growing sector where the aforementioned potentialities are defining new solutions for the geomarketing policies. In particular, customized delivery of advertisements (*ads*, for short) on mobile devices is recognized to be a promising approach to overcome limits of traditional and digital media and capture users' interests in certain domains [11,12]. However, further attention is still required in order to investigate an adequate tradeoff between the requirements elicited from advertising companies and the degree to which a user may accept to be bothered while performing some tasks.

With respect to conventional advertising, online advertising may benefit from a promising advantage, namely the measurability. Whenever an *ad* is published on the web, several parameters can be measured, such as how many times a page with the *ad* has been seen, how many users have clicked on a suggested link with respect to those that have visited the page, how long a user has been watching a commercial video on the web. Obviously, there are several forms of online advertising and for each of them a proper measure exists, which can be specified according to the geomarketing policies. Moreover, mobile phones are extremely personal devices and this makes them a precisely targeted communication channel, improving the effectiveness of advertising campaign. Then, when designing a mobile advertising campaign, it is important to remember to provide a non-intrusive consumer experience and to ensure that ads are effectively displayed on the majority of mobile phones, without compatibility problems [4].

In this paper we present an approach for presentation of personalized ads on mobile devices, based on a user model that takes into account user's interests over time. The purpose is to create a community of people, where they write up a profile indicating their commercial interests. By means of this profile, they receive information on sales promotions targeted to certain shops that have an agreement, directly on their mobile device, also by using advanced techniques of geo-localization and augmented reality.

We describe the personalization component of the advertising management system and we explain how a user model is dynamically created, saved and updated on the basis of the latest interaction history and on the delivered contents. In order to achieve this goal our studies have been especially focused on advanced modalities of geo-targeting that allow to control where ads are displayed based on parameters like geographic point coordinates. In particular, we have decided to combine the potentials of mobile devices with augmented reality features [3] that allow customers to invoke several functionalities, ranging from traditional functions such as obtaining latest news about a product, getting an indication on a map and getting information on other products of interest.

Our proposal combines benefits coming from both approaches and is based on a geo-localization technique which integrates GPS and WiFi Positioning System (WPS). The latter allows to accurately determine the location of wireless access points (APs) to estimate the physical location of a WiFi enabled device, thus overcoming limits of GPS in terms of coverage and accuracy. As a result, users are provided with services where mobility comes into, such as services that support navigation and location based services (LBS), which may further enhance their activities inside a wide community of people.

The paper is organized as follows. In Section II we give a system overview by illustrating the available visualization modalities. In Section III, the dynamic user model for ads delivery is described. Section IV specifies the system architecture, while Section V describes the client application, Some conclusions are drawn in Section VI.

II. SYSTEM OVERVIEW

The system provides an integrated platform of services to promote commercial business that have an agreement. It is a client/server application where service requesters are consumers with a smartphone.

Users logging on a website have to create a personal profile that stores their buy preferences (see Fig. 1). By using a wizard it is possible to select commercial categories of interest by assigning a score to each item. For our purpose, we considered the classification of 35 product categories adopted by the ebay© platform. As an example, the car domain has the item subcategories sportive, limousine, city car, etc.. The score is automatically and continuously updated according to some actions taken by the user. A user may change her/his profile to add, delete a category, and update the score. Once the profile is complete user may use the client application. Through the mobile device s/he receives real-time advertisements of associated stores according to her/his position and preferences. Augmented reality may contribute to the advertisement delivery.

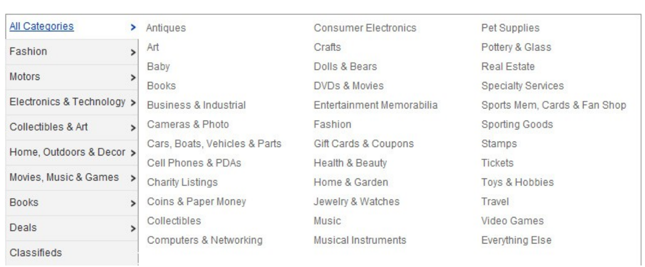


Figure 1. User profile

Whenever the user launches the application s/he can choose between two operational modes, namely Map mode and Live mode. The former corresponds to the classic two-dimensional map view, where points of interest are drawn on the map, as shown in Fig. 2, the latter consists of a real visualization enhanced by ad *hoc* textual information about item within the camera visual field, as shown in Fig. 3.

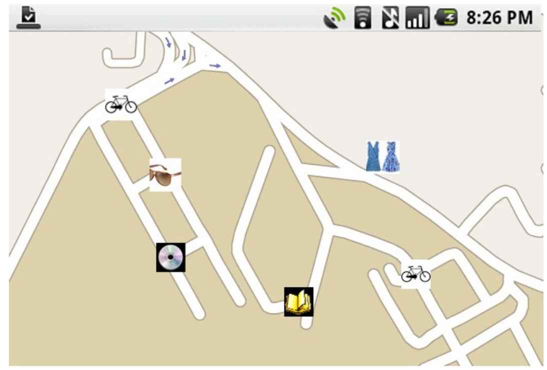


Figure 2. The Map Mode visualization modality

While running the application in Map mode, the user can see all shops having an agreement around her/his current position. In this modality s/he can zoom on a selected shop and can obtain additional information, such as discounts on some products. All the data sent to the user are first compared with preferences set in her/his profile thus giving the vendor the possibility to select only information of interest.

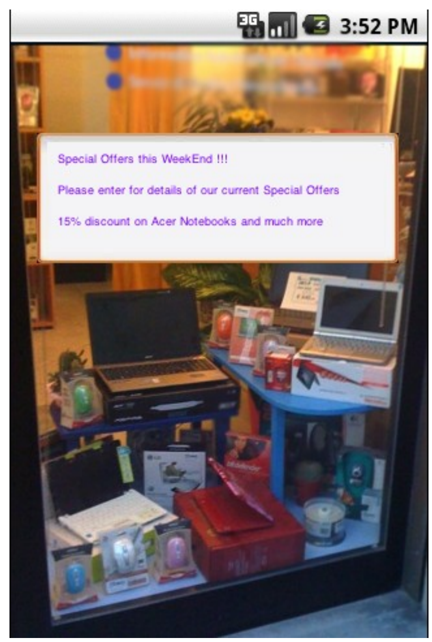


Figure 3. The Live Mode visualization modality - Outdoor

While running the application through the Live mode, the user can exploit augmented reality to improve her/his sensory perception about objects of interest around her/him, captured by the video camera, such as a product shown in the window. Fig. 4 illustrates an indoor use of the Live Mode.

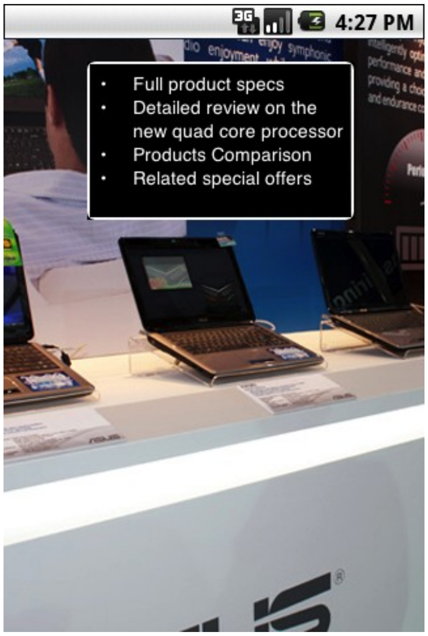


Figure 4. The Live Mode visualization modality - Indoor

In such a modality, if the user points the camera on a specific product the system provides her/him with detailed information and where available, additional information from the Internet such as reviews from specialized sites, comparison with similar products offered by the shop, additional special offers. Also an outdoor modality is available, as shown in Fig. 3. By pointing the camera towards the shopping window, the system provides the user with information about offered products selected according to the user profile. Switching between modalities depends on a threshold which combines the available number of satellites and WiFi signal strength.

III. A DYNAMIC USER MODEL FOR ADS DELIVERY

The *ads* available for the publication on a mobile device through the system prototype are chosen or discarded by associating them with a rank value which both takes into account long-term as well as short-term user interests, and periodically verifies if the dynamically updated ranks satisfy a specific pleasure threshold.

Based on results described in [4], in this Section we describe the dynamic user model and explain how long-term and short-term interests are computed and then combined to derive a user-oriented rank.

III.I LONG TERM USER INTEREST FRAMEWORKS

Long-terms user interests are modeled with respect to two reference frameworks. The former is based on a classification of the domains and the latter is based on *ad* contents mapped onto user's general interests.

As for the first framework, when building a user profile, the system stores for each domain, the score the user assigns to each of its derived subcategories. Such a score is stored as a matrix, where rows correspond to the subcategories and columns corresponds to users (see Table. 1).

CAT	**u_1**	**…**	**u_m**
C_1	val_{11}	…	val_{1m}
…	…	…	…
C_n	val_{n1}	…	val_{nm}

Table 1. User specified scores for CAT item subcategories

As each *ad* has a pre-assigned sub-category, selection with respect to this reference framework is immediate. Each *ad* a is assigned the score associated with the corresponding specific category in the corresponding user profile. Thus, the relevance of the *ad* a for a user profile u, classified as belonging to subcategory Ci, corresponds to the score assigned to Ci by the user u. Namely:

$$Rel_u(a) = S_{Ci}(u). \qquad (1)$$

A similarity value sim(a, CAT) is then computed between the *ad* a and the general category CAT to which Ci belongs, by exploiting the cosine similarity formula for the vector space model.

Consequently, the relevance between an *ad* a and all the general categories of a user model u is computed using the next formula:

$$REL_u(a) = \frac{\sum_{i=1}^{35} sim(a,CAT_i)S_{CAT}(u)}{\sum_{i=1}^{35} S_{CAT}(u)} \qquad (2)$$

The second reference framework for long-term interests further specializes user's profile. It is based on a set of user-specified keywords, which are weighted on the basis of their relevance for the user. For each user, these keywords are stored as a term weight vector (ku), as illustrated in Table 2.

k_u	**k_1**	**…**	**k_t**
U	k_{1u}		k_{tu}

Table 2. User specified scores for keywords

Again, the relevance between the *ad* and the keywords of a user model is given by the similarity cosine of the vector space model:

$$rel_{Ku}(a) = sim\ (a,\ k_u). \qquad (3)$$

Thus, the long-term interest LT(u,a) of a user u in a given *ad* a can be computed by combining formulas (1) and (3) as follows:

$$LT(u,a) = \frac{w_1 rel_u(a) + w_2 REL_u(a) + w_3 rel_{Ku}(a)}{w_1 + w_2 + w_3},$$

where w1, w2, w3 are the weights representing the importance assigned to the three relevance measures, referred to the specific category, to the general category and to user's keywords, respectively.

In order to dynamically update the long term interests of user profile, each current subcategory is georeferenced and two functions are calculated, namely $diff_s\ (lat_2, lon_2, lat_1, lon_1)$ whose output is the difference in meters between two couples of latitude and longitude coordinates corresponding to two different user's positions, and $diff_t\ (timestamp_2, timestamp_1)$ which calculates how long the user stays in the first position. In order to associate a meaningful size to the user's position, a buffer zone is performed inside which movements are not relevant.

A new variable, *ptw* (position-time weight) is defined by each customer of the system, according to his needs, with the following procedure :

Chosen the desired distance x, in meters

if $diff_s\ (lat_2, lon_2, lat_1, lon_1) \leqslant x$
then

if $diff_t(timestamp2, timestamp1) \geqslant 600$ *sec*

then

$$ptw = \frac{1}{\sum_1^n S_{Ci}}$$

else

$$ptw = 0$$

Based on the subcategory-location association, the *ptw* value is then summed to the proper subcategory, thus dynamically updating its relevance to the user.

Finally, the relevance of a subcategory for a user may be decreased of a $1/S_{Ci}$ in case either the user or the system did not update it in the last three months.

The *LT(u,a)* value is successively updated by combining it with a short-term rank which is based on the feedback the user provides during the navigation.

III.II SHORT TERM USER INTEREST FRAMEWORK

Short-term interests are represented by means of feedback terms. Such terms are obtained from the user-provided feedback over the web documents s/he receives while browsing. That is to say, the user provides positive or negative feedback (f) over the document, and a set of representative terms is extracted from them. This information is processed and the resulting value is a term weight vector (t). By fixing a *p* value, representing the number of the last web documents which should be considered for the short-term value, a similarity degree between an advertisement a and the web document d is defined as follows:

$$r_{ad} = sim(a, t_d),$$

and the current short-term interests of the user u are specified as follows:

$$r^t_{au} = \frac{\sum_{i=1}^{p} f_i r_{adi}}{p}$$

Short-terms interests tend to correspond to temporary information needs whose interest for the user vanishes after the connection session.

By combining long-term and short-term interests, the total relevance between an *ad* a and a user model u is computed by the following formula:

$$Int(u,a) = \frac{w_1 rel_u(a) + w_2 REL_u(a) + w_3 rel_{Ku}(a) + w_4 ST(u,a)}{w_1 + w_2 + w_3 + w_4}$$

where w4 is the weight representing the importance given to the short-term interest in the given user model.

IV. SYSTEM ARCHITECTURE

The system architecture is client-server based (see Fig. 5). The server receives information from clients, such as user position and time of her/his permanency on a given place.

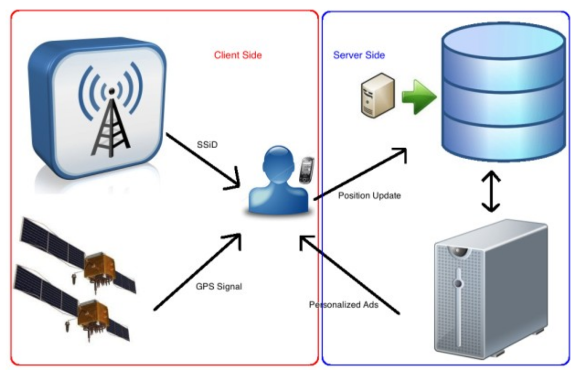


Figure 5. System Architecture

In particular, based on fixed buffer zones, which can be specified by customers in agreement with their needs, the client collects information about current user's activity and, when the position changes significantly, i.e the user has moved to a different buffer zone, a new record is sent to the server. Fig. 5 illustrates two different user positions captured by the client, and Fig. 6 shows how such data are stored on the server database.

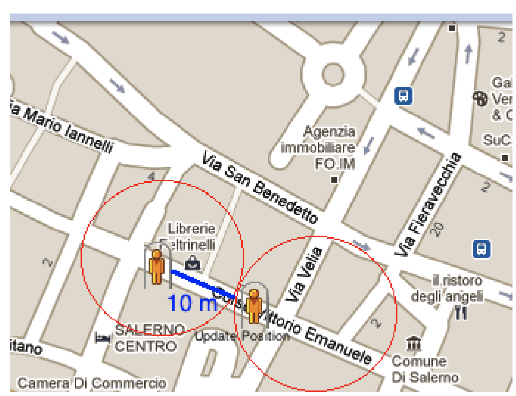


Figure 6. Tracking user position

The user's localization is obtained through two methods. The former corresponds to the classical localization through the GPS standard. The latter calculates the position through the triangulation of the WiFi hotspot signals by sniffing for the surrounding WiFi networks, then measuring signal strength and comparing those results with a threshold. The location accuracy is proportional to the number of hotspots present in the area around the device. The client collects data related to the signal strength and its SSID, and transmits them to the server, which in turn calculates the position of client according to the data contained in the database.

ID Track	ID User	Latitude	Longitude	Elevation	ID Area
1	1	36.124015°	-115.207160°	678	2
2	1	36.124187°	-115.207801°	678	
3	1	36.124190°	-115.207721°	678	4
4	1	36.124192°	-115.207689°	678	
5	1	36.124185°	-115.207671°	678	22
6	1	36.124183°	-115.207638°	678	22
7	1	36.124184°	-115.207624°	678	
8	1	36.124183°	-115.207591°	678	
9	1	36.124138°	-115.207573°	678	
10	1	36.124117°	-115.207552°	678	
11	1	36.124114°	-115.207502°	678	
12	1	36.124102°	-115.207426°	678	5
13	1	36.124070°	-115.207423°	678	5
14	1	36.124053°	-115.207424°	678	
15	1	36.124026°	-115.207425°	678	

Figure 7. Tracking user position on server database

Data stored on server database are used to make estimations to dynamically enrich user's profile and to provide more detailed hints on products of interest. User position is updated according to the information sent by the client, that allows the system to understand if the user is moving rapidly or s/he is near a shop having an agreement. Based on the position and user profile the server sends specific commercial ads.

V. THE CLIENT

The client application has been developed on a device that incorporates a software stack for mobile. In this case, Android platform [7] has been used that includes an operating system, middleware and key applications which allow autonomy in the development of application for mobile.

The Android platform is made up of:

- An hardware reference design which describes the physical characteristics that a device must satisfy to support the software stack
- A Linux based kernel (2.6 version) for processes, memory management, network stack and driver management.
- A set of Open Source libraries such as SQLite, WebKit, OpenGL, Open Core
- A run time environment used for the execution of the applications.
- A framework which provides the applications with all system services, such as telephony or geo-localization.

From a technological point of view, the system prototype combines an integrated camera for the video-image capture, a GPS device to detect the position, a compass and motion sensors to detect the user point of view. Moreover, WiFi allows to identify user position with a good accuracy inside the area not covered by GPS signal. This is possible using a technique that calculates the position through the triangulation of the WiFi hotspots signal. Obviously, this service can be offered only within WiFi covered areas, where positions of access points are known.

Android framework includes a set of core libraries that provide developers with most of the functionality available in the core libraries of the Java programming language. It provides applications with access to the location services supported by the device through the classes in the android.location package. The central component of the location framework is the LocationManager system service, which provides an API to determine location.

The access of a map is provided by MapView that displays a map with data obtained from the Google Maps service. MapView captures key-presses and touch gestures to pan and zoom the map automatically, including handling network requests for additional maps tiles. It also provides users with all of the UI elements necessary for the map control. The applications can also draw a number of Overlay types on top of the map. In general, the MapView class provides a wrapper around the Google Maps API [8].

Android also supplies 3D libraries, an implementation based on OpenGL ES 1.0 APIs. The libraries use either hardware for 3D acceleration (where available) or the included highly optimized 3D software rasterizer.

LocationManager is used for identifying the current position when the GPS signal is accessible, otherwise it exploits the WiFi hotspot triangulation technique when users get into a shopping centre.

The Map mode uses the MapView framework to view data either on a street plan or on a satellite image. The Live Mode displays data in a 3D environment by using the OpenGL ES and superimposes them on video-captured by camera.

Typically mobile devices have some computational limits that can slow down the application performances. To overcome such limits some Web Services have been exploited, thanks to the Internet permanent connection of the device. In particular, for outdoor navigation a route is provided by the GPX Driving Direction [10]. As for the indoor mobility, an ad *hoc* cartography and a service have been realized.

Both services receive two couples of coordinates representing source and target points, calculate the path between them and transmit it in GPX format (the GPS Exchange Format)[9]. GPX is a light-weight XML data format for GPS data interchange between applications and Web services on the Internet, such as waypoints, routes and tracks.

VI. CONCLUSION

The goal of the research we are carrying out is meant to realize advanced solutions to support users in their daily marketing activities, ranging from usual services to extraordinary facilities, from mobility to geo-marketing.

In this field, recent literature underlines the role that specific advanced tools play from a technological point of view, such as PDAs and mobile devices, which demonstrate to be the best common and pervasive technological solutions. This has suggested us to adopt them as the underlying platform on which the development of the proposed applications could be based.

The initial results of our research have strengthened our choice and have formed the basis for the proposal presented in this paper. Here, we have described an integrated solution for geo-targeting and location based services, which exploits an advanced visualization technique. In particular, we have combined the potentials of mobile devices with augmented reality features for simplifying the process of planning and implementation of marketing activities. We have created a social network that allows members of a group to invoke several location based functionalities. Moreover, we have associated each user with a profile that indicates her/his preferences on marketing activities. This profile is dynamically updated with information resulting from the various clients that constantly send information about customer behaviour.

In the future, we plan to improve the prototype in order to both automate its behavior in reply to some stimuli and better satisfy users' requirements on the basis of feedback obtained through a usability study meant to improve the application.

REFERENCES

[1] Paolino, L., Sebillo, M., Tortora, G., Vitiello, G.: Framy: Visualizing Geographic Data on Mobile Interfaces. Journal of Location Based Services, 2(3), 236--252 (2008)

[2] Sebillo, M., Vitiello G., De Marsico, M.: Multimodal Interfaces. In: Liu, Ling; Özsu, M. Tamer (Eds.) Encyclopedia of Database Systems, pp. 1838—1843. Springer US (2009).

[3] De Chiara, D., Paolino, L., Romano, M., Sebillo, M., Tortora, G., Vitiello, G. : LINK2U : Connecting Social Network Users through Mobile Interfaces. PCM 2010 11th Pacific Rim Conference on Multimedia September 21-24 2010 Shanghai, China

[4] Paolino, L., Sebillo, M., Tortora, G., Vitiello, G., Martellone, A.M., Tacconi, D.: Dynamic User Modeling for Personalized Advertisement Delivery on Mobile Devices.

[5] Hollerer, T., Feiner, S., Terauchi, T., Rashid, G., Hallaway, D.: Exploring mars: Developing indoor and outdoor user interfaces to a mobile augmented reality system Computers and Graphics, 23, pp. 779–-785 (1999)

[6] Norman, D.: The Psycology of Everyday Things, Basic Books, New York (1988)

[7] Android Platform : http://developer.android.com/sdk/index.html

[8] Google Maps API : http://code.google.com/intl/it-IT/apis/maps/

[9] GPX, http://www.topografix.com/gpx.asp

[10] GPXdrive direction, http://gnuite.com/cgi-bin/gpx.cgi

[11] Locately location analytics : http://www.locately.com/

[12] Watch Where You're Going : http://www.technologyreview.com/business/26643/

Automatic Generation of Multi Platform Web Map Mobile Applications

Marta Cimitile[1], Michele Risi[2], Genoveffa Tortora[2]
[1]Unitelma Sapienza University
Rome, Italy
marta.cimitile@unitelma.it
[2]University of Salerno
Fisciano(SA), Italy
{mrisi, tortora}@unisa.it

Abstract — **The development of current mobile applications is a challenging task because mobile devices are characterized by a variety of advanced services respect standard computers. In particular, these services as an example Map visualization, Web access, GPS localization, Camera, Accelerometer, and so on, interact with the applications in different ways, depending on the used mobile device platform. These challenges are further increased by the fact that each platform needs a different development process and provides a different framework to implement these mobile applications. In this paper we propose a common architecture and an unified development process implementing portable applications based on mobile services. This architecture is based on the Model-View-Control design pattern and provides a framework that generates the code starting from a formal algebraic specification. This specification integrates different formalisms such as LTL formulae and functional programming, permitting the description of structure and dynamic aspects of a mobile application. Moreover, we show how the proposed framework allows to generate Web map applications integrating different mobile services.**

Keywords- Mobile Computing; Platform-Independence; Formal Specification; Web Map Applications.

I. INTRODUCTION

Nowadays the increasing and intensive use of mobile devices can be compared to a technological explosion that obviously represents a determinant factor for the success of this segment. The fast diffusion of mobile communications according to [1] is determined by increasing technology and competition. Organizations use new technologies to offer a highest number of services and functionalities to their customers in the way to obtain competitive advantages. This goal is mostly and quickly achieved considering the wireless or mobile communications. Consequently we assist to the proliferation of software applications for mobile devices. These applications must be developed considering and exploiting the several and different characteristics of mobile devices.

According to this aspect, we can observe that differently from a computer, mobile application must be adaptable to the device operational systems and characteristics. In the last years, different systems have been proposed, wherein each one provides different characteristics and functionalities. The more diffused operational systems are: iPhone OS X, Android, Windows CE Mobile and Symbian (Nokia). Mobile devices characteristics that may influence the functionalities implementation can be:

- Screen: The screen dimensions and resolution for different devices can change. It means that developer must design a user interface adaptable to all devices screens;
- Memory: It is a critical resource, changing dependently from device capability. For this reason application developers strive to reduce the application's memory footprint by, for example, eliminating memory leaks, making resource files as small as possible, and loading resources lazily;
- Application switching: Different devices can switch applications in different way. For example in iPhone only one application can run at a time, and third-party applications never run in the background. This means that application developer must consider switching device characteristics in the way that users can leave their phone application and returning to it later, easily.
- Internet connection: The research browsers, information broadcasting, advertisement, etc. are managed in different way by mobile devices. All the applications that use internet connection must be adaptable to the different protocols, browser, and so on.
- Geographical Positioning System (GPS): Applications must be adaptable to the device Web map management and GPS system to perform geocoding and/or reverse geocoding operations.
- Webcam: Video, biometric recognizing and identification applications are influenced by device webcam features.

From the user's point of view, the problem of dependence between mobile device characteristics and application implementation is mitigated by the availability of an high number of device applications. In fact they can choose the application that permits to obtain the required service and is easier to be implemented. For the developers, there is a high difficulty to develop some applications that must be adaptable to the several devices. In particular developers are more focused to application adaptability topic than to the creative aspects.

Moreover, each platform uses a different programming language, such as Objective-C for iPhone or Java for Android, and needs a different development process and a different framework to implement mobile applications. This highlights a high degree of challenge in development of mobile native application, especially for novice programmers. A solution provided in literature [6] [7] is to develop a mobile application as a Web application. The advantages regard the portability. This approach allows exploiting the current Web design and development skills. Basing on the proposed approach developer doesn't use a specific operating system or a specific programming language and his application will run on any device having a Web browser. The main disadvantages regard the impossibility to access directly to hardware features of the mobile device and the difficulty to achieve sophisticated user-interface effects. Furthermore, programmers are in charged to manage and implement all the aspects of the application, such as the data structure, the data flow, the control flow and the objects behavior, and finally the visualization and the interaction control. During the maintenance activity the application will be changed and enhanced for several reasons [9]. It could affect different aspects of the application, so the programmers have to modify the whole implementation. Moreover, some parts of the code hardly can be reused in different applications with the same context.

In this work, we proposed an unified architecture based on a suitable design pattern which captures the essential aspects and the specific characteristics of mobile devices permitting to adapt in an easy and rapid way different mobile applications to specific device systems. Such a proposal is based on the Model-View-Controller (MVC) design pattern, which is a well-established and compelling approach to develop software. Moreover, we provide a framework that enables to integrate within an algebraic specification different formalisms such as LTL formula, permitting the description of the structure and the dynamic aspects of the mobile. This framework allows specifying a mobile application independently from a specific device or operating system and supports programmers managing and controlling all the aspects of the mobile application easily. The proposed approach is particularly suitable for implementing interactive Web map applications.

The rest of the paper is organized as follows: related works are described in Section 2, whereas Section 3 illustrates the development process and Section 4 presents the algebraic framework; a Web map case study is presented in Section 5 and finally conclusions and future works are discussed in Section 6.

II. RELATED WORKS

The proposed framework consists of an architectural model based on a design pattern and algebraic specification formalism aiming to obtain a rapid mobile applications development process.

This topic is largely explored in the last years according to the increasing diffusion of mobile devices and mobile applications [1]. In [1] a code generator to develop mobile applications is proposed. The approach allows generating the code independently from the required language. Although the code brings several benefits, its design has been limited to creating applications that can perform http connections and database interactions. In [11] a framework to face the problem of device fragmentation of mobile applications is presented. Even if the approach is of interest, it is very new and there are not applications in real context. This framework generates a client application that communicates with a remote server to run the developed application. Respect our approach, the generated application requires a net connection to run correctly.

For this reason own proposal aims to generalize the development process starting from a more high level. In fact the proposed approach supports the development of adaptable applications from the architectural level. The main architectural models adopting an interactor based view are the MVC [13].

MVC software pattern is widely used in web and mobile application [15][8]. It is a three-layer application; each layer corresponds to the following components: Model, View, Control. The Model component includes the core of application data and logic domain functionality. The View obtains data from the Model and displays them to the user. The Controller receives and interprets input into the requirements for the Model or the View. MVC decouples data access code from business logic code, and presentation code, allowing us for greater flexibility and possibility for reuse. Moreover, MVC provides a powerful way to organize systems that support multiple presentations of the same information.

The most common algebraic specification formalisms (temporal/modal logic) are widely described and compared in [2][5] on the base of their expressivity, readability and efficiency capabilities. We used Linear Temporal Logic (LTL), an algebraic model incorporating the concept of transaction. This logic is the basis for formalizing dynamics. It constitutes, together with an algebraic part (which serves as semantics for structural aspects), a firm mathematical framework to give semantics to object specifications including transactions [3]. In own approach, the use of simple version of LTL allows to define events and possible user interactions occurring during the execution of the application.

Finally, in the way to work with Web applications, we need to a Web application framework that implements the MVC design pattern. This topic is largely discussed in literature. An interesting solution is the Struts framework [15]. Apache Struts is a free open-source framework for creating Java Web applications. The Struts framework is designed to help developers to create Web applications that utilize MVC architecture. In particular, Apache Struts exploits different technologies to implement these applications, such as Java Server Pages, XML configuration and Java programming language. We chose to refer to this framework because it allows implementing Web applications by separating the control flow with the main business logic of the application to be developed. Differently from the Apache Struts, our approach defines a mobile application exploiting an integrated algebraic specification.

In the literature, different approaches have been proposed to integrate Web maps into mobile applications [10], or into Web applications [14] [17].

In [10] a mobile digital maps system based on Google Maps is defined through system topology architecture and key realization technologies. The approach allows deploying and using a digital map on the mobile device by using an ad-hoc network structure.

A rapid software prototyping approach based on new Web tools has been presented in [14]. In particular, the approach uses AJAX to integrate Google Maps into a Web application. Differently from our approach, the application functionalities have to be completely implemented by the developer.

III. THE DEVELOPMENT PROCESS

The main process that we have defined to model and develop a mobile application is shown in Fig. 1. It is composed of three phases: Definition of the Application, Platform Configuration, and Integration and Deployment. The rounded rectangles represent process phases whilst the rectangles represent the intermediate artifacts produced at the end of each phase.

The first phase allows defining the object data and behavioral aspects of the application. In this phase the views of the application are defined on the object data. The output is a description of the static and dynamic aspects of the mobile application to be developed. This description enables to integrate algebraic specifications with different formalisms such as linear temporal logic and state diagrams, permitting the description of static and dynamic aspects of the application. The algebraic framework enables the specification of both business logic and user interface of a mobile application.

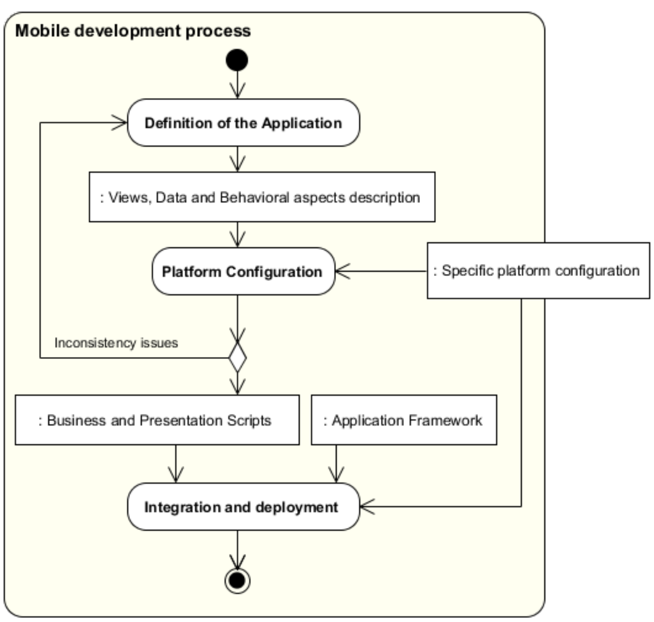


Figure 1. The mobile development process.

The second phase takes as input this specification and generates the business logic and the user interface of the application described using scripting languages. In particular, the first scripting language is used to manage the objects and the states of the mobile device whereas the latter allows managing the user interface widgets (i.e., the view). This phase takes as input a specific platform configuration file that specifies the available objects and the visualization constraints for a specific mobile platform. In particular, we consider Android, iPhone and Symbian platforms.

In case of inconsistency issues, the developer can repeat the definition phase in order to eliminate the inconsistency modifying the description of the objects and views. Finally, the latter phase integrates the generated code with the application framework. This framework specifies a generic mobile application customized with the objects, views and behavioral aspects defined through the algebraic specification. The output of the integration phase is a customized mobile application framework.

Moreover, this phase is in charge to select the definitive platform architecture. The customized application framework is linked to the specific components of the selected platform using an abstract layer that maps the application objects and widgets into the specific platform objects and widgets. The platform dependent architecture is as follow in Fig. 2.

The object data and behavioral aspects of the generated application are defined in the first layer through the business and presentation scripts and the Application Framework. These aspects use the user interface widgets, objects and APIs defined in the second layer.

Successively, the abstract layer, represented by the Platform to Objects/Widgets Bridge, maps the objects defined in the above layer into the dependent objects of the specific mobile device (i.e., Android, iPhone, Symbian). The proposed architecture enables developers to produce mobile applications which are independent from the underlying platform.

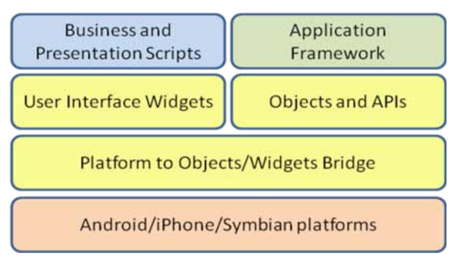


Figure 2. The platform dependent architecture.

In this paper we describe the topmost two layers of the architecture. The third and the fourth layers are implemented using the PhoneGap framework [12]. It is a standards-based, open-source development framework for building cross-platform mobile apps with HTML, CSS and JavaScript for iPhone/iPad, Google Android, Palm, Symbian, BlackBerry, Windows Mobile and more.

In the following we provide details about the phases of the development process. In particular, we propose the algebraic framework to specify the Definition of the application. Moreover, we describe the Platform Configuration phase and finally we introduce the Application Framework component that is based on the MVC design pattern.

IV. THE ALGEBRAIC FRAMEWORK

This section presents the algebraic framework. It integrates different formalisms in order to define the structure of the mobile application together its dynamic aspects. The proposed framework allows specifying in a complete and rigorous manner the mobile application. In fact, the structure of the application is modeled with state diagrams and a behavioral description, whereas events can be expressed by using the temporal logic formalism. Finally, the user interface is automatically obtained from state diagram and arranged using a constraint solver. We used a constraint solver algorithm that is in charge to displace the graphical widgets with respect to the limits of the screen device. The algorithm uses the simple list layout that visualizes the widgets in the order they are defined.

A. The Algebraic Specification

The state of a mobile application is described using the algebraic framework specification. This framework allows the developer to:

- Describe data types, objects and events that the device can manipulate through their communication functions.
- Describe the control flow and the behavioral aspects of the application.
- Describe the user interface.

In a mobile application, an object is the fundamental unit of program. It is made up of three main parts: the structure, the behavior and the events. The first is a collection of named data (attributes) identified by a unique qualified name, the second is a set of methods working on these attributes whereas the events represent the messages that occurs when a particular condition is satisfied. The methods are defined using a functional loosely typed programming approach.

A mobile program is defined using objects and actions. The former is specified when specific functionalities are needed. As an example, we consider the Map object that is in charge to manage the map visualizations. However, an action is a particular object that implements and exports the execute and view methods. The execute method describes the action behavior and how the containing objects are used and managed. The view method provides the objects' list that will be shown and arranged in the user interface.

The attributes are divided into internal (private) and external (public) type. The former type is invisible from the external and is not accessible outside the object that defines it, whereas the latter is accessible for reading and writing from other objects/actions. The public attributes are automatically saved and restored by a common storage. This storage provides a persistence of objects structure over the changing actions. In particular, these attributes are stored when an action is disposed whereas they are restored when the same actions are recreated, or when the attribute is referenced by another action. The consistency of these attributes is obtained taking into account a qualified unique name.

The mobile application is defined arranging the actions by specifying a state diagram. The state diagram allows to describe the dynamic behavior of the application. In particular, it describes all the reachable states that implement the application and how the events affect each state.

The state diagram is implemented by a state machine that executes the states and the state transitions. We can consider a state executing whereas the state machine waits until an event occurs for that state. In particular, an event occurs in the case a specific triggered condition is satisfied or when user interacts with the application. An event is a relation that links a starting with an ending state, which are not necessarily distinct. If more than one event occurs in the same time, they are serialized using a priority scheme in order to execute one event at time.

The definition of the application also needs a definition of the user interface, which is generated as shown in Fig. 3.

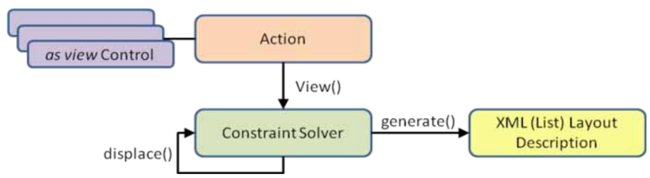


Figure 3. The User Interface.

Inside an action the proposed framework allows to define the controls which have to be displayed in the user interface. This is performed using a special keyword *as view*. Thus, these controls representing graphical widgets are exported by the method View, implicitly provided by the action. Successively, the constraint solver displaces these widgets in the graphical user interface by using a list layout visualization providing an XML description of the layout. This XML description is used to generate the final HTML user interface.

```
abstract action actionbase
{
 data section:
        //defines attributes
 behavior section:
        init();          //initializes the action
        prepare();       //prepare data before the execution
        execute();       //executes the action business logic
        dispose();       //deinitializes the action
        loadObjects(); //restores the object saved in the repository
        saveObjects(); //saves the objects in the repository
        view();          //exports the graphical controls
 event section:
        //defines events by exploiting LTL formulae
}
```

Figure 4. The *actionbase* definition.

Respect the MVC design pattern, the actions represent the Model of our application, whereas the state machine represents the Control, and the XML descriptions are used to implement the View of the model. The communication between the actions becomes through a message passing technique.

Each action is defined by inheriting and implementing a specific abstract definition, shown in Fig. 4. Generally, only the first three methods of the behavior section are overridden. The other methods are automatically implemented by the framework.

A state machine is defined by providing the definition of the actions and a set of messages that link each action with the others. In particular, the state machines are initialized by calling the *init* method of all the considered actions. When an action is invoked, the *loadObjects* method is invoked in order to restore all the public attribute values of the current action. Successively, the *prepare* method is invoked to set up the action and then the *view* method is called to show the user interface widgets organized by the Constraint Solver component. Then, the action waits for an event. When an event occurs the method *execute* which implements the business logic of the action is invoked, and finally the *saveObjects* method saves the public attribute values in a common repository. During the execution of the business logic, a message is set for the current action, thus the state machine can perform the properly transition to another state.

There are two kinds of events: local event and global event. The former is an event that is defined for a specific action. It can occur only when the defined action is active and is running. The latter, can occurs independently from the active executed action.

B. Managing the events through the LTL formula

The proposed framework uses LTL formulae to define and manage events. In particular, we consider a simple temporal logic for quantitative reasoning about mobile events. This temporal logic is an event-based linear-time logic; a time frame is considered to be an infinite, discrete and linear-ordered set of events, where an event corresponds to the occurrence of an execution action. There are several temporal operators for linear-time logic that can be used for the algebraic specifications. We consider the following temporal operators:

□ *always*, it is an event that has to occur.

◊ *eventually*, it is an event that can occur.

○ *nexttime*, it is a chaining of two actions that are executed consecutively. Generally, this event disables the user interaction.

∪ *until*, it repeats the current action until an event occurs.

The combination of these operators allows defining a complex events management.

V. Case Study

This section intends to clarify possible usages of the proposed approach by means of an example, trying to cover all the major features of the proposed framework. We propose a simple mobile application that uses the Map, GPS, Geocoding, Calendar and Camera objects. The application shows a Google Map in the main screen of the mobile application and triggers the GPS change events in order to trace a red polyline on the visualized map. In case the user clicks the *Photo* button, the application allows to take a picture by using the Camera widget. The user can come back to the initial state by clicking the *Back* button or accept the picture. In this case, other information are computed, the current date is taken from the Calendar widget, and the street name (where the user is located) is computed by using a reverse Geocoding object. These information are successively stored in a database in order to make persistent the computed data. Moreover, the data are used to build a marker and an associated tooltip that the user choose to add on the map. The state diagram of the proposed example is shown in Fig.5.

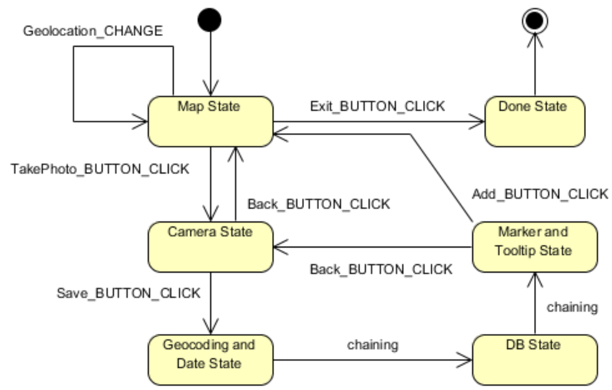


Figure 5. State diagram of the Web map application.

The example has been implemented by using the proposed algebraic framework. Fig. 6 shows the formal specification of the main action Map State.

```
action MapState extends actionbase
{
data section:
        public Map map as view;
        public Button takePhoto as view;
        public Button exit as view;
        private Geolocation gps;
behavior section:
init {
        map= Map.addMap("m_id"); //creates and adds the map to the view
        map.setCenter(gps. getCurrentPosition(); //sets the center of the map
        takePhoto.text = "Photo"; //sets the button labels
        exit.text = "Exit"
}
execute {
        map.addPolyline(gps. getCurrentPosition(), "red");
        map.setCenter(gps. getCurrentPosition());
}
event section:
        ◊gps.change → MapState
        ◊takePhoto.click → CameraState
        ◊exit.click → ○DoneState;
        ◊map.marker.click→map.marker.showTooltip();
        //no message is set, thus no change state is performed
}
```

Figure 6. The implementation of the Map State action.

The Map and the button objects are defined in the data section as widgets. Moreover, a private Geolocation object is defined in order to take the continuous GPS positions. The *init* function initializes the action by creating and adding a Google Map in a specific area of the user interface, identified by the id *m_id*. The map is centered by using the current GPS position whereas the button labels are opportunely set.

When an action is executed the function *execute* is performed. This function first adds a colored polyline on the map by exploiting the current position and then centers the map.

The event section defines the behavior when the Geolocation object indicates that the GPS position is changed. In this case, the Map State is executed again without call the *init* function. The *click* events defined on the two buttons allow to perform a transition in different states by setting the *CameraState* or the *DoneState* messages. For the latter message, the state machine performs a chaining by executing the Done State and terminating the application without user interaction. Finally, the *click* event defined on a marker shown on the map, visualizes the tooltip associated to the selected marker. In this case, no message is set thus the state machine does not perform any transition.

Once, the formal specification is completely defined, the developer chooses a specific platform (i.e., Android) to implement the Web map application. In case, the objects used in the formal specification are compatible with the selected platform, the framework generates the business and presentation scripts. These scripts are embedded into a single HTML file.

Finally, the deploy phase is performed by using the PhoneGap [12] framework and installing the generated application on an Android device. The screenshots of the generated Web map application are shown in the Fig. 7. With respect to the other methods, our approach does not need the knowledge of a specific programming language and the Google Map API specifications.

Figure 7. The Web map application screenshots.

VI. Conclusions and Future Work

Starting from the consideration of the necessity to find new solutions to the problem of dependence between mobile device characteristics and application implementation, this paper presents a framework based on an unified architecture, which captures the essential aspects and the specific characteristics of mobile devices. It permits to adapt in an easy and rapid way different mobile applications to specific device systems.

This architecture is based on the Model-View-Control design pattern to describe interactive applications and provides a framework that generates the code starting from a formal algebraic specification. This specification integrates different formalisms such as LTL formulae and functional programming.

A case study has been presented in order to highlight the main functionalities of the proposed approach. The case study shows the applicability of the framework and its capability to manage and control all the aspects of the proposed mobile application easily. In particular, we implemented a Web map application based on the use of Google API features.

Future work will aim at improve the proposed approach and apply it in several applications and contexts. Moreover we will evaluate the advantage deriving from the adoption of the proposed approach in real contests. For this reason we are available to support other research groups in applying the framework and exchanging research results.

References

[1] N Amanquah, O. T. Eporwei. “Rapid Application Development for Mobile Terminals”. In Proceedeings of the 2nd International Conference on Adaptive Science &Technologies, 2009.

[2] L. G. Bouma, N. D. Griffeth, K. Kimbler. “Feature interactions in telecommunications systems”, 32(4). Computer networks, Elsevier, 2000.

[3] G. Denker, J. Ramos, C. Caleiro, and A. Sernadas. “A Linear Temporal Logic Approach to Objects with Transactions”. In Proc. of the 6th international Conference on Algebraic Methodology and Software Technology. Lecture Notes In Computer Science, vol. 1349. Springer-Verlag, London, 1997, pp. 170-184.

[4] H. Gruber and F. Verboven. “The Diffusion of Mobile Telecommunications Services in the European Union”. European Economic Review, 2001, pp. 577-588.

[5] M. Huth and M. Ryan. Logic in Computer Science (Second Edition). Cambridge University Press. 2004.

[6] S. Jonathan, Building Android Apps with HTML, CSS, and JavaScript Jonathan Stark, O'Reilly Media, 2010.

[7] S. Jonathan. Building iPhone Apps with HTML, CSS, and JavaScript Jonathan Stark, O'Reilly Media, 2010.

[8] S. Kang, W. Manyi. “Key Technologies of the Dual-mode Electronic Commerce”. In Proc. of the Second International Symposium on Electronic Commerce and Security, 2009, pp.348-351.

[9] M. M. Lehman, “Program Evolution”, Journal of Information Processing Management, 19(1). 1984, pp. 19-36.

[10] H. Li, X. Yi. “The Study and Implementation of Mobile Digital Maps System”. In Proc. of the 2nd International Conference on Information Engineering and Computer Science (ICIECS), 2010, pp. 1-4.

[11] P. Miravet, I. Marin, F. Ortin, and A. Rionda. “DIMAG: a framework for automatic generation of mobile applications for multiple platforms”. In Proc. of the 6th International Conference on Mobile Technology, Application and Systems (Mobility '09). ACM, New York, NY, USA, pp. 1-8.

[12] PhoneGap. http://www.phonegap.com/

[13] D. Plakalović, D. Simić. “Applying MVC and PAC patterns in mobile applications”. Journal of Computing, 2(1), 2010, pp.65-72.

[14] F. Rousseaux and K. Lhoste. “Rapid Software Prototyping Using Ajax and Google Map API”. In Proc. of the 2nd International Conferences on Advances in Computer-Human Interactions (ACHI '09), IEEE Computer Society. 2009, pp. 317-323.

[15] Struts framework. http://struts.apache.org/

[16] K. Yi, W. Liwen. “Application methods of struts framework based on MVC designed pattern”. Computer Applications, Dalian, Liaoning, 2003, pp. 93-95.

[17] A. Wang, Z. Jun, W. Jiang. “Useful resources integration based on Google Maps”. In Proc. of the 4th International Conference on Computer Science & Education (ICCSE '09). 2009, pp. 1044-1047.

A Matching-Algorithm based on the Cloud and Positioning Systems to Improve Carpooling

S. Di Martino, R. Galiero, C. Giorio
University of Naples "Federico II"
80127, Naples, Italy
sergio.dimartino@unina.it

F. Ferrucci, F.Sarro
University of Salerno
84084, Fisciano (SA), Italy
{fferrucci | fsarro}@unisa.it

***Abstract* – Sustainable Mobility is a key goal to reduce pollution and improve quality of life. Information Technology can highly support Sustainable Mobility in many ways, such as highlighting to commuters the most suited means of transport for a given trip, or optimizing the vehicle's occupancy rates by dynamically arranging carpools. In this paper, we present a solution for a Cloud computing-based platform, meant to improve Sustainable Mobility. In particular, the proposed system helps users in choosing a transport solution according to its ecological footprint, matching his/her needs, preferences, and actual location. Moreover, it exploits a geosocial network to improve the users' confidence in the rides arranged with other passengers. From a technological point of view, the platform takes advantage of the Cloud infrastructure, resulting in a Software plus Services solution. In particular, the route calculation algorithm is based on a composition of different services available in the Cloud.**

I. INTRODUCTION

The reduction of the global emissions of carbon dioxide (CO_2) represents a crucial goal of our society [17]. According to the International Energy Agency, CO_2 emissions released by the fossil fuel combustion was 30 billion metric tons in 2007, with an increase of 3.2% on 2006 [8]. Among these, 23% of the CO_2 emissions refer to the transport sector that is also the main cause of oil dependency worldwide. *Sustainable Mobility* can represent a valuable approach to address the above goal. The term has been introduced to refer to any means of transport with low impact on the environment. Among the most common means for Sustainable Mobility we can find Bike- and Car-sharing, i.e. rental from a shared fleet of bikes or cars respectively, and carpooling. Carpooling (also known as car-sharing, ride-sharing, lift-sharing and *covoiturage*) is one of the most simple and effective enhancements that can be done on environmentally friendly transport solutions. It is based on the observation that if some travellers have similar departure and destination points they could share vehicles. Carpooling allows for improving the vehicles occupancy rate, that was only 1.06 persons per car in 2005 [16] with significant benefits in reducing environmental impact, decrease in traffic, reducing moving costs, promoting social relationships, and reducing the stress due to driving.

Despite the above benefits and the many efforts and initiatives to support it, presently carpooling has reached only a partial success. This is due to both technological and social issues. Many web portals are being developed that provide basic features such as the possibility to ask for a ride in a given date/time, Nevertheless, to effectively support carpooling it is necessary to define tools able to coordinate "on-the-fly" ride offerings and requests, taking into account users' positions and preferences [9]. To this aim, several technological-computational issues need to be addressed, such as:

1. Heuristics should be defined to address computational issues. Indeed the long term scheduling of carpooling has been proven to be an NP-complete problem[19]. Some heuristics have been defined for constrained versions of the problem [3], but in any case if the community of users becomes wider, there are still significant computational issues to be addressed.
2. Location awareness should be adequately employed to improve the service, also exploiting the GPS sensors in the smartphones. To this aim, integration within mobile devices should be considered as a priority to improve the diffusion of the service. Indeed, a platform working only on the internet is very limited [16], being able to serve in real-time only a fraction of potential carpoolers.

Moreover, there are also several not trivial social challenges that need to be adequately addresses. Among these, it is worth mentioning:

1. Lack of security for participants [9]. Sharing the car with unknown participants may discourage people to use these systems.
2. Participants may highly vary in terms of cultural, educational, and financial backgrounds. People with different profiles may feel uncomfortable in a ride arranged based only on their location and trip details.
3. Lack of reliability for pre-arranged rides. Often existing systems require that passengers arrange the ride and do not provide location-based features to check if a ride is on time, or if there are any problems.

In this paper, we describe a solution to effectively support *Sustainable Mobility* by means of carpooling, able to address part of the above issues. The proposal is based on the idea to coherently integrate, within a Cloud-based architecture, a set of sustainable means of transport with typical social network concepts. The use of a Cloud-based architecture is meant to address the computational issues. In particular, the route calculation algorithm is based on a composition of different services available in the Cloud. The use of several means of transport aims to increase the number of solutions that can be suggested to users while some social challenges are addressed by social network concepts. Finally, to improve the capillarity of the information distribution and the quality of the provided location-based service the services are accessible thanks to three different types of clients (desktop-based, mobile, and in-car telematics). The back-end of the described platform has been developed upon the Microsoft Azure solution, while the clients exploit the Microsoft Silverlight technology.

The paper is organized as follows. Section 2 provides an overview of existing proposals. Section 3 illustrates the ideas behind the proposed approach. Then, in Section 4, an overview of the employed technologies is provided, while in Sections 5 and 6 the route calculation algorithm and the multichannel clients are described. Some final remarks and future work conclude the paper.

II. A PLATFORM FOR CARPOOLING

In this section, the platform, we named *Lift4U*, is described illustrating its services and features, while technical details are provided in the next one.

A. *The offered features*

As described in the previous section, presently several utilities supporting users to arrange a trip are available on the web. Nevertheless, usually these applications deal with a single modal choice. The result is that, in order to find out, compare, and choose the most viable plan, user has to spend a lot of time, gathering fragmentary information and then "squeeze out" the solution. Thus, in our opinion, there is a critical lack of centralized software applications able to compare various transportation solutions and propose the one that will fit better some given user's needs and location. To this aim, the platform is proposed to support users in arranging environmentally sustainable trips that best fit their profiles, by identifying and comparing different transportation modalities. Indeed, the innovation in the proposal is due to the use of a platform that embraces many modal choices with a low environmental impact (e.g., carpooling, car sharing, bike sharing, walking, mass transit, and taxi sharing) and is able to compute the ecological and economic impact of each solution. In particular, once user set departure and destination points, Lift4U gives as output several mobility solutions among those fitting user preferences, characterized by different estimated costs, commuting time, and CO_2 emissions. Moreover, Lift4U is able to arrange carpools, where passengers are chosen taking into account also social networks, in order to mitigate social issues. Moreover, many different channels can be adopted to access the Lift4U platform, such as a web browser, a smartphone, or an in-car telematics system. Each channel has its peculiarities, allowing users to exploit a different range of services. As an example, a web browser will support user in specifying his/her preferences, while a GPS-equipped smartphone will be useful to provide a real-time location-based transportation solution. Thus, the platform interacts with a set of external data sources to get information about non-private transport solutions and many different external social networks are involved to get information about users' connections. The GIS services are handled by an external map service (in our case the combination of Microsoft Bing Maps and MapPoint). The core of the platform consists of three main modules running on the Cloud, namely a handler able to dispatch requests and to compute the different solutions (i.e., the *Sustainable Mobility Module*), a *GIS Module* which limits the interaction towards the external data sources to only those geographically involved in a given trip, and a *Carpool Module* exploited to arrange the pools of passengers. A remote database is also employed by the platform to store some information about users. Finally, a set of different intended clients is made available to improve the spreading of the platform.

B. *Exploiting GeoSocial Network Information*

To date the spreading of car sharing and pooling is limited by some social issues. Among them the lack of trust in unknown people that could share a ride is probably the biggest one. A key point of the Lift4U proposal is to take leverage on the concept of "community". Indeed, we believe that exploiting the social links established among users in a community is crucial to increase confidence in other users, by means of:

1. a friendship mechanism, where *friends,* or *friends of friends,* are preferred in the arrangement of a carpool or other mobility solutions that involve the sharing of a vehicle with other people;
2. a feedback mechanism, such as the one included in *eBay*, checking the behavior of a user in past rides. Indeed, at the end of each trip, passengers can rate each other and add also comments that will be valuable for future carpoolers. This concept holds also for other services, such as public transport, car sharing or bike-sharing, where users can leave ratings and comments on their satisfaction.

Since there are available on the web many global-coverage social networks with a huge number of users, it makes no sense to define from the scratch a new community for our platform. Rather, the Lift4U community is based on other diffused social networks to highly increase the potential catchment area of the platform and disseminate the benefits of sustainable mobility. Indeed, Lift4U can exploit social networks, such as Facebook, MSN Live, or Twitter, to access users' profile information. As a consequence, a GeoSocial Network is considered to take into account not only the user profile and preferences, but also his/her actual location, to provide more tailored location-based services, as described in the following section.

C. *The Location-Based Services*

The platform can be exploited through two main different channels: (i) a Rich Internet Application, declined for both a nomadic or a stationary users, or (ii) a web service, planned for any enterprise or public institution willing to provide sustainable mobility solutions within its own web portal. To overcome some computational issues and to scale accordingly to the number of users both these solutions are based on a Cloud infrastructure, thus we can talk of a Rich Cloud Application (or RCA) and of a Software-as-a-Service (or

SaaS). Details on these paradigms are provided in Section 4. Regarding the Rich Internet Application, to date it has been implemented in two different versions, namely a standard desktop application and a lighter mobile solution, able to exploit the positioning sensors of newer mobile phones, to provide Location-based services. The latter offers new interesting scenarios, leading to a conceptual shift from the old *travel planning* to a newer real-time dynamic arrangement of sustainable solution, optimizing users' mobility, by knowing his/her current position and preferences. As an example, if the transport solution is the carpool, the platform will automatically find "compatible" people from both spatial and user profile points of view. Once a user proposes him/herself as driver, he/she specifies the departure and destination points, the number of available seats, and a maximum allowed distance of detour to pick-up other passengers. Then, the platform creates an area of interest (or a *tunnel*) around the original route and checks the list of waiting passengers, looking for all the departure and arrivals points falling within this tunnel. This allows identifying not only the passengers making the same trip but also those who share only a part of the journey. Once all the involved people have agreed on the arranged trip, the mobile clients of the passengers will be notified of the carpool position, alerting their owners for the estimated pick-up time or for delays. These location-based services exploit the Hybrid Positioning Systems (XPS) technology that complements satellite navigation systems, such as the GPS, with Local Positioning Systems (LPS), exploiting other technologies to understand user position. In our case, a lookup mechanism on the IP of the client is used to infer its location. The combination of these technologies helps us in overcoming the limitations of common Assisted-GPS (A-GPS), limiting the "urban canyon effect" between tall buildings and working also indoors with an acceptable precision in the calculation of position. In the future we envision a third version of the client (which is under development), intended as a plug-in to be installed in Automotive Telematics Devices equipped with Microsoft Auto. In this way, a driver interested in creating a carpool could post the availability of seats on the Lift4U platform in a straightforward way, since some information are automatically inferred by the vehicle and in-car navigator status.

III. ARCHITECTURE AND INTERFACE OF LIFT4U

In this section the employed architecture of Lift4U is presented, then we illustrate how the integration with external services was realized. Finally, two main usage scenarios of the proposed system are described.

1) *The Lift4U Architecture*

The main design criteria underlying the definition of the Lift4U architecture were to get an agile application infrastructure able to:

1. easily integrate newer sustainable solutions, as soon as they become available;
2. scale up as the number of users increments;
3. be reliable and highly available.

To address these non-functional requirements and cope with the heavy computational needs of Lift4U's adaptive route matching algorithms induced us to realize the back-end on a Cloud Computing infrastructure. Cloud Computing has been defined as "[…] *a model for enabling convenient, on-demand network access to a shared pool of configurable resource (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction*" [12]. Indeed, outsourcing IT commodities to a Cloud Computing provider brings many benefits such as a higher availability of computational power offered by data centers and an excellent scalability due to the pay-per-use concept, where additional computational resources are easily allocated on demand. Also from an ecological point of view, the Cloud solution can bring many benefits, since it allows different enterprises to share computing power as well as data storage, consequently improving energy efficiency. In particular, to realize the Lift4U back-end, the Microsoft Azure technology was exploited; it is a Platform-as-a-Service (PaaS) [12] which provides a distributed operating system (mainly based on Windows 2008 Server) useful to develop on top custom Software-as-a-Service (SaaS) [12]. Other than server computation, the necessity to provide Location Based Services, obtained exploiting the Hybrid Positioning System (XPS), and in general an appealing user experience, requires that some logic runs on the client devices. To this end the *Software plus Services* (S+S) paradigm was employed [12], since it combines the appealing features of the server-side *SaaS* model with the advantages of software running on clients. Indeed, S+S is one of the most advanced architectural patterns since the back-end is implemented by a Service Oriented Architecture (SOA) running in the Cloud to offer scalable computational power, and front-end is provided by a Rich Internet Application (RIA) [1], able to fully exploit the potentialities of newer client devices, such as positioning sensors, accelerometers, and so on. More specifically, when dealing with the Microsoft technology stack, an S+S system is known as *Rich Cloud Application* (RCA), where basically the back-end is based on Microsoft Azure and the front-end on Silverlight. Concerning the clients, they are based on the Microsoft Silverlight technology, a development platform proposed by Microsoft which offers interactive user experience. Among the features offered by this technology, one of the most appealing for our purpose is the possibility to run Silverlight interface on several platforms (i.e., Web, desktop, embedded, automotive and mobile applications), with a reduced effort due to the porting between them.

2) *Integration with External Services*

The integration with the external data sources, providing information on public transport was a challenging part. The main issue is that to date, to the best of our knowledge, there is no widely accepted standard to specify and exchange public transport information, such as timetables, status, and so on. So the definition of a new interchange standard could be of great utility to improve Sustainable Mobility. Our solution to afford this problem was the adoption of the *Adapter* Design Pattern

[7], which allowed us to specify our own interface for the specification of public transport information. Then, we had to realize a wrapper -for each available web service we found - which is able to convert its actual signature into our defined interchange interface. For each provider, the system keeps track of the area the service covered. (e.g., the information that a specific bike-sharing service *BS1* covers the urban area of Rome, while *BS2* the area of Milan). This is crucial information to reduce the number of providers to involve in computing a solution. In a similar way it was also realized the integration with the social networks that offer an API and allow communications over a REST XML interface, such as Facebook, MSN Live, or Twitter. As an example, the integration with Facebook APIs allows users to post on their walls ride offers, as well as to the system to access profile information. It is worth noting that thanks to the integration with other social networks, only application-dependent user information, (e.g., preferences, feedbacks, comments, etc…) are stored in our database (i.e., Microsoft SQL Azure). Thus, leaving the most of user profile onto the main social networks' infrastructures allowed us to keep track of a very small amount of information per user. The traceability of a profile is handled by means of *OpenID* [14], a decentralized standard with over one billion enabled user accounts and over 50,000 websites accepting it. In this way a user does not need to create and maintain multiple credentials, accelerating also the SignUp process. This brings also the advantage that a user's password is not stored in our platform. Finally, the geospatial and route mapping services are orchestrated with the Representational State Transfer (REST) paradigm, described by Web Services Description Language 2.0 (WSDL). In the platform we developed, the *BingMaps* and the *MapPoint* Web Services provided by Microsoft were exploited to manage the geographical information, to get data about route directions and route mapping, and to render maps onto the web clients.

3) *Main usage scenarios*

In the following two main usage scenarios of the proposed system are described, i.e. when the solution is based on public transport and when a carpool is arranged. These two different tasks require involving different modules and 3rd party services. In Figure 1 it is reported a simplified schema showing the modules involved in a trip based only on external mobility solutions. The process is triggered by the user that specifies his/her departure and destination points through a client. This textual information is georeferred by querying an external map service provider, such as *BingMaps*. Exploiting this spatial information, the GIS module discards all the external service providers whose coverage area is too far from the requested locations. Then, each potential service is queried to retrieve plausible transport solutions. The retrieved solutions are merged and forwarded to the geospatial mapping service, now responsible to render them onto a map. Then, a ranking mechanism, based on user's profile, ecological and cost factors is applied to put in the highest positions the solutions that should best fit user preferences. Finally, the core module will transmit the set of solutions to client that requested the service.

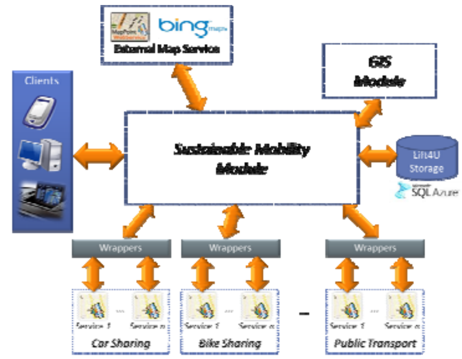


Figure 1: The modules involved in a trip based only on external mobility solutions

When dealing with carpooling, the schema slightly differs. Rather than querying external data sources, the platform searches for all the users waiting for a pick-up, whose path falls within the "tunnel", computed around the carpool driver's route. Then, a ranking mechanism promotes passengers that are in the driver's network of contacts, or those whose profile is most similar to the driver's one. Once the ride is arranged, information about status, pick-up time and location, and others are broadcasted to all the involved passenger clients.

IV. THE ROUTE CALCULATION ALGORITHM

The route calculation algorithm is one of the key challenges of the proposed solution. Indeed, its optimal solution is an NP problem, requiring to compute all the possible arrangements of rides involving all the friends in the social network. As a consequence, some heuristics are required to make the problem addressable.

The basic idea was to exploit the Cloud to search for different solutions, each of them uses a greedy approach on a different starting item. In particular, the approach explores up to 50 different solutions, differing in the people involved in the computed ride. Each solution is elaborated on a different *Worker* in the Cloud. Once all of them have computed the total distance to cover, the best solution is picked. More in details, the resulting route calculation algorithm is based on the following five steps:

1. We start by computing the route for the user who set-up the trip, by invoking the Bing Maps Web Service on his/her starting and ending destinations. The service returns the path, intended as a sequence of manoeuvres. In it is represented as the blue line.
2. We define a circle around the starting point of the route, whose radius is defined according to the user preference about the maximum allowance for a detour (the green circle in Figure 2).
3. We look for friends (and then friends of friends) within the social network connections, whose departure point fall within the above defined circle (yellow points in Figure 2). If there is a candidate, we compute the new route to match friend's destination. If the detour is bigger

than driver's preferences, the friend is discarded, and another solution is searched. If more than one candidate is found, we compute different solutions on different "workers" on the Cloud, beginning from the closer one. We limit our search to the 50 candidates whose starting points are close to the driver's one. Among all the found solutions, we look for the one which minimizes the detour.

4. If the computed detour is shorter than the maximum allowed by the driver, we consider the next potential travel-mate, moving the centre on the friend's starting point and reducing the radius of the research consistently. This is shown in Figure 3, where the green circle has been moved to the closer mate, and its radius has been reduced.
5. We iterate, until (I) the maximum number of passenger is reached, or (II) the maximum detour distance has been reached, or (III) no feasible solution is found. In the latter case, the system asks the user if he/she is interested in involving unknown people in the search.

Further heuristics are included in the algorithm, suited to prefer friends rather than friends of friends in the arrangement of the ride.

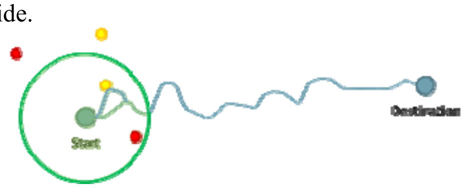


Figure 2: The Route Calculation algorithm at step 3

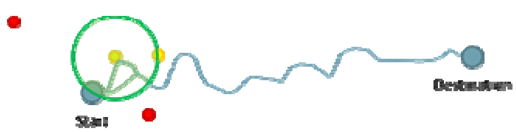


Figure 3: The Route Calculation algorithm at step 4

V. THE DEVELOPED CLIENTS

In this section the user interfaces (web and mobile) developed for the Lift4U platform are illustrated. A scenario-based description is adopted to present the interfaces, showing a real example of use. User Bob lives in Milan and needs to travel to Warsaw for a business trip in a given date. Bob accesses the login page of Lift4U website and inserts his credentials. Once authorized, he can specify his source and destination addresses. As previously described, different sustainable means of transport will be considered by the system. As a first step, the system queries the map service, to get a total distance of the trip. This information is useful to prune some commuting modalities. For instance, since the distance between Milan and Warsaw is about 1,000 miles, of course walking and bike sharing are not considered. In particular, for such a long distance, only carpooling is proposed. Once selected carpooling, Bob specifies he agrees to drive a carpool. In order to match user's convenience, the system asks the maximum distance in miles Bob is willing to spend for "pick-up" detours, the number of available seats, an estimated starting period, and the type of vehicle fueling, to compute the trip costs and CO_2 emissions. Once all this information has been gathered, Lift4U starts searching available traveling companions in the "tunnel" whose size is given by the defined maximum distance of detour. This is the most computationally expensive part of the process which is realized in the Cloud by querying the Bing Maps route calculation service. Companions are searched initially in the Bob's network of friends. If no match is found, other potential passengers are identified, ranked by similarity between profiles. At the end of this step, Bob can browse the list of potential passengers on that route (Figure 4), checking their personal details, feedbacks and comments left by previous carpoolers who have already traveled with these users.

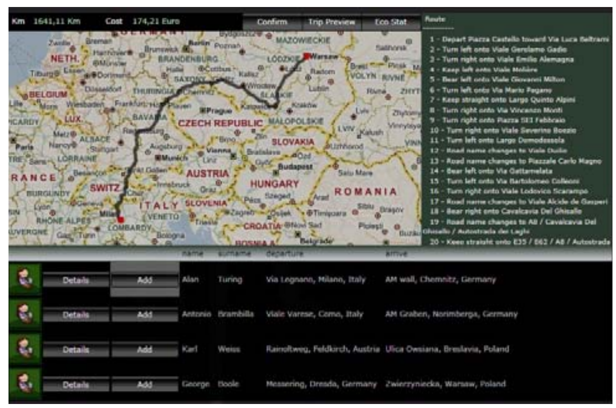


Figure 4: Selection of passengers and trip details

Once he feels confident about the reliability of a passenger, Bob can add him/her to the carpool. The system updates the trip preview on the map, with the driving directions and the travel cost per head, to reach Bob's destination. Then, the final route is previewed, its length with detours and cost effectiveness are displayed and Bob can check the travel savings and the statistics about CO_2 reduction thanks to car-pooling. Alan and other selected passengers will receive a notification via the channel specified in their preferences, namely a text message, an e-mail, or an alert when activating the Lift4U mobile client. In the latter case the platform notify Alan through his mobile device that Bob chose him as companion for the Milan-Warsaw trip. Alan is free to confirm or reject the trip plan, also considering Bob's feedbacks and comments.

VI. RELATED WORK

Many solutions concerning Sustainable Mobility issues have been proposed in the literature and on the market, but at the best of our knowledge, none of them is able to provide the set of services we are proposing. As a matter of fact, the majority of the existing works take into account only one means of transport (e.g., comparing different car rental rates) or a specific category of intended users with a repetitive transport pattern (e.g., employees, students) [9]. Among them, it is worth mentioning the online community of *NuRide* [13], where users earn rewards from sponsors as they arrange rides exploiting the system. Other interesting proposals are *eRideShare* [5] and the one by the Italian Highway agency [2]. Nevertheless, we believe that two weaknesses arise from an IT solution limited to a single means of transport: (i) user cannot compare various transportation solutions and choose the most suitable for his/her aims; (ii) user is not made aware of the existence of other Sustainable Mobility solutions, other the one suggested by the employed IT solution.

Concerning carpooling, several works tackling the problem from an algorithmic point of view mainly aiming at optimizing the selection of passengers and the route for their pick up. As an example, [18] proposed an optimization algorithm to reduce the idle time for carpoolers, while minimizing the number of vehicles on the road. Unfortunately, these works do not consider the social issues previously described. Integration between carpooling and social networks was proposed in *TravelRole* [15], but this solution does not provide support for mobile users, neither exploits location-based services.

Other interesting approaches have been presented in the domain of the Intelligent Transportation Systems, such as [20]. However, to the best of our knowledge, none of these solutions merge social network benefits with a smart route planning.

Finally, many carmakers are moving toward Sustainable Mobility solutions, such as the Daimler *car2gether* system [4], but they handle only car sharing/pooling and are often restricted to small geographical areas.

VII. CONCLUSIONS

Reducing the ecological footprint due to transportation is one of the key goals to cut the emissions of carbon dioxide. Advanced information technologies can heavily support this task, by suggesting to the users "smarter" mobility solutions among the various means of *sustainable* transportation.

In this paper, the Lif4U platform has been described. It is meant to overcome the limitations of current sustainable mobility solutions, by coherently integrating a wide range of new technologies. Among them, the Microsoft Azure platform allowed us to develop a Cloud-based back-end, where the computation of route arrangements can be done exploiting the massive elaboration power of the Cloud infrastructure. As for the services, given a user's departure and destination addresses, together with other preferences, the proposed framework computes the transport solution with the lowest ecological footprint, considering various mobility alternatives, namely carpooling, car sharing, bike sharing, walking, mass transit, and taxi sharing. As for carpooling, the use of a Geosocial network is included, in order to overcome some social issues that can arise when sharing a trip with unknown people. Indeed, the system tries to match users' profiles and preferences when combining a ride, taking into account locations and social links among users. The current location of the participants, obtained through Hybrid Positioning System, is taken into account to improve the affordability of the proposed arrangement. Many other issues have to be still addressed. First of all, an usability assessment is planned in order to evaluate the effectiveness of the proposed interfaces. Then, as any community-based solution, there are strong issues about the privacy of the information, especially when dealing with tracking user movements. Moreover, to date the use of the Cloud helped us in overcoming problems in computing the most suited arrangement for carpooling, but other heuristics should be defined, for instance to allow for mixing different means of transport within the same trip, that presently is too much computationally onerous. In particular, it could be very interesting to investigate the use of Search-based approaches, such as Genetic Algorithm or Ant Colony Optimization, for finding preference-based shortest path [6][11].

REFERENCES

[1] Allaire J., Macromedia Flash MX – A next-generation rich client; 2002.pp 1-4

[2] Autostrade Carpooling, available at http://www.autostradecarpooling.it/

[3] Coppersmith D., Nowicki T., Paleologo G., Tresser C., Wu C., *The Optimality of the On-line greedy algorithm in carpool and chairman assignment problems*, IBM Research Report, (2005)

[4] Daimler car2gether, available at http://www.car2gether.com/

[5] eRideShare, available at http://www.erideshare.com/

[6] Fu, L. Sun, D., Rilett, L.R., Heuristic shortest path algorithms for transportation applications: State of the art. Computers & OR 33(11): 3324-3343 (2006)

[7] Gamma E., Helm R., Johnson R., Vlissides J., *Design Patterns: Elements of Reusable Object-Oriented Software*, ISBN 0-201-63361-2, Addison-Wesley, 1994

[8] International Energy Agency, *CO2 Emissions from Fuel Combustion – Highlights*, 2009 Edition, available at http://www.iea.org/co2highlights/CO2highlights.pdf

[9] Lalos P., Korres A., Datsikas C., Tombras G., Peppas K., *A Framework for Dynamic Car and Taxi Pools with the Use of Positioning Systems*, computation world, pp.385-391, 2009 Computation World: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns, 2009

[10] Lift4U website, available at: http://www.lift4ucarpooling.com/

[11] S. Ok, W.J. Seo1, J.-H. Ahn, S. Kang, B. Moon, An Ant Colony Optimization Approach for the Preference-Based Shortest Path Search, in Proceeding of FGCN/ACN 2009

[12] P. Mell and T. Grance, Draft NIST working definition of cloud computing, v15, 21 Aug 2009

[13] NuRide, available at http://www.nuride.com

[14] OpenID, available at http://www.openid.net

[15] Selker, T., Saphir, P.H., *TravelRole: A carpooling / physical social network creator*, in Proceedings of the 2010 International Symposium on Collaborative Technologies and Systems (CTS), pp 629 – 634.

[16] Steger-Vonmetz, D.C., *Improving modal choice and transport efficiency with the virtual ridesharing agency*, in Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE, pp 994 - 999

[17] United Nation Millenium Goal website: http://www.un.org/millenniumgoals/ . Last visited on September 2010.

[18] Vargas, M. A., Sefair, J., Walteros, J. L., Medaglia, A. L., Rivera, L. *Car pooling optimization: a case study in Strasbourg (France)*. In Proceedings of the 2008 IEEE Systems and Information Engineering Design Symposium, pp. 89-94

[19] Varrentapp K., et al. The Long Term Car Pooling Problem – On the soundness of the problem formulation and proof of NP-completeness, November 2002

[20] Winter, S. Intelligent Self-Organizing Transport. Knstliche Intelligenz, 08 (3): 25-28

Efficient Co-Salient Video Object Detection Based on Preattentive Processing

Duan-Yu Chen and Chung-You Lin
Department of Electrical Engineering, Yuan-Ze University, Taiwan
dychen@saturn.yzu.edu.tw, *s994633@mail.yzu.edu.tw*

ABSTRACT

Automatic video annotation is a critical step for content-based video retrieval and browsing. Detecting the focus of interest such as co-occurring objects in video frames automatically can benefit the tedious manual labeling process. However, detecting the co-occurring objects that is visually salient in video sequences is a challenging task. In this paper, in order to detect co-salient video objects efficiently while maintaining the correctness, we first use the preattentive scheme to locate the co-saliency in a pair of video frames and then measure the similarity based on KL-divergence. Finally we improve the correctness of the matching across all video frames using our proposed filtering scheme. As a result, we are able to describe the significant parts of a video sequence based on the detection of co-occurring video objects. Our experiment results show that the proposed co-salient video objects modeling achieves high precision value about 85% and reveals its robustness and feasibility in video.

Keywords— Co-saliency, video object

1. INTRODUCTION

Methods for annotating video content as well as related video retrieval techniques have attracted a great deal of attention in recent years. Since video content contains much richer information than other types of media, the annotated content can be used in many fields, such as video surveillance systems and entertainment applications. Among the different types of video annotation approaches, detecting visual salience in a video is considered a good way to understand the video's content. Moreover, detecting visual saliency successfully can substantially reduce the computational complexity of the video annotation process. According to a study conducted by cognitive psychologists [17], the human visual system picks salient features from a scene. Psychologists believe this process emphasizes the salient parts of a scene and, at the same time, disregards irrelevant information. However, this raises the question: What parts of a scene should be considered "salient", especially for the scene in videos? To address this question, several visual saliency (or attention) models have been proposed in the last decade [1-5, 16, 18]. In videos, salient parts may depend on the contextual information in successive frames. Therefore, detecting the co-occurring objects across video frames would make the saliency process more specific. Therefore, in this work, co-salient video objects are the target that we aim to locate within each video clips.

In the literature [6-9, 10-11], the co-saliency problem in an image pair might be solved using some unsupervised algorithms to find the common parts among all possible correspondences. Previous approaches in the detection of co-salient objects generally would first extract the features in the regions of interest, e.g., SIFT [12], and then perform feature matching accordingly. However, the computation cost makes the process unpractical to be conducted in videos.

To speed up the process of the co-saliency computing in image pairs, Goldberger *et al.* [13] proposed two methods that approximats the KL-divergence between mixture densities. The first (match-based) method can be applied to any mixture density while the second (unscented) is tailored for mixtures of Gaussian densities. The efficiency and the performance of these methods were demonstrated on image retrieval tasks on a large database. In all the experiments conducted, the unscented approximation achieves the best results, results that are very close to large sample Monte-Carlo based ground truth. However, the KL-match based approximation is faster but less accurate than the unscented based method.

Therefore, in this paper, in order to detect co-salient video objects efficiently while maintaining the correctness, we first use the preattentive scheme [3, 15-16] to locate the co-saliency in a pair of video frames and then measure the similarity based on KL-divergence, and finally improve the correctness of the matching across all video frames using our proposed filtering scheme. As a result, we are able to describe the significant parts of a video sequence based on the detection of co-occurring video objects.

The remainder of this paper is organized as follows. In the next section, we introduce the proposed co-saliency modeling. Section 3 details the experimental results. Then, in Section 4, some concluding remarks are drawn.

2. CO-SALIENT VIDEO OBJECTS DETECTION

2.1. A Set of Preattentive Base Functions Used to Compute Salient Map

For detecting salient regions in video frames effectively, we shall have a representative set of base functions. Therefore, a set of sparse feature patches proposed by Hou and Zhang [15] is selected as the set of preattentive base functions that is of 192 8×8 image patches in RGB color space trained from a dataset of 120,000 images. Some patches are shown in Fig.1(c). The salient map for each video frame can then be obtained by applying the patch set. An example of salient map is demonstrated in Fig.1(b).

(a) (b) (c)

Fig. 1. (a) an original video frame (b) The salient map of (a) obtained by finding the incremental coding length using the set of preattentive base function [15] (c) a part of the base functions.

2.2. Frame Response Corresponding to the Base Functions

Let $\boldsymbol{B}$ denote the matrix representation of the set of base functions, where each column of $\boldsymbol{B}$ corresponds to a patch of the base functions. With the set of base functions, a given image patch $\boldsymbol{x}$ can be represented by a linear combination of the base functions, such as $\boldsymbol{x} = \boldsymbol{B}\alpha$. The coefficient set α of the linear combination can be regarded as the response corresponding to each base function. Accordingly, α can be obtained by $\alpha = \boldsymbol{B}^{-1}\boldsymbol{x}$.

To compute the response of a video frame, patches $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ are obtained by shifting a sliding window of size 8×8 in the video frame and then the normalized response p_i of each path is computed by

$$p_i = \frac{\sum_{j \in w_j > \delta} \left| F_i^T \boldsymbol{x}_j \right|}{\sum_k \sum_{j \in w_j > \delta} \left| F_k^T \boldsymbol{x}_j \right|}, \tag{1}$$

where $\Sigma \mathbf{p_i}=1$, and w_n is the saliency weighting that is set 1 initially. F_i^T is the i^{th} row vector of $\boldsymbol{B}^{-1}$. The detail of weight updating scheme is described in the following section. δ is a predefined threshold in the range [0..1] that is used to filter out patches that are of relatively small response. The absolute value of the product $F_i^T \boldsymbol{x}_j$ is adopted since the magnitude of the response is the major concern. An example of the response using the set of base functions is demonstrated in Fig.1(b). It is clear that the highlighted area would generally be the focus of interest and is thus reasonably regarded as salient regions.

2.3. Response Similarity Estimation Using KL-Divergence

For a pair of frames X_S and X_T, we then have a response probability distribution for each frame, say p and q and then compute the dissimilarity between X_S and X_T by KL-divergence defined as

$$KL(p \parallel q) = \sum_i p_i \log \frac{p_i}{q_i} \tag{2}$$

Notice that we have to compute KL(p||q) and KL(q||p) respectively since these two values may not be the same and both KL(p||q) and KL(q||p) are all equal or greater than zero. In order to indicate which patches in base functions are more important to the minimization of the KL-divergence, the partial derivative of KL-divergence for X_S and X_T is computed respectively, i.e., for example, p_i for X_S is obtained by

$$\begin{aligned} \frac{\partial}{\partial p_i} KL(p \parallel q) &= \frac{\partial}{\partial p_i} \sum_j (p_j \log p_j - p_j \log q_j) \\ &= p_i + \log p_i + p_i \log p_i - \log q_i - q_i \log q_i - KL(p \parallel q) \end{aligned} \tag{3}$$

Considering the characteristic of KL-divergence, the smaller value KL(p||q) or KL(q||p) represents the higher similarity between X_S and X_T, i.e., more co-occurring patches present in the frame pair. Therefore, those patches leading to a decrease of the KL-divergence is the features of interest and the patch selection process is based on the refined KL-divergence and is defined as:

$$KL'_{p_i}(p \parallel q) = \min\left(\frac{\partial}{\partial p_i} KL(p \parallel q),\ 0 \right). \tag{4}$$

Eq.(4) means that patches with increase of the KL-divergence would be ignored. For updating the patch weight, w_n is updated based on the KL' and is defined by

$$w_n = \frac{\sum_i KL'_{pi}(p \parallel q)(F_i^T x_j)}{\sum_i KL'_{pi}(p \parallel q)}. \tag{5}$$

The updating process is terminated until the weight value is converged.

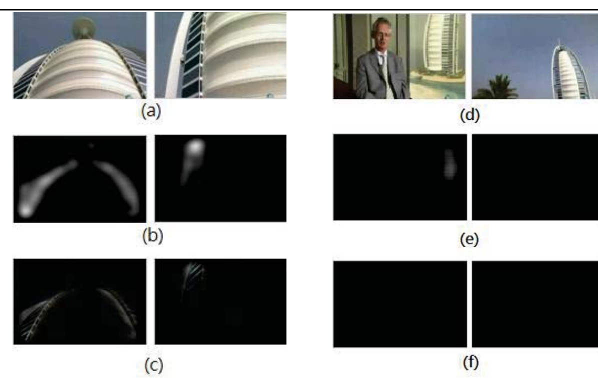


Fig. 2. Demonstration of the detection process of co-salient video objects. (a)(d) original video frame pair (b)(e) co-salient maps (c)(f) detected co-saliency objects.

An example of co-saliency in a frame pair is illustrated in Figs.2(b)(e), where co-salient regions denote that the extent is covered by some patches with small value of KL'. In Fig.2(b), co-occurring video objects can be addressed in co-salient map using the proposed approach. However, to find major co-salient objects across consecutive frames, a measure that can indicate the degree of similarity between co-salient objects is needed. The detail of this measure is described as follows.

2.4. Similarity Measure between Co-salient Objects

For co-saliency object detection, $KL'(p \| q)$ is used to select the patches that are co-salient in a frame pair. Considering a frame pair that is of certain co-saliency, $KL'(p \| q)$ would generally be similar to each other. Therefore, based on the difference of KL' for each updating and the convergence rate, the similarity between co-salient objects can be estimated.

To compute the difference between co-salient objects, we select β patches that are of the smaller $KL'(p \| q)$ and the saliency reflected by the β patches is defined by

$$S_{KL} = \sum\nolimits_{p_i \in \text{the } \beta \text{ smallest values of } KL'} \left(1 + 1/e^{|KL_{pi}|}\right). \tag{6}$$

Accordingly the distance between two co-salient objects is defined as

$$D_{KL}(p,q) = |S_{KL}(p) - S_{KL}(q)|. \tag{7}$$

To model the difference of convergence rate between co-salient objects, the difference of KL' in iteration t and $t\text{-}1$ is modeled by

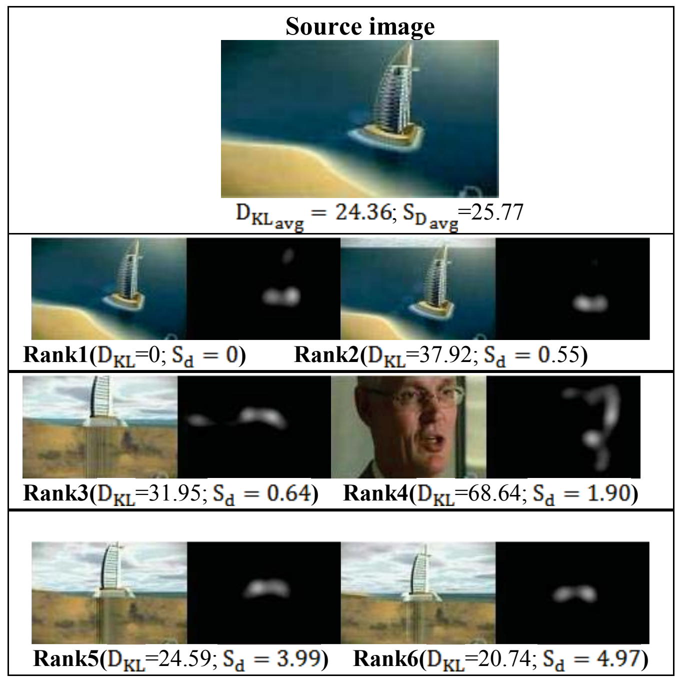
$$S_d = \sum\nolimits_{t=1}^{\tau} \sum\nolimits_{p_i} \left(KL'_{p_i(t)} - KL'_{p_i(t-1)}\right), \tag{8}$$

where τ is the number iterations. Notice that co-salient objects in a frame pair would have small value of S_d if they are similar in appearance. An example of computing similarity between co-salient objects is illustrated in Fig.2(c) and Fig.2(f). As the case of Fig.2(c), it is clear that real co-salient objects can be addressed using our measure and the distinct ones shown in Fig.2(f) can be eliminated.

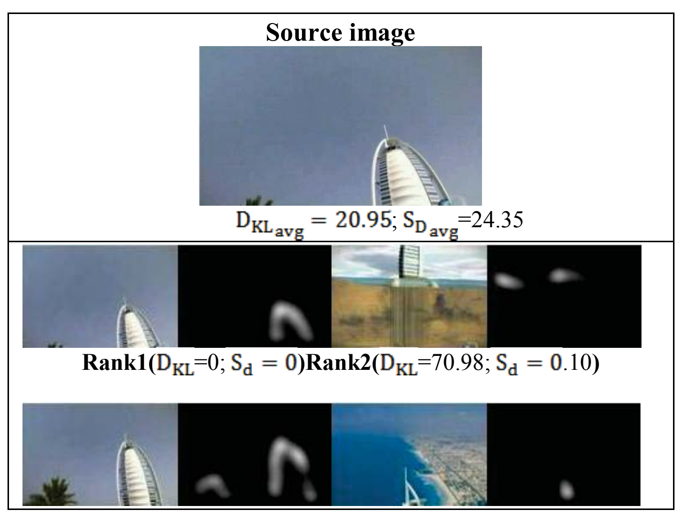
3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed visual saliency model, we conducted experiments on different kinds of videos of size 320×240 downloaded from Youtube. Some sample frames of the dataset are shown in Fig.3. In order to illustrate the qualitative evaluation of our proposed scheme, some detected co-salient objects are demonstrated. In Figs.3(b)(c), six most similar co-salient objects corresponding to source image are shown, i.e., from rank 1 to rank 6. The co-salient map for each frame pair is also shown next to its original frame to indicate where the co-salient objects are located. We can see that frames with similar co-salient objects corresponding to the source frame are detected. Particularly in Fig.3(c), co-salient objects, the roof of the hotel, of distinct size across different video frames can be correctly detected.

(a)



(b)



(c)

Fig. 3. (a) Some sample frames of the video sequence - Dubai Burj Al Arab Hotel (b)(c) top six co-salient objects are illustrated.

To show the quantitative evaluation of system performance, precision is used as the measure. The ground truth is defined manually frame by frame to see if every pair of video frames is of similar co-salient objects. Two major types of objects, buildings and people, are selected for testing since these two categories are the focus of interest for general users. In Table 1, our proposed approach achieves around 85% detection rate in average. Comparing the approach that only uses S_d for similarity measure of co-salient objects, the approach using both S_d and S_{KL} can have about 11% performance gain. For the computation cost, in average we need 0.6348 sec for each frame pair.

Table 1. The system performance of our proposed scheme. R(.) denotes ranking based on corresponding measure(s).

	Categories	$R(D_{KL})$	$R(S_d + D_{KL})$
Case 1	building	66%	83%
Case 2	people	83%	89%

4. CONCLUSION

In this paper, in order to detect co-salient video objects efficiently while maintaining the correctness, we first use the preattentive scheme to locate the co-saliency in a pair of video frames and then measure the similarity based on KL-divergence. Finally we improve the correctness of the matching across all video frames based on our proposed filtering scheme. As a result, we can describe the significant parts of a video sequence based on the detection of co-occurring video objects. Our experiment results have shown that the proposed co-salient video objects model has achieved high precision value about 85% and reveals its robustness and feasibility in video.

5. REFERENCES

[1] J. H. Elder, S. J. D. Prince, Y. Hou, M. Sizintsev, and E. Olevskiy, “Pre-attentive and attentive detection of humans in wide-field scenes,” *International Journal of Computer Vision*, 2007, vol. 72, no. 1, pp. 47–66.

[2] O. Boiman and M. Irani, “Detecting irregularities in images and in video,” *International Journal of Computer Vision*, 2007, vol. 74, no. 1, pp. 17–31.

[3] C. G. Healey, K. S. Booth and J. T. Enns, “High-speed visual estimation using preattentive processing,” *ACM Transactions on Human Computer Interaction*, 1996, vol. 3, no. 2, pp. 107–135.

[4] R. Rosenholz, “Significantly different textures: A computational model of pre-attentive texture segmentation,” Proc. *European Conference on Computer Vision* (2), 2000, pp. 197–211.

[5] M. Yang, J. Yuan and Y. Wu, “Spatial selection for attentional visual tracking,” Proc. *Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[6] H. K. Tan and C. W. Ngo, “Common pattern discovery using earth mover’s distance and local flow maximization,” Proc. *International Conference on Computer Vision*, 2005, pp. 1222–1229.

[7] A. Toshev, J. Shi and K. Daniilidis, “Image matching via saliency region correspondences,” Proc. *Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[8] J. Yuan, Z. Li, Y. Fu, Y. Wu, and T.S. Huang, “Common spatial pattern discovery by efficient candidate pruning,” Proc. *International Conference on Image Processing* (1), 2007, pp. 165–168.

[9] J. Yuan and Y. Wu, “Spatial random partition for common visual pattern discovery,” Proc. *International Conference on Computer Vision*, 2007, pp. 1–8.

[10] Z. Su, S. Takahashi, “Real-Time Enhancement of Image and Video Saliency Using Semantic Depth of Field,” Proc. *International Conference on Computer Vision Theory and Applications*, 2010, pp. 370-375.

[11] J. Li, Y. Tian, T. Huang and W. Gao, “Probabilistic Multi-Task Learning for Visual Saliency Estimation in Video,” *International Journal of Computer Vision*, 2010, pp. 150-165.

[12] D. G. Lowe, “Distinctive image features from scaleinvariant keypoints,” *International Journal of Computer Vision*, 2004, vol. 60, no. 2, pp. 91–110.

[13] J. Goldberger, S. Gordon, and H. Greenspan, “An efficient image similarity measure based on approximations of KL-divergence between two gaussian mixtures,” Proc. *International Conference on Computer Vision*, 2003, vol. 1, pp. 487–493.

[14] A. Treisman, “Preattentive processing in vision,” *Computer Vision, Graphics, and Image Processing*, 1985, vol. 31, no. 2, pp. 156–177.

[15] X. Hou and L. Zhang, “Dynamic visual attention: searching for coding length increments,” *Neural Information Processing Systems*, 2008, pp. 681-688.

[16] H. T. Chen, “Preattentive Co-Saliency Detection,” Proc. *International Conference on Image Processing*, 2010, pp. 1117-1120.

[17] W. James. The Principles of Psychology. Harvard University Press, Cambridge (1980/1981).

[18] D. Y. Chen, "Modelling Salient Visual Dynamics in Videos," *Multimedia Tools and Applications*, vol.53, no.1, pp.271-284, May 2011.

Painting into Music

An interactive multimedia edutainment tool for painting

Adam Jareh, Tom Parish, Cansin Rathge, Elina Vasina, and Kia Ng
ICSRiM – University of Leeds
School of Electronic and Electrical Engineering and School of Music
Leeds LS2 9JT, UK
mme2011@icsrim.org.uk, www.icsrim.org.uk

Abstract— Interactivity is an integral part of engaging children with what they are learning. We address this engagement by creating an interactive entertainment and learning tool for children to use in the classroom or at home using a painting scenario. We do this by coalescing music, multimedia and electronics together through the real-time mapping of painting gestures into musical output and feedback. The system analyses and objectifies aspects of the painting process, measuring painting gestures with multiple sensors and using the data to create a meaningful musical feedback that is controlled by the actions of the painter on the interactive "canvas".

The system integrates various software modules, sensor inputs and hardware systems. The paper describes several different mapping strategies, from simple one to one relationships, to complex webs of interaction between gestural data and sonic processing and controls. Input features include brush stroke analysis such as speed and position, shape and angles of drawn items. They can be configured and mapped to offer different mapping strategies, focusing on varied characteristics of the painting gesture, depending on the application context.

Keywords— multimedia; interactive; painting; mapping; sensor

I. Introduction

The bridge between artistic expression and mathematically mapped music is perhaps a difficult one to build as they are different forms of expression – visual and sonic. This project aims to explore the relationship between the two expressive domains by systematically analysing several different aspects of painting gestures and "translating" the painting movements into a musical output whilst providing an entertaining and interactive tool for children. The idea being that the user paints onto a canvas, whilst the system measures appropriate physical motion created by the action of painting (e.g. shape of the painter's strokes, speed of strokes, paint colour, etc.). Gestural data is then captured and transformed into multimedia feedback. The mapping strategy follows a set of rules to recreate similar sounds when similar painting actions are performed that can convey some insights into the strokes that are not easy to determine by eye, such as, the consistency and variations of stroke speed.

The system offers two different ways of interaction for the user. With this system, children are able to listen to their painting in addition to just looking. They can either paint as usual and listen to the music produced, or paint strategically to try and create a certain style of music.

Significant control over the system is given to the user. For example, a teacher would be able to configure appropriate mapping (or select one from a preset library) for the student in accordance to the exercise, and the user/student has the ability to choose colours and mix them in a realistic paint-based fashion.

The paper discusses related background, presents the design and development of the system and focuses on the mapping of various geometric features of paint strokes that are used to affect the musical output.

II. Related Research

Significant developments within trans-domain mapping and sonification of movement have been made in the last decade with a lot of research and industrial applications [1]. The sonification of painting gestures for technology-enhanced learning has not been fully explored. We have adopted and modified methods and existing research for our system from the following related systems.

Knoerig *et al.* [2] examined the conducting gesture, paying particular attention to the relationship between gestural expression within painting and the playing of a musical instrument. By using a physical brush on a canvas that was linked to a conductor program, they were able to show the importance of physical movement in relation to music. The focus was on a metaphorical visual representation of sound and the opportunity to learn a musical instrument without the need to comprehend the complexities of an actual instrument. This idea became the focus of the early stages of this system.

Rozin [3] proposes a method of capturing the painting process into a digital format. He used rear-projection with an easel that gradually revealed a video stream as the user painted onto the canvas, with movement of the "paint brush" being captured with the use of video cameras. Rozin's setup is adapted for our project.

Motoglyph [4] used similar methods of input, however, unlike Rozin's project, it provided the user with a musical output. Pseudo-spray cans were implemented to graffiti three glass panels, each of which had its own unique sample library. The graffiti created on the panels were measured and mapped in order to generate ringtones. Motoglyph used ultrasound as a tracking method.

The initial idea for our system was to bridge painting and music for children in an edutainment application context. [5] described a study on infants' perception of consonance and dissonance shows the importance of the musical output. It examined the "universal" perceptions of music and the reactions of infants to two different versions of a melody, one consonant and one dissonant. The article concluded that infants looked significantly longer at a source of sound that played a consonant passage, compared to a dissonant passage, where they tended to turn away and fret. The overall results suggested that infants are biologically inclined to treat consonance as perceptually more pleasing than dissonance. Whilst this project is aimed at slightly older children, who will be more susceptible to a consonant output, thus suggesting that it would be best to avoid using dissonant or unpleasant tones as the main output of our system [5]. However, the usefulness of dissonant outputs cannot be ignored, therefore, they will be implemented within the system, e.g. as a way of "rewarding" the user for consistency.

When considering the musical output, 'In B Flat' [6] gives an interesting example of how numerous sound sources can be combined to sound pleasing and avoid dissonance. 'In B Flat' uses a large number of pre-recorded videos with music played in b flat on a wide variety of instruments. The instruments can be played in any order, at any time, and sound coherent and musical. This is one of the mapping strategies that we adapt for this project, i.e. by using a set of samples in the same key which will ensure a pleasant output when played simultaneously.

III. DESIGN AND DEVELOPMENT

A. Requirements

User requirements were sought from a group of expert users consisting of 5 artists and 5 musicians. A formal survey was carried out and a summary of key points is provided. The artists believed that receiving any feedback from the process of painting would be more engaging, as it could provide a real-time feedback feature to painting that is not normally associated with the process. They were also interested in the final piece of music being composed by an artist; this was seen as a potential bridge between those art forms. The musicians were more prone to believe that a musical output could be a distraction whilst in the process of painting. However, there was interest in the detailed aspects of composing a piece of music using painting techniques and the idea of using the process of painting as a musical instrument. Overall, both groups were positive in regards to the interaction between the two art forms and would be personally interested in using or testing out the final device. We used this information and setup our program in a strategic way so that a musical piece could be easily composed using artistic gestures. This was achieved by creating a mathematical grid across the canvas. The frequency of the sample changes dependant upon the area in which the user paints with the idea that this visual representation of the notes within a scale would be interactive and logical.

The musicians were also asked about their perceptions of colour in accordance to the musical output, with a general consensus of warm shades (reds, yellows, etc.) corresponding to major, happier and more energetic sounds and instruments, and cool shades (blues, greens, etc.) to minor and mellower sounds. We used this information to setup our samples in accordance to these results. Warmer shades of colours used in the system trigger happier and more energetic sounds where as the cooler shades trigger mellow samples and ambience.

Another survey was carried out to provide some understanding of the techniques used by children when painting. 60 preschoolers were asked to paint on a piece of A3 paper whilst monitored by a teacher. Different parameters were then noted. It was found that in regards to the initial placement of the drawing (i.e. where the child starting drawing) a larger percentage of children began at the bottom of the page. Using this information, we positioned the first note of the scale at the bottom of the canvas with the following notes increasing higher up the canvas.

In terms of initial colour selection, yellow, blue and green were the most popular colours. A tally was also made for the number of times the children dipped their paint brushes into specific colour, with the results showing that yellow was the colour used most frequently and for the longest periods of time. This information was used in such a way that the more popular colours trigger the more common instruments/samples. It also triggers the more simple samples so that the piece is less likely to sound too complex.

The averaged time spent painting (continuous) was 5 minutes. Comments on techniques used were also provided by the teachers, showing that singular long lines and dabbing were used most frequently. The audio effects controller and mapping was designed to ensure that these more common artistic gestures would create common and easily noticeable effects; delay and filtering. This meant that the child may be more aware that their gestures are producing the relative effects.

B. System Overview

The end product of the design will closely replicate an artist's actual utensils and environment (see Figure 1). This is done to ensure that the end user has a familiar experience when compared with using real paint.

The design uses a wooden easel, measured to be comfortable for the average height of a 7 year old (122cm [7]) which is also usable for a seated adult. A projector is mounted onto the back of this easel, pointing at a perpendicular angle toward a sheet of smoked Perspex which is used as the screen. Using rear projection the user is presented, on screen, with the interactive "canvas" to draw on (see Figure 2). The draw tool is a modified paint brush containing sensors that enable it to interact with the canvas and software. Using a Nintendo Wiimote and an infra red LED embedded on a paint brush, the canvas is turned into a "touch screen" [8]. Colours are selected via the graphical user interface and then placed into "pots" which act as real paint pots, in that, the user dips the paintbrush in to cover it with the colour selected. The main software is developed using Max/MSP (Max hereafter) [9], which controls all the mappings, processing, analysis and feedback (visual and audio feedback), see Figure 3. The program can end at any time by selecting the 'Finish' option. Colours can be selected and if necessary mixed dependant on the number of colours chosen. If clear is selected/activated the colour choices are deactivated. If it has not been chosen the colour or colours selected will remain stored. If there is a load on the bend sensor the relevant colour image will appear on the screen and the relevant musical output is produced dependant of the point based and geometric based mapping. This is done through an audio effects patch that is manipulated relative to the output values from the mapping. These audio parameters are also effected from the shape-based, similarity-based and feature-based mapping. The visual data is sent through a projector that projects the image onto the screen. The audio data is sent to an audio output. The sensors communicate with Max using an Arduino board [10] (see Figure 4).

Figure 1. Prototype side view.

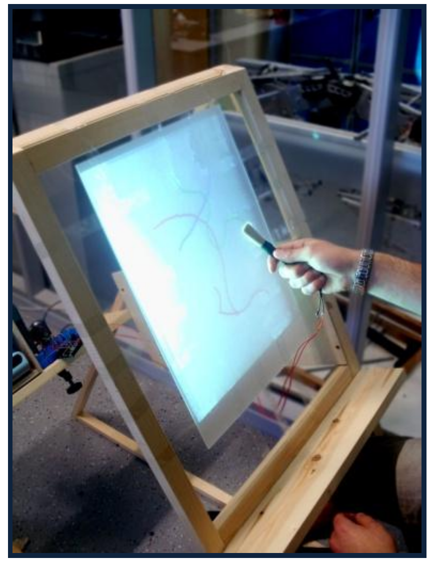


Figure 2. Interactive "canvas".

C. *Inputs*

Physical movement of the paint brush provide the key inputs to the system. A bend sensor is embedded inside the brush to measure the "pen-down" and "pen-up" action which triggers the paint and the amount of the paint, as in, controlling the thickness of the stroke depending on applied pressure.

The position of the pain brush is measure by a Wiimote tracking an infra red LED embedded on the paint brush. The Wiimote is clamped to the back of the easel and pointed toward the Perspex. After calibration, it is used to provide the coordinates of the infra red LED to a computer via Bluetooth.

When the bend sensor is triggered, the infra red LED activates. This acts as the activity trigger, which starts the position tracking and enables drawing onto the canvas. This input gives a variable value of pressure on the brush to use within mapping.

In the faux paint pots there are photocell resistors, with LEDs, that send the software a signal through Arduino when the light beam is broken. This signals that the colour picked has been added to the paint brush. The option is also present to mix the two coloured pots together to create different shades.

The position tracking of the painting gestures are derived from measuring x and y co-ordinates, when the activity trigger is on. All of these values are stored in a stream of data which we interpret with the mapping strategies and eventual output. With the 2D co-ordinates of the paint brush, its speed and acceleration are computed to objectify the speed of the paint stroke. Stroke/line length is also tracked and used as one of the key input elements to the mapping module.

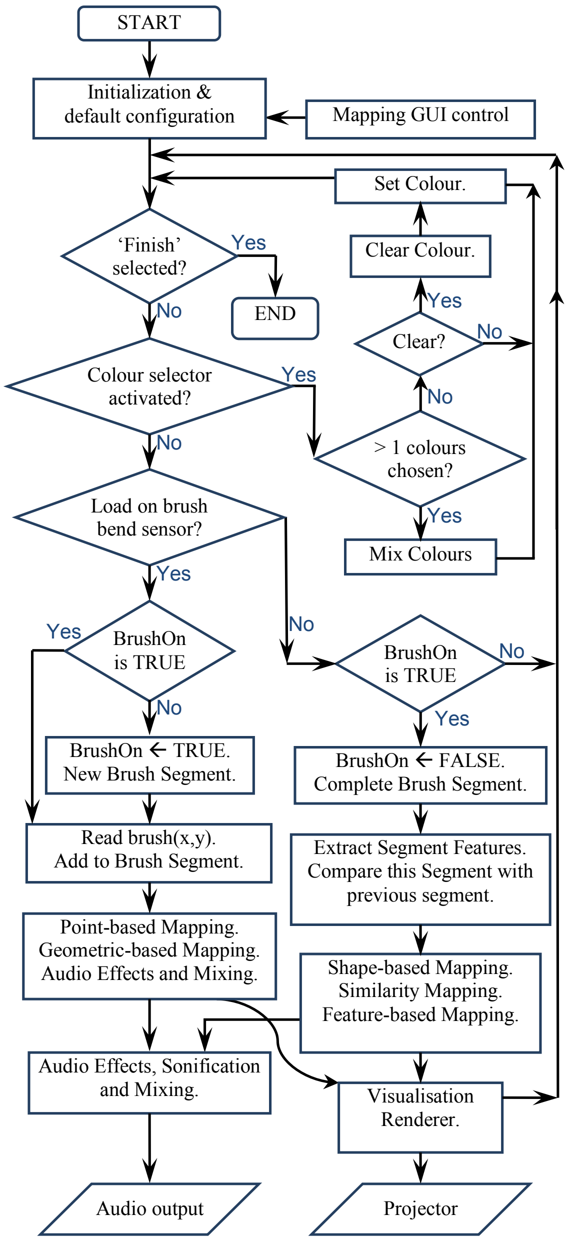


Figure 3. Software processing in Max environment.

D. Data analysis and mapping strategy

The mapping of the data is managed through a master patch in Max, which is further divided into several sub-patches. It has been designed in such a way so that the input data streams are setup independent of the output (audio) parameters that are to be manipulated. This provides more flexibility and allows us to manipulate various audio parameters by any visual ones. In doing so, it is easy to produce various system setups and mapping strategies to create a mixture of mapping possibilities. The mapping techniques that apply in the system fall under four categories.

The first category is a point-based mapping strategy, where the output is dependent on the acting point of the brush, i.e. wherever the brush is touching at any given time. This includes the coordinates and, in turn, a mathematical grid that can be determined as another input variable.

The second category is stroke-based mapping which includes variables including speed, acceleration and direction. These values are worked out within Max by using the 2D positional coordinates of the LED on the paint brush as they move over time.

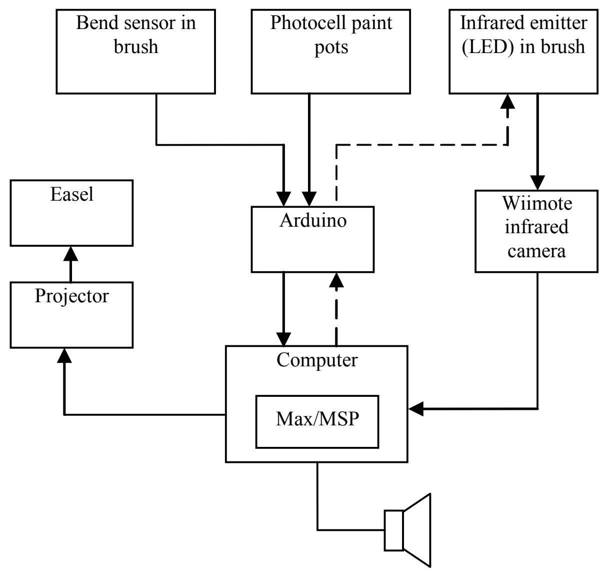


Figure 4. System set-up.

The third category is feature-based mapping. This area focuses on textural analysis. The painting techniques that the user produces are measured in a variety of ways, i.e. the number of angles and their size within a single brush stroke are quantified. In doing so, it is determined whether circular motions or sharp 'saw tooth'-like motions are being produced. The line distance is also measured to determine whether a longer or shorter line is produced, this will also pick up on any techniques such as "splodges" onto a page. Other textural techniques are noted, e.g. compact blocks of one colour being created or the size of a particular part of the painting. This mapping strategy is particularly important for pedagogical applications working on specific techniques.

The final category is class-based mapping, this focuses on analysing the shapes and sizes created on the canvas. Various regular shape forms such as squares or rectangles can be discerned from more obscure irregular shapes which can give musical feedback relevant to the types of shapes created. All of these variables can work both independently and also with other variables to produce data outputs that are then fed accordingly into a separate system that contains all of the audio parameters. This link between the two systems is easily manageable and interchangeable allowing more control over the desired output.

E. An example multimedia mapping setup

The overall system can be configured to support different mapping using different input, analysis and features depending on the application context and personalised for the individual student/user. For example, the overall length of a line (or the area of a drawn object) creates reverb, sustain and release effects. The average value of the angles within a stroke affects high and low pass filtering (with a large overall angle creating a low pass filter and a smaller angle creating a high pass filter), the idea being that a larger angle creates a larger "arc" which would let through lower and "stronger" frequency range. The number of curves in a line creates either delay or vibrato effects, i.e. if a user paints a large number of sharp angles, the visual lines are representative of the consistent frequency fluctuations within vibrato effects. The angle of the line relative to x and y axis, in relation to the length of a line, creates real time reverb and echo effects. Similarly to the high and low pass filtering effects, the visual aspects of a large angle create a larger space, therefore, the idea that the sound needs to be enlarged or elongated is appropriate. The circularity of a line alters vibrato effects and dotting on the canvas creates pizzicato effects.

IV. Prototype

A. Interactive canvas

Two main components make up the interactive canvas interface. The first component is in the form of a large white box. This is the drawing environment (the canvas) that the user draws onto to create a painting. The second component is a menu bar located underneath the drawing environment. This contains a number of options for the user to choose. There are two virtual paint palettes each containing all eight of the main colours accessible to the user. Choosing a colour will store that colour into the desired paint pot. There are also two colour indicators located on the screen above each paint pot to show the user which colour is currently stored in each pot. Also, on the menu bar is a master volume controller and a restart option. Once a colour has been assigned to each paint pot, the user can select an individual colour by dipping the paintbrush into a single pot. To mix the colours together, the user must dip the brush into both of the paint pots. The third pot is a clear function that will remove the current colour stored on the brush.

B. Output

For accessibility and to avoid musical dissonance mapping (for start up initial stage), the system applies a non-genre specific output by default, although the user can of course configure it to other behaviour as required.

The main intention of the system is not to "compose" a musical score, but to keep a child engaged for the duration of their painting and to offer meaningful feedback.

Outputs involve musical instruments that have been recorded and re-sampled to comply with requirements. The output will also contain synthesised instruments that are controlled and processed based on the mapping strategies and configuration of the system.

V. Evaluation

Qualitative evaluation of the system was conducted, internally and externally. A questionnaire was designed to capture participants' expectations for, and their experience with, the system.

A pre-usage questionnaire asked participants to rate their interest in the idea of the system, their predicted enjoyment of it and their predictions for the way they would use the system. The results showed that people were generally interested in the idea and thought they would find it enjoyable. They did not choose to attempt to predict how consonant or pleasant the output would be or whether it would accurately represent inputted painting gestures, which shows that, whilst there is understanding about what the system does, its actual output would be unclear to a first-time user until they used the system. The subjects said that they were very likely to mix colours and use a larger number of them (5-6 colours on average).

A post-usage questionnaire asked the subjects to rate their experience with the system, say how they thought they used the system and provide further comments. Questions were asked regarding how easy each tool was to use, how realistic they were and how they enhanced the experience, also, questions related to the operation of the system.

The results showed that, in terms of realism, the easel was unanimously rated as the most realistic. The paint pots, canvas and brush were rated neutrally, but, considering the additional comments, it is clear that the paint pots were a particularly distinctive feature. Whilst some noted the need to both choose a colour in the GUI and dip the brush into a paint pot, it was also noted as an interesting feature of the system. The physical tools were also, on average, rated both comfortable and easy to use, as was the overall system.

In terms of the comparison of the two surveys, participants found the system more interesting and more enjoyable than they had anticipated. They also found the output to be both more pleasant and more consonant than expected and to represent their visual inputs more accurately. It can also be noted that the time they thought they would spent using the system, was, on average, significantly shorter than the actual time they spent using it and, furthermore, their understanding of the time spent using the system and the actual timed result differed, which would suggest that the system is engaging and interesting. In terms of usage of colour, there was a clear trend for using more colours than the subjects had predicted in the pre-usage survey.

It is clear that whilst the sample group used in this questionnaire is not of a significant size, there are already some clear results emerging and this system, in its present state, has shown its potential to be interesting, engaging and realistic. It would be exceedingly beneficial to not only enlarge the sample group, but to test this system with its target user, i.e. children, and compare the results.

VI. CONCLUSION AND FUTURE DIRECTION

In this paper we present an interactive multimedia painting interface with musical feedback for technology-enhanced edutainment. It discusses the application context, requirements, a brief overview of the system design and development, and mapping strategies.

In addition to optimisation, future developments include additional sensors, such as accelerometer and pressure sensors on the pain brush to offer more detailed understanding of the painting gesture and the development of a wireless paint brush for better accessibility.

REFERENCES

[1] K. C. Ng, “Music via Motion (MvM): Trans-Domain Mapping of Motion and Sound for Interactive Performances,” in Gunnar Johannsen (ed.) Proceedings of the IEEE, Special Issue on: Engineering and Music, Supervisory Control & Auditory Communication, 92(4): pp. 645-655, April 2004.

[2] A. Knoerig, B. Mueller, and R. Wettach, “Articulated paint: musical expression for non-musicians,” *Proceedings of the 7th international conference on new interfaces for musical expression (NIME '07)*, pp. 384-385.

[3] D. Rozin, ‘Daniel Rozin Easel’, *Daniel Rozin Interactive Art* (1998) <http://www.smoothware.com/danny/neweasel.html> [15 November 2010].

[4] Motoglyph, “Motoglyph, Interactive Installation,” 2006, <http://www.thessa.eu/2006/09/01/motoglyph-interactive-installation/> [15 November 2010].

[5] M. R. Zentner and J. Kagan, “Infants’ perception of consonance and dissonance in music,” *Infant Behaviour and Development, vol.* 21.3, pp. 483-492, 1998.

[6] D. Solomon, *in Bb 2.0,* 2010, <http://www.inbflat.net/> [3 December 2010].

[7] World Health Organisation, “Height for age (5-19 years),” <http://www.who.int/growthref/who2007_height_for_age/en/index.htm> [20 March 2011].

[8] J. C. Lee, Whiteboard, <http://johnnylee.net/projects/wii/> [10 January 2011].

[9] Max/MSP, <http://www.cycling74.com/products/maxmspjitter/> [10 January 2011].

[10] Arduino, <http://www.arduino.cc/> [10 January 2011].

Multimedia Mobile Motion

Augmented Handheld Mobile Device for Motion Control

Matt Benatan, Ian Symonds, and Kia Ng
ICSRiM – University of Leeds
School of Computing, School of Electronic and Electrical Engineering and School of Music
Leeds LS2 9JT, United Kingdom
mobilemotion@icsrim.org.uk, http://mobilemotion.leeds.ac.uk

Abstract— **As the popularity and capabilities of mobile devices increases, there are many more opportunities arising for innovative applications on the mobile platform. Over recent years a trend has emerged with the majority of handsets utilizing direct user interfaces such as touch screens. However, more current mobile platform developments have also resulted in the incorporation of a number of on-board sensor systems, such as accelerometers. This project looks to augment these existing on-board sensor technologies to facilitate user-centric motion detection, and from this develop a comprehensive gesture-based motion control system for the mobile platform. This paper discusses position detection techniques for the mobile platform utilizing additional sensor augmentation, and presenting a comparative study with a computer vision-based motion tracking technology to analyze the motion data.**

Keywords- multimedia; sensor; accelerometer; ultrasonic; augmentation; interactive

I. INTRODUCTION

There are a number of existing technologies that have been augmented to facilitate motion control. Many of these, such as Sony's PlayStation Move [1] and Microsoft's Kinect [2], have been developed for commercial gaming platforms. This level of commercial investment demonstrates that there is currently a significant interest in motion-control interaction.

Motion-control interfaces, such as the game controllers mentioned above, and professional motion capture systems like Vicon [3] require static hardware units to measure the subject's movements. By definition, static reference points are not an option for mobile applications. Hence mobile devices typically use accelerometers to sense changes in movement, but not to acquire positional data.

Through this project, an add-on has been developed that utilizes ultrasonic sensors to enable a mobile device to measure user-relative positional information. This has been used as the foundation for a portable motion control interface.

There are a number of critical requirements that needed to be met in order for a successful solution to be developed. These were broken down into three primary groups:

- Portability: the resulting system would have to be portable, and not inhibit the mobile nature of the augmented device.
- Responsiveness: the system would have to function at low latency, in order to provide an optimal user experience.
- Accuracy: the system would have to accurately and consistently interpret user gestures.

II. EXISTING MOTION CONTROL TECHNOLOGIES

To date there have been few developments involving mobile device augmentation. However, there has been an interest in this area for a number of years – with systems such as Tanaka's augmented PDA [4] being developed in 2004, before mobile devices began to increase in popularity. Outside of the mobile platform, motion control technologies are widely utilized within console gaming, such as with the Nintendo Wii, PlayStation Move and Microsoft Kinect. The approaches used by each of these systems were surveyed during the development of the mobile motion control interface.

A. Nintendo Wii

The Nintendo Wii [5] is an excellent example of a multi sensory motion control system, utilizing both accelerometers and infrared to detect user gestures. The controller is able to determine its position through the use of tracking sensor technology developed by PixArt. This technology uses an infra-red camera to detect the location of two infrared beams projected by the Wii's sensor bar (serving as the static reference). The Wii console then calculates the Wii remotes position through triangulation. The infrared tracking system can then be couple with the controller's accelerometer, which provides data on the forces acting on the Wii remote, resulting in a multimodal gesture control mechanism.

B. PlayStation Move

Unlike the Wii, Sony's PlayStation Move relies on visual light, rather than infrared, for its tracking mechanism. The system tracks an orb of colored light attached to the 'wand controller', which is held by the player. As the orb is of a known size, its position can be tracked, simply by analyzing the size of the received orb image. The system is also capable of dynamically changing the orb's color to allow consistent tracking against a variety of backdrops, and in a range of different lighting conditions. The Move also incorporates a 3D accelerometer, rate sensor and magnetometer. These are used, as with the Wii, to monitor forces acting on the device and provide gesture data. The magnetometer enables the device to correct for cumulative error through calibrating its orientation against the earth's magnetic field.

C. *Microsoft Kinect*

The Kinect system uses imaging sensors, enabling the user's movement to be tracked without the need for a handheld device. The system relies on time-of-flight; transmitting infrared light and bouncing it off of the player and surrounding surfaces. The system provides a depth-map together with a color video stream for gesture analysis. User body models are used to follow and interpret the movement of the users.

D. *Other Technologies*

Other technologies for position tracking systems were also investigated, such as near magnetic field tracking [6] and ultrasonic transducers [7]. Despite the high accuracy, the near magnetic field approach proved unsuitable for mobile device augmentation due to its limited range. In contrast to this, ultrasonic transducers proved to be ideal, due to their low cost, light weight and low power requirements.

E. *Discussion*

Although many of the technologies above have proven to be highly successful in console gaming, they all suffer from one common drawback: their susceptibility to noise. Though reliable indoors, and thus suitable for console gaming applications, the data becomes less consistent in intense or constantly changing lighting environments. This makes them impractical for mobile device augmentation, as the system is required to work reliably both indoors and outside.

Accelerometer sensors on mobile devices, such as the Wii remote and Move, are both economical and widely available. This project utilizes these sensors as one of the main input sources, as their performance is not affected by lighting conditions. They also have a proven record in gesture-based interface development, having been used in a range of projects, such as Young's Hyperbow [8], the i-Maestro project [9] and the Augmented Violin [10].

III. Design and Development

The system design utilizes a multi sensor approach using the device's onboard accelerometer, and augmenting it with ultrasonic sensors to allow user-relative positional data. This data is acquired through comparing ultrasonic and radio signals. The Mobile Motion prototype is shown in Figure 1.

Figure 1. Mobile motion prototype in action.

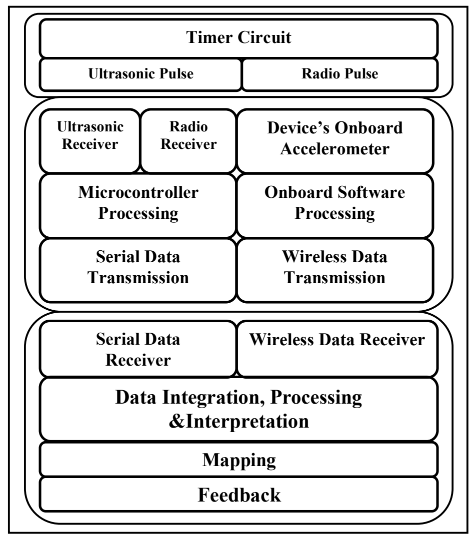


Figure 2. System architecture.

The system architecture (Figure 2) is largely modular in design with each subsystem passing information to subsequent layers resulting in output data that can be used within motion control applications. The lowest level subsystems consist of the hardware augmentation components (ultrasonic and radio units) and the device's onboard accelerometer. The data from these is then processed through hardware and software systems before it is transmitted to a laptop, where the data can be interpreted and used within prototype applications.

A. *Ultrasonic Sensor Implementation*

The theory behind the system relies on a time-of-flight principal. The mobile device is fitted with an ultrasonic transmitter and radio unit. These transmit pulses at a frequency of 50Hz to a receiver unit. This unit consists of two ultrasonic receivers and one radio receiver. The two ultrasonic receivers are worn by the user and placed 25cm apart, as illustrated in Figure 3. Upon receiving the radio and ultrasonic signals, a microcontroller determines the time delay between the received signals. This results in the time taken for the ultrasonic signal to travel to receivers one and two. These two values can then be compared to determine the position of the device.

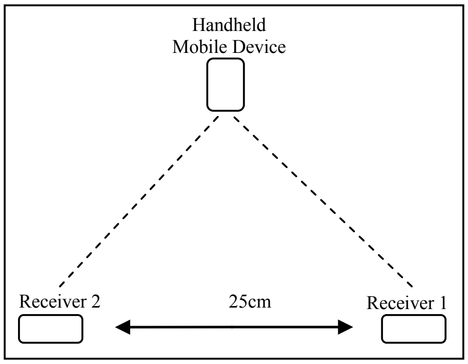


Figure 3. Ultrasonic sensor setup.

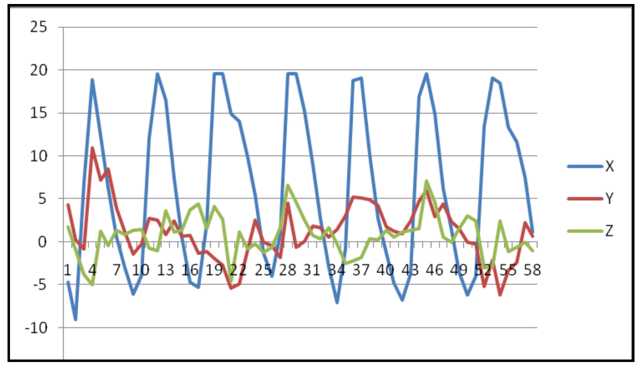


Figure 5. Graph depicting accelerometer data during conducting gesture.

B. *Mobile Device Sensors*

The device's onboard accelerometer was used to provide data from the forces acting on the phone. This provides three-dimensional information in the form of x, y and z force data from the accelerometers three axes. To use this information within the system, a custom Android application was developed (Figure 4). The application, Sensor Control, accesses the phone's sensors and transmits the data via wifi, allowing it to be used for system testing and prototyping.

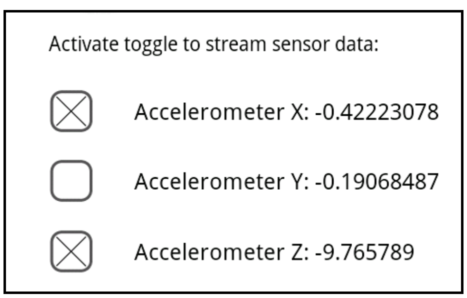


Figure 4. Sensor Control Android application.

1) *Hit Detection*

One of the first test-cases of multimedia feedback to using the motion-based interface is to use the mobile device to virtually play a percussion instrument such as a drum or a xylophone. For the system to trigger audio samples in response to the user's movement, a hit detection technique was developed. This allowed samples to be triggered dynamically in response to user gestures.

As Figure 5 demonstrates, percussive gestures have very distinct patterns. The maximum force of the gesture produces a distinct feature when compared to the resting force on the axis. Due to this, it can be measured and exploited to allow for more dynamic feedback. This is done by comparing the maximum force to a number of threshold values. Different audio samples are then triggered according to the range of values the accelerometer falls within; resulting in more realistic, dynamic feedback.

As well as velocity-sensitive playback, the hit detection process also utilizes the ultrasonic positioning information. Using this, the system is able to select the correct sample set according to the device's position – moving up in pitch towards the right, and down towards the left, just as with a real xylophone.

The multi sensory approach opens a wide range of multimedia mapping strategies; enabling simple one-to-one mapping to many-to-many mapping using different types of measurements and analysis to control different output parameters [11].

IV. Understanding Gestures

A custom application was developed to validate the accelerometer sensor as a form of capturing gesture data. This program uses Microsoft's Kinect to compare the computer vision data with the accelerometer output. Through correlating this information, it is possible to compare the performance of the accelerometer with a widely used commercial motion control system.

The application is programmed to track the user's hand position in 3D. This allows the data to be compared with the device's accelerometer output.

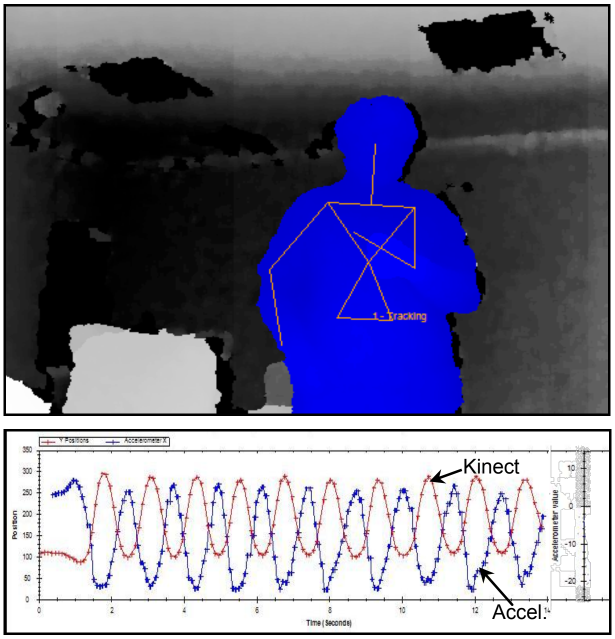


Figure 6. A syncronised view of the Kinect (vertical position coordinate only) and accelerometer (corresponding axis) output.

Figure 6 illustrates one of the dimensions as measured by both sensors, in this case using the vertical axis. It shows a logical correlation between the accelerometer and computer vision data sets. The acceleration increases with the downward movement (decreasing vertical position), and vice versa. The acceleration can also be seen to decrease as the user approaches the apogee or perigee of the gesture before the change of direction.

The correlation between the data shows that the device's onboard accelerometer is capable of providing detailed gestural information. To fully utilize this, cluster analysis algorithms have been developed to provide more detailed gesture recognition. As well as increasing system accuracy within the xylophone application, the enhanced pattern recognition algorithms have also broadened the potential applications of the system giving it the functionality required for use within a wider range of gesture-based applications.

V. Conclusion and Further Direction

This paper proposes a motion control interface for the mobile platform by augmenting an existing mobile device with additional sensors. It describes the needs of a user-relative positioning system, discusses related works and presents the design and development of the system. The paper also proposes a system that integrates ultrasonic sensors to provide time-of-flight-based positional data within a mobile setup.

The system is currently undergoing validation, and recent results prove promising. Comparing the system's performance directly with a commercial motion control unit has shown that this approach is capable of capturing complex gesture data.

Beside optimization, current works in progress include longer range ultrasonic sensing. This capability does not have significant impact on the device's main functionality; though improving the coverage area will broaden potential application domains.

Mobile Motion demonstrates that a portable solution to motion control can be developed and refined for use in a range of applications. These could include game development, alternative forms of device control, musical interfaces and gestural research applications. The system could also be further expanded to utilize other technologies, such as GPS, to enable the development of combined user-relative positioning and global positioning systems.

References

[1] A. Mikhailov, "Sony reveals what makes PlayStation Move tick," 2009, http://www.viddler.com/explore/sceablog/videos/546/

[2] J. Shotton, "Real-Time Human Pose Recognition in Parts from a Single Depth Images," IEEE Computer Vision and Pattern Recognition 2011, Colorado, USA, 21-25 June 2011.

[3] Vicon, www.vicon.com (18/04/2011).

[4] A. Tanaka, "Mobile Music Making," in Proceedings of New Interfaces for Musical Expression, Hamamatsu, Japan, 2004.

[5] Nintendo of Europe GmbH, Wii – Nintendo, wii.nintendo.co.uk (31/03/2011).

[6] M. Bezdicek and D. G. Caldwell, "Portable Absolute Position Tracking System for Human Hand Fingertips," in Proceedings of Virtual Concept 2006, Cancún, Mexico, November 27 - December 1, 2006.

[7] T. B. Greenslade, "Experiments with Ultrasonic Transducers," in The Physics Teacher, vol. 32, pp. 392-398, October 1994.

[8] D. Young, "Wireless Sensor System for Measurement of Violin Bowing Parameters," in Proceedings of Stockholm Music Acoustics Conference, KTH Speech, Music and Hearing, Stockholm, Sweden, 6-9 August 2003.

[9] D. Bradshaw and K. Ng, "Tracking Conductors Hand Movements using Multiple Wiimotes," in Proceedings of the International Conference on Automated Solutions for Cross Media Content and Multi-channel Distribution, Florence, Italy, 2008.

[10] N. Rasamimanana, F. Guedy, N. Schnell, J.-P. Lambert, and F. Bevilacqua, "Three Pedagogoical Scenarios using the Sound and Gesutre Lab," in Ng (ed.) Proceedings of the 4th i-Maestro Workshop on Technology-Enhanced Music Education, co-located with the NIME 2008 International Conference, Università degli Studi di Genova, Italy, 4 June 2008.

[11] K. Ng, "Music via Motion (MvM): Trans-Domain Mapping of Motion and Sound for Interactive Performances," in Gunnar Johannsen (ed.) Proceedings of the IEEE, Special Issue on Engineering and Music — Supervisory Control and Auditory Communication, vol. 92(4), pp. 645-655, April 2004.

Virtual Drum Accompanist

Interactive Multimedia System to Model Expression of Human Drummers

Elliot Danby and Kia Ng
ICSRiM – University of Leeds
School of Computing and School of Music
Leeds LS2 9JT, UK
vdrum@icsrim.org.uk, www.icsrim.org.uk

***Abstract*— This paper presents a system that makes use of machine listening techniques to analyse an incoming musical audio signal and determine various musical properties such as dynamics, tempo, and style. With this knowledge, the system generates an expressive drumbeat to accompany the input audio signal in real time. These drumbeats are created with a set of genetic algorithms which are designed to model basic human creativity in a musical sense. The system is intended as a practice tool as well as a means to observe the musical interaction that may occur between humans and computers.**

Keywords- multimedia; interactive; drum; accompaniment; beat tracking; genetic algorithm; creativity; machine listening

I. Introduction

A. Background and Motivation

The interaction between human and computerized musicians is an interesting concept to consider. The marimba-playing Shimon and percussion-playing Haile [7] provide an interesting look at how human and robotic musicians can create music in a cooperative manner. Drum machines are a very prevalent tool available to musicians but they do not typically create beats based on what is being played nor do they respond to subtle changes in the performance.

This project will provide a computer simulation of a musical performance with a human drummer in an improvisational style. Not all musicians have immediate access to band members with whom they can practice and compose music with. This system would allow for individuals to see how their musical compositions will sound with a drummer who is able to follow along and adapt to changing musical conditions. Newer musicians can also benefit from the experience of taking the leading role in an ensemble rather than playing along with favourite song or to a metronome. This project will initially focus on guitarists and pianists though additional instruments may be tested at a later stage.

The overall goal of this project is to develop a system which receives a musical audio signal as it is generated, determines the tempo and beat pattern of the signal, then proceeds to output a percussion accompaniment to the signal in real-time.

First, a beat detection feature needs to be implemented. This feature must derive the beat structure from an incoming audio waveform by detecting onset events within the wave. These onsets are defined by sudden energy increases within the waveform and typically coincide with a note or chord being played on the instrument that produces the waveform.

Once the beat structure has been determined, a fitting drum part will be generated in a predictive fashion so that it will be playing along with the human musician in real time. This drum part will be modelled on the components of a basic rock/jazz drum set.

B. Key Concepts

The concept of beat tracking refers to how a computer determines the tempo of a piece of music, much in the same way a human may tap their foot to the beat of a song. A computer accomplishes this through a waveform analysis in which sudden increases in energy are observed. These increases are known as onsets and the waveform can be translated into a format which displays the relative strength of these onsets [4]. Because these onsets typically coincide with initial hit of a note or chord, a tempo can easily be derived from a pattern of equidistant onsets as shown in Figure 1.

Artificial creativity is a rather abstract concept in which a model of human creativity can be represented algorithmically. Collins [2] employs a method of finding empty spaces within a musician's rhythm and creates counter-melodies within the spaces. This provides an interestingly layered and complex melodic structure that is meant to inspire the human musician into exploring new musical ideas. This results in a constant trade of ideas between human and computer rather than having the computer restricted to a supporting role. Another approach makes use of evolutionary computation and genetic algorithms to generate artificial creativity [6]. This method creates a number of randomly generated musical segments, picks from the most suitable segments and uses them to seed new segments. This process is repeated until an acceptable segment is found.

C. Key Requirements

This project will require the development of a robust and accurate beat tracking algorithm capable of tracking songs with non-constant tempos. It requires a suitable onset detection method from which the beat tracking algorithm can derive a tempo. The beat tracking process must work in real time with a streaming input so it must be able to accurately predict approaching beat locations.

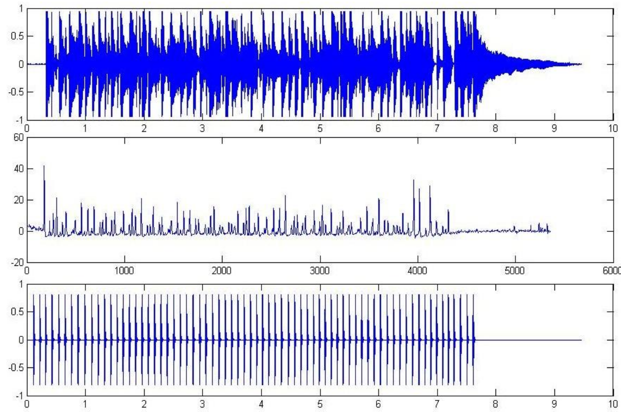


Figure 1. The top graph shows the raw waveform of an 8 second guitar part. The graph below it displays the waveform after it has been processed to display onset strength envelope. The bottom plot displays the beat derived from the onset envelope. Output derived from [4].

A suitable drumbeat template must also be created with a strong reliance on the derived tempo. This template will allow for the creation of basic and relatively complex beats while remaining flexible to changes in tempo. This will prevent the generated drumbeat from playing out of time.

In order to avoid repetitive and emotionless drumbeats, an algorithmic approach to mimic human expression will be implemented. This will also allow the drumbeat to respond to a musician's playing style accordingly.

II. Related Research

A. *Onset Detection*

Onset detection is a key aspect of the overall beat tracking algorithm. Bello *et al.* [5] discuss various approaches involving differences in energy and phase within the waveform. Each method has a potential application that is largely based on the type of input that will be received. One effective method is to observe a waveform's spectral features by performing a Fourier transform. The features derived from this method are useful in detecting onsets amidst relatively noisy and layered inputs.

Temporal features, which relate to a wave's amplitude, may also be observed. A valid onset typically occurs during a sudden increase in the waveform's amplitude. It is described in [5] how rectifying and smoothing the signal can help to accentuate these onset features. It is also argued that the wave should be filtered so that high amplitudes which are not part of a sudden rise will be lowered. This makes it easier to identify where the onsets actually occur. It is stated that analysis of the temporal features is a fast and efficient method of onset detection and is particularly useful when the audio signal is being produced by single, accented instrument such as a guitar or piano.

Ellis [4] discusses using an onset strength envelope which is essentially a filtered representation of the original waveform. The onset strength envelope used in [4] locates sudden energy increases within the waveform and represents them as individual spikes in the processed waveform. Higher spikes tend to represent valid onsets. The filtering process involves performing a short-time Fourier transform similar to the spectral feature analysis approach described in [5]. The signal is also passed through a high pass filter and is then convolved with a Gaussian envelope [4]. This seems to be an effective approach though it would need to be modified for real time applications.

Goto and Muraoka [1] attempt to recognize chord changes in a piece of music in order to determine the beat. By focusing on the lower end of the frequency spectrum, the onset events are likely to occur on chord changes rather potentially complex melodies which are typically of a higher frequency. While this method is relevant due to its execution in real time, it may not be robust enough to accommodate pieces of music which do not feature clear and discernable chord changes.

B. *Beat Tracking*

Ellis [4] describes a method to derive the basic beat based on the onset strength envelope. By finding a pattern of reoccurring and equidistant peaks in the onset envelope, a general tempo can be easily derived. The tempo calculation is weighted to prevent extremely distant and near peaks from forming improbable tempos. The derived tempo is biased towards 120 beats per minute that acts as probable middle ground for human created music. Extremely fast and slow tempos, while still possibly in time, would not represent the common human interpretation of the beat. This system has been implemented in Matlab and produces clicks to indicate beat locations in relation to the input waveform.

Collins [2] proposes a multi-agent beat tracking algorithm which follows a human musician in real time. This method uses a number of agents which are instances of predicted beat locations. Each agent has a score and weight to determine its accuracy, with these values increasing for each correct prediction Poorly performing agents are eliminated and new ones are generated based on the onset time. Scoring is weighted depending on the amount of time since the last onset to prevent subdivisions of a beat influencing the tempo. Collins' onset detection method is dependent on MIDI onsets rather than a waveform so it does not use the onset methods described in [5].

C. *Computer Music Composition*

Collins [3] also touches on the creation of a computer generated music. Collins describes how a series of tables with probability values which can represent different beats. Table I displays a template for a single measure of 4 beats, each divided into a set of 16th notes. With a probability value of 1.0, the note will be played on every iteration while a 0.0 indicates no chance of activation.

TABLE I. Probability Template

beat:	1		&		2		&		3		&		4		&	
kick:	0.7	0.0	0.4	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.3
snare:	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.5	0.0	0.2	0.0	0.0	1.0	0.0	0.0	0.3
hi hat:	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.7	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.7

Adapted from Collins [3]

This method allows for a non repetitive and more expressive drumbeat. Additional values can be attached to each location which affect volume and pitch if desired. The template will be set to synchronise with the beat pattern and will continuously update the time interval between notes as tempo changes in the musician's playing occur.

Ramirez *et al.* [6] provide an approach rooted in the concept of genetic algorithms. The authors use recordings of a professional jazz saxophonist as a training set which the computer generated composition will in the creation of basic rules. Because these rules are rooted in a particular style, the program will create a melody that is similar to that style while still being relevant to song it was created for. This method may be useful for the creation of drumbeats that are intended for a particular style of music.

III. System Design and Development

The general outline for the system is shown Figure 2. It receives an incoming audio signal from a musician then processes the signal to determine the beat. From there, the basic drumbeat template can be created, populated, and then output back to the musician. This process is continuous so that changes in tempo can be properly accounted for.

There are four primary sections which determine the structure of the application: input signal analysis, the beat tracker, the drumbeat generator, and an artificial expression simulator.

The input signal analysis stage processes the incoming musical waveform into a format the rest of the program can interpret. This step is vital to the beat tracking aspect. The beat tracker will determine the tempo of song based on a consistent pattern of onset moments occurring in the input signal.

A sampling of guitar and piano performances will be used to evaluate the beat tracker. Input is captured with an inexpensive computer microphone to ensure lower quality signals are acceptable. The performances will feature sample songs with different speeds, volume, and levels of complexity. Additional samples will contain non-consistent levels of these features in order to gauge the programs adaptability. This will accommodate musicians who are not able to play perfectly in time as well as pieces of music which require a varying tempo.

The drumbeat generator makes use of a probability template to generate a simple but varying drumbeat. This provides the groundwork for the artificial expression simulator to create interesting and relevant beats. This aspect makes further use of the input signal analysis to determine dynamics and the overall style of the piece.

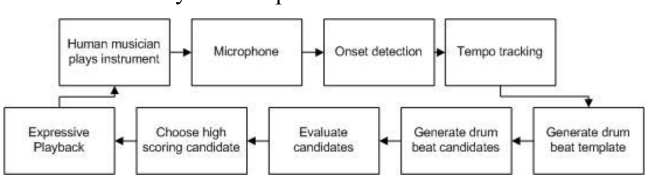


Figure 2. General flow of application.

IV. System Features

A. Input Signal Analysis

First, a suitable onset detection method must be implemented in order to correctly derive the tempo. By running the signal through a short-time Fourier transform, followed by a high pass filter, then convolving the signal with a Gaussian envelope, a suitable onset strength envelope can be created [4]. These methods accentuate possible onsets so that beat tracking is just a matter of peak selection and pattern detection. The input signal analysis must be executed in real time as there is no guarantee that a musician will be playing in a consistent style and tempo.

The waveform is monitored for gradual changes in amplitude, as these will determine the overall volume of the drumbeat. The rate of decay in amplitude between peaks will also be noted as it may contain information on the current style the musician is playing in. This information is contained in an average intensity value which is used by the artificial expression simulator. High rates of decay will increase the intensity value while lower rates will decrease it.

B. Beat Tracker

The methods described in [4], deriving a tempo from repetition of local peak patterns in the onset strength envelope, work well but would need to be further optimized to reduce latency. This is accomplished by having the beat tracker predict the beat for the next few seconds based on a short segment of the incoming signal. The tracker can then observe future segments as they are received and adjust its timing accordingly.

C. Drum Beat Generator

The drumbeat generator creates a drumbeat template similar to the probability template described in [3], but with additional features. Each beat in the template is determined by the beat tracker to ensure that any generated drum part remains in time with the human musician. In Table II, each beat contains four evenly spaced subdivisions which model a string of potential 16th notes. The template has been expanded to include a tom and cymbal track, volume adjustments for each track, and tone parameters for relevant tracks. The volume parameter allows for accents and ghosts notes while the tone parameter can create the effect of multiple toms and cymbals within one instrument track.

TABLE II. Expanded Drumbeat Template

Beat Template		Beat															
intr.	parameter	1		&		2		&		3		&		4		&	
kick	probability	1.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.4	0.4	0.0
	volume	0.8			0.8					0.8					0.7	0.7	
snare	probability	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	1.0	0.0	0.0	0.3
	volume					0.8						0.6		0.8			0.6
tom	probability	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.8	0.0
	volume														0.8	0.8	
	tone														0.6	0.4	
hi hat	probability	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.5
	volume			0.8		0.8		0.8		0.8		0.8		0.8		0.8	0.9
	position			0.3		0.3		0.3		0.3		0.3		0.3		0.3	0.7
cymbal	probability	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	volume	0.8								0.8							
	tone	0.3								0.5							

The distance between the two hi hat cymbals is represented by the position value. A value of 0 indicates tightly closed high hats while completely open high hats correspond to a 1. Values between 0 and 1 cover the remaining distance and allow for interesting crescendo effects and accents to set the overall style.

The template is initially populated by a genetic algorithm with a series of 1's and 0's, representing hits and non-hits. The simplest fitness function for this algorithm takes the number of kick and snare drum hits occurring on their expected beats (1 and 3 for kick and 2 and 4 for snare) over the number of offbeat hits, rewarding correct hit placement while punishing overuse. The time keeping track is scored for consistency. This fitness function works well for simple drumbeats, however, more intricate fitness functions are currently being evaluated, balancing complexity with how easy the drumbeat is to follow.

D. *Artificial Expression Simulator*

The values contained in the drumbeat template are assigned by the artificial expression simulator. These values are assigned in three separate stages in order to avoid unnecessary values being assigned to unused cells. Probability values are assigned first, followed by the volume, tone, and position values.

a) Assigning Probability Values

The probability values in the drumbeat template dictate how likely a hit in a given position will occur. These values, falling between 0.0 and 1.0, are determined by a genetic algorithm with a different scoring system for each individual drum or cymbal. Hits that are determined to be vital to the beat are more likely to occur on each iteration of the beat, thus having a higher probability value. The track designated as the time keeper (consistent hits) will also have higher probability values unless a particular hit falls outside the time keeping pattern. Any non essential hits will normally have lower probability values so they are not overused.

The number of onsets directly corresponding to beats can also influence a template's probability values. If most onsets are occurring directly on the beat, it can be assumed that the musician is playing a relatively simple rhythm. In this case, excessive and offbeat hits will be less likely to occur. The opposite is true in cases where many offbeat onsets are occurring, indicating a slightly more complex rhythm. A complexity value is assigned to this parameter to determine whether a simple or intricate drum beat is called for.

New templates will be compared with previous ones to prevent both drastic changes and excessive repetition within the various drumbeats. The probability values ensure a template can be used multiple times in sequence without too much repetition.

b) Assigning Volume Value

The next stage is to set the volume levels for each hit. A base volume level will be set based on the relative amplitude of the input waveform. This base value will adjust to whatever the current dynamic level of the input signal may be, much in the same way a human drummer will respond to another musician playing louder or softer. Another genetic algorithm will then proceed to adjust the volume levels in order to create accents and soft pickup hits. Scoring is slightly different for each component and fitness functions are still being verified. Volume levels will generally remain consistent unless changes are called for.

c) Assigning Tone and Position Values

The tom and cymbal tone levels will be set to a low value if the track is being used as the primary time keeper. This is to remain consistent with human drummers who typically will use lower-toned toms and cymbals as time keepers isolated tom and cymbal hits can have completely random tone levels otherwise.

The position value of the hi hat track largely relies on the input (specifically, the intensity rating). A waveform with a low decay rate following onsets is most likely corresponding to an instrument that is being played with smooth, sustained chords. If the high hat is being used as the time keeper in this section, the position value will be somewhere near the middle. If the waveform features a high rate of decay after an onset, then this section of music will be more suited to a closed high hat (lower position value). The higher position values are used for accents, which may also be indicated by a higher volume level for the hi-hat at that moment in time.

V. Conclusions

This paper presents the design and development of a virtual drum accompanist for musical composition and practice purposes. The ultimate goal is to have the virtual drummer seamlessly provide accurate and appropriate accompaniments to pieces of music as they are played. The system modelled on a human drummer who has no prior knowledge of a song but can still create a suitable and dynamic drumbeat to suit the piece. Through detailed waveform analysis and accurate beat tracking, the program can quickly compose and execute drumbeats while modelling some forms of expression exhibited by human musicians.

This is an ongoing project with a prototype system currently being validated. The artificial expression simulator is also undergoing constant analysis and tweaking due to its influence on drumbeats candidates.

References

[1] M. Goto and Y. Muraoka, Real-time beat tracking for drumless audio signals: Chord change detection for musical decisions, *Speech Communication*, Volume 27, pp. 311–355, 1999.

[2] N. Collins, Contrary Motion: An oppositional interactive music system. *NIME Conference,* Sydney, Australia, June 2010.

[3] N. Collins, Algorithmic composition methods for breakbeat science, in *Proc of Music Without Walls*, De Montfort University, 2001.

[4] D.P.W. Ellis, Beat tracking by dynamic programming, *Journal of New Music Research*, 36:1, pp. 51-60, July 2007.

[5] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, A tutorial on onset detection in music signals, IEEE transactions on speech and audio processing, vol. 13, no. 5, 2005.

[6] R. Ramirez, A. Hazan, E. Maestre, and X. Serra, A genetic rule-based model of expressive performance for jazz saxophone, Computer Music Journal, 32:1, pp. 38-50, Spring 2008.

[7] G. Weinberg, B. Blosser, T. Mallikarjuna, and A. Raman, The creation of a multi-human, multi-robot interactive jam session, NIME09, 2009.

A SMIL player for any web browser

Ombretta Gaggi and Luca Danese
Dept. of Pure and Applied Mathematics, University of Padua
via Trieste, 63, 35121 Padua, Italy
gaggi@math.unipd.it, ldanese@studenti.math.unipd.it

***Abstract*—SMIL, *Synchonized Multimedia Integration Language*, is a W3C markup language for the definition of complex multimedia presentations. SMIL documents need a specific player for its playback and cannot be rendered by modern browsers. In this paper we present *SmilingWeb*, a first tentative to implement a cross-platform player for SMIL presentations contained in web pages. We implement a JavaScript library, based on the web standards, which allows to solve the synchronization of media items contained in multimedia presentations through the use of any available browser. The proposed solution is cross-platform and cross-browser, therefore it can be potentially used by any user. The player has been tested with the SMIL Testsuite, provided by W3C and with a set of very complex multimedia presentations in order to check its support to the standard and its scalability. All the tests reported positive results.**

I. Introduction

SMIL [1], *Synchonized Multimedia Integration Language*, is a W3C standard that allows to create hypermedia presentations. SMIL describes both the temporal behavior of media items contained in a multimedia presentation and their spatial layout. Moreover, the playback of the media objects may be completely changed by the user, who can follow a link, click on an image, or even simply move the mouse over or out of an item.

Even if SMIL was not intended as a substitute for Adobe Flash®, it can be a valid alternative for simple animations or other synchronization of multimedia objects contained into web pages. However, more than ten years have passed after the first definition of this language in 1998, SMIL is used, as examples, to describe interaction in brazilian digital TV, as language to describe the MMS, *Multimedia Message System*, and in the *Daisy digital talking books*, i.e., book accessible to visually impaired users [2], but its diffusion in the web pages is still lower than expected. One of the key problems of this low diffusion is due to the complexity of the authoring activity [3], but also to the lack of a fully-featured SMIL player is a strong obstacle to its adoption. In fact, a SMIL document cannot be rendered by web browsers but needs a suitable player. At the moment, the only available player for the third version of the standard is Ambulant Player [4]. Unfortunately, Ambulat player is a stand alone-application, therefore it cannot be used to render SMIL presentation inside a web page. There exists a plug-in, but it is available only for a limited subset of browsers. Therefore at the moment, the multimedia presentations created with the standard SMIL cannot be rendered on web pages, but only as stand-alone applications. Moreover, the available players are often platform-dependent, therefore it does not exist a solution which works on all operating systems. Since web users are very heterogeneous in terms of used device, operating system and browser, this is a big problem.

In this paper we present *SmilingWeb*, a first tentative to implement a cross-platform player for SMIL presentations contained in web pages. We implement a JavaScript engine, based on the web standards, to allow the playback of multimedia presentations through the use of any available browser. The proposed solution is cross-platform and cross-browser, therefore it can be potentially used by any user.

The possibility to use SMIL in web pages is very important since, differently from Adobe Flash®, this standard can be used to improve accessibility of web pages. As an example, SMIL is suggested to create accessible subtitles for audio or video files. Moreover, SMIL allows to consider different alternative inside a presentation in order to better match users profile.

Our player does not fully support the third version of the standard, since a limited subset of features are not currently implemented, but the state of the implementation allows to render complex, interactive, multimedia presentations. The player has been tested using the SMIL 3.0 Testsuite with positive results.

The paper is organized as follows: Section II discusses background and related work. The implementation of the player is described in Section III and Section IV describes the tests made on the player. Finally, we concluded in Section V.

II. Background and related work

A. The SMIL language

A SMIL file is divided in two sections: the `layout` section defines the regions, i.e., rectangular areas on the user screen in which media items are visualized, and the `body` section contains the definitions of media items involved in the presentation, and the temporal relationships among them. The language SMIL also allows to define *transitions*, i.e., visual effects between two objects, and *animations*, i.e., modifications to the value of some attributes (e.g., the color, the size, the position, etc.) of a media item.

SMIL does not define a reference model for the data structure, but only tags for describing media objects behavior. Synchronization is achieved essentially through two tags: `seq` to render two or more objects sequentially, one after the other, and `par` to play them in parallel.

Using attributes `begin`, `end`, and `dur` it is possible to fix the start and the end time of a media item. Consider a media item inside a `par` block. If its attributes `begin`, `end` or `dur` are undefined, the media item starts at the beginning of the `par` block and ends with its natural termination. Otherwise it begins (ends) a certain amount of time after the beginning of the `par` block, given by the corresponding attribute value. In script 1, the attribute `begin = ''intro.activateEvent''` makes audio *artwork* start when the user clicks on video item *intro*. Therefore, the beginning and the end of a media item can be defined with reference to the tag in which it is contained, or according to a particular event, even completely changing the semantics of tags `par` and `seq`, as in the case of script 1. The attribute `dur` defines the duration of an object.

⟨par id="par1" dur="60s"⟩
⟨video id="intro" end="20s"/⟩ (1)
⟨audio id="artwork" begin="intro.activateEvent" dur="30s"/⟩
⟨img id="picture" begin="artwork.begin+5"/⟩
⟨/par⟩

The tag `excl` is used to model some user interactions. It provides a list of children elements, and only one of them may play at any given time. We refer to [1], [2] for more details about this standard.

B. Related work

Since SMIL's first appearance, many authoring tools have been implemented [2] offering their users different facilities like visual editors or preview windows and some players. The first available player was RealNetworks *RealPlayer*® [5]. RealPlayer, and its successor RealOne, implements the SMIL standard up to version 2.1. It was a commercial product, available both *freeware* and as a paid version (RealPlayer Plus). It is a stand alone application available both for Microsoft and Apple operating systems, even if in the second case, the users reported a longer list of bugs. Unfortunately, RealPlayer does not support the third version of the standard.

The only player that support SMIL 3.0 is *Ambulant* [4]. This player is available *freeware* for all operating systems as a stand alone application, or as a plug-in for the Mozilla FireFox and Apple Safari[1]. Tests made has shown that, despite the large support of the numerous features of the language, the application often does not work as soon as the complexity of the synchronization of media items involved in the presentations increased, therefore, the player is often unusable.

Valente et al [6] consider the problem of finding out temporal conflicts into SMIL documents and implement a player, based on a formal description of SMIL elements through automata, which enables the generation of a valid scheduling for rendering, considering QoS problem. The authors compare different players' behaviors in case of temporal inconsistencies and denote that they are implementation-dependent. This approach is not extensible: the automaton which describes the obtained behavior needs to be re-built after any changes.

Microsoft Internet Explorer® supported a selection of SMIL Boston[2] components till version 6.

Unfourtunately, the implementation of SMIL was abandoned in Internet Explorer 7 in favor of SAMI, *Synchronized Accessible Media Interchange* [7], a technology to provide closed captioning to a wide range of multimedia products, which works with Microsoft software only.

Some tentative approaches to create a cross-browsers SMIL player are documented in [8], but the proposed solutions, Soja and S2M2, are limited to the first version of the standard, therefore of no practical use.

Therefore at the moment, SMIL multimedia presentation can be played as a stand alone application with Ambulant player in any operating system, but it cannot be incorporated into web page, even if, both XHTML and SMIL are XML languages and therefore can be used inside the same document. This limitation is due to the state of software, i.e., browsers, implementation and not to the SMIL specification, therefore we think that future implementations of the standard will overcome this limitation.

We must note here that supporting SMIL is a very complex task, since this language provides a good expressivity, i.e. it allows to define rich and interactive multimedia documents, thus mastering its timing relations is not easy. To the best of our knowledge, no browsers support SMIL natively (only versions 5.5 and 6 of Microsoft Internet Explorer partially did in the past), but some of them declare to plan to support this standard in the future. Moreover, there exist JavaScript implementations of few SMIL modules (e. g. JavaScript implementation of smilText is available in the web page of the Ambulant Player) which make documents, using that modules, accessible with any browser.

III. SmilingWeb

SmilingWeb is a web player for multimedia presentations designed with the SMIL standard implemented as a JavaScript library. We aim at developing a solution suitable for all available browsers, therefore we paid particular attention to their differences in terms of support to HTML5 and other solutions like CSS3 and AJAX. The first problem that we have to consider was the possibility to play a video or an audio file without the need for a plug-in. HTML5 [9] offers the ability to easily embed media into HTML documents, using <audio> and <video> elements, but it is still a W3C working draft, and its support is still limited. Moreover, modern browsers provide support for these tags, but for a very limited subset of audio and video codec. For this reason, our player uses Modernizr [10], a small and simple JavaScript library that helps to take advantage of the emerging web

[1]A plug-in for Internet Explorer is currently available in the Ambulant web pages, but at the time of writing it cannot be installed on all available versions of Internet Explorer.

[2]SMIL Boston was a preliminary version of SMIL 2.0, supported by Microsoft, which aims at introducing transition effects and animations that are not included in the first version of this standard.

technologies (CSS3, HTML5) while still maintaining a fine level of control over older browsers that may not yet support these new technologies. Modernizr allows *SmilingWeb* to play a video or an audio file natively, thanks to HTML5 features, or to require a plug-in only if the browser does not support it. Modernizr also allows to avoid code forking and to design an application compatible with future development of browsers, since it does not test the browser name and version, but it tests its capability. This means that, when all browsers will support HTML5 features, *SmilingWeb* will not need any redesign.

A similar consideration can be done with the support for CSS3, which should be very useful to implement some transition effects. Differently from HTML5, new browsers begin to support CSS3 properties, but not using the standard notation, but adding a particular prefix related to the specific rendering engine[3]. Moreover, the set of transition effected supported deeply varies from browser to browser. This means that, the use of CSS3 implies that the transition effects applied to media objects can be present or not according to the browser. For this reason we decide to use another solution, the well-known JavaScript library jQuery [11], since the browsers support to CSS3 is still inadequate. Future versions of *SmilingWeb* will probably adopt the CSS3 standard as soon as browsers support will be sufficient. Moreover, jQuery simplifies HTML document traversing, event handling, animating, and Ajax interactions, therefore it also used to implement SMIL animations.

The *SmilingWeb* player has been designed to support SMIL 3.0. It does not implement all the tags of the language, but the subset of unsupported tags is limited. In particular, it supports:

- all tags and attributes of the Layout Module, with the only exception of the tag `regPoint` and the attribute `soundLevel`[4];
- all tags and attributes of the Linking Module, with the only exception of attributes used to control the volume;
- tags `img`, `video`, `audio`, `brush`, `text` and `ref` for definition of media items;
- all tags of the Timing Module, with the only exception of `priorityClass`;
- all tags of the Animation Module and
- the tag `transition` for the transition effects.

The complete set of supported tags and attributes is detailed in Table I.

Moreover, the player supports events handling and the temporal synchronization of text through smilText. We must note here that the player does not support all video and audio codecs, but this limitation is due to the set of codecs supported by the browsers and the plug-in. In fact, if the browser does not support HTML5, the flowplayer plug-in [12] is used, an Open Source (GPL 3) video player for the web.

[3]E.g. the property `border-radius` becomes `-moz-border-radius` for Mozilla FireFox while Apple Safari and Google Chrome have recently replaced their `-webkit-border-radius` with the standard property.

[4]We must note here that this attribute does not work even in commercial products like RealPlayer.

Supported SMIL elements		
LEGEND: LM = Layout Module, MM = Media Module, LkM = Linking Module, TM = Timing Module, AM = Animation Module, TrM = Transition Module, ST = SmilText Module, *= partial support		
Tags	LM	root-layout, region
	MM	text, img, video, audio, animation, brush, ref
	LkM	a, area
	TM	par, seq, excl
	AM	animate, animateMotion, animateColor, set
	TrM	transition*
	ST	smilText, tev, br, clear, div, p, span
Attributes	LM	height, width, backgroundColor, top, left, bottom, right, z-index, backgroundOpacity*, backgroundImage, backgroundRepeat, fit, showBackground
	MM	src, id, region, fill*, color, transition
	LkM	shape, coords*, sourcePlaystate, href, alt, tabindex, accesskey, show*
	TM	begin, end, dur, repeatCount*
	AM	targetElement, attributeName, from ,to, values
	TrM	transIn*, direction, dur, subtype, type
	ST	next, textAlign, textBackgroundColor, textColor, textFontFamily, textFontSize, textFontStyle, textFontWeight
Admitted Values	MM	fill: transition
	LkM	coords: a list of positive integer
	TM	begin: a positive time value t, an event ev, ev+t end: a positive time value t, an event ev, ev+t, indefinite dur: a positive time value t, indefinite

TABLE I
SMIL TAGS, ATTRIBUTES AND VALUES, SUPPORTED BY *SmilingWeb*

A. The scheduler

The most important module of *SmilingWeb* is the scheduler, i.e., the engine that solves the synchronization constraints of the presentation to find out the correct begin and end time of each element to be scheduled. By *element* we mean all SMIL objects which can be synchronized, i. e., media items, smilText tags, animations or transition effects. A *task* is therefore the set of operations needed to start an element.

The scheduler must be *correct* and *efficient* to avoid the user lies in wait for long time. Moreover, efficiency is particularly important in case of synchronization due to user interactions: it may happen that the scheduler had to re-calculate the entire scheduling after an event like the user clicking on an image. In this case, the playback could be paused or be subject to delays if the computation does not end in time. While the user is used to wait the beginning of playback of multimedia presentations, long response times after interactions are poorly tolerated.

The scheduler engine loads the SMIL file only once and saves all the information about media, transitions and animations into a tree structure. For each element, the collected information are the id, a pointer to the tag in which it is contained, if available, and the start and end time. If the element `begin` and `end` attributes contain a time value, they are simply saved in the tree structure, otherwise, they will be calculated according to their definitions.

SMIL language allows to synchronize elements according to two kinds of events, *internal* events, i.e., the begin or end time of another element, and *external* events, i.e., user's interactions, e.g., the user clicking on an image or keying a key on the keyboard. The first kind requires the scheduler to search the tree for the id of the referred object and simply retrieve the time value of its start (or end). The second kind is very different since, if the definition of a SMIL tag is bound to a user event for its start or end time, they must be solved *on-the-fly* during playback, therefore the scheduler does not return a point in time but the event contained in the definition.

Before the calculation of the start and end time of each element, the player pre-loads the images and other media files contained in the presentation, so to avoid pause during presentation rendering to wait for distributed media items[5].

Synchronization constraints are solved by the function $findRendering()$ which recursively considers all the tags of the presentation. Initially, it receives as input the entire SMIL file, and creates the tree structure. Then it calculates the start/end times of elements, starting by the first synchronization tag contained in `body` and current time instant equal to 0. More precisely, if the element el does not contain any references to external events, $findRendering(el, currentTime)$:

1) calculates el start time,
2) calculates el end time,
3) recursively applies $findRendering(c_i, time_i)$ to all the element's children c_i, with current time instant
   - equal to the one received as input in two cases, if el is a `par` or an `excl` tag, i.e., $\forall i\ time_i = currentTime$, or the current child is the first child of a `seq`, i.e., $c_1 = currentTime$;
   - equal to the time instant in which the previous child ends otherwise,
4) adjusts, if necessary, the end time of el[6].

Calculating the start and the end time of a SMIL tag is a complex task, since both values may depend on offset, internal and external events. A detailed description of different possible combinations of attributes' values are reported in [13]. We report here only a brief, and not complete, description of the algorithm chosen to solve the synchronization constraints introduced by SMIL tags.

The start time of the element is obtained by adding the offset contained in the attribute `begin` to the current time instant received as input. If the attribute `begin` refers to an internal event, the schedule retrieves the time instant in which that event occurs and calculate the start time of the element. The end time of an element is calculated as follows:

1) it is equal to the current time instant plus the value of the attribute `end`,
2) it is equal to the start time plus the value of the attribute `dur`,
3) it is obtained resolving the time instant in which the referred event occurs and adding the offset (if present).
4) if both the attributes `end` and `dur` are undefined, the element ends together with its father, or if not possible, when all its children have terminated their rendering, if the element is a time container, i.e., a tag `par`, `seq` or `excl`.

The scheduler engine creates a hash table, named the *scheduledTasks*, to insert all the start and end times that are calculated. The scheduledTasks is chronologically sorted and contains, for each time instant returned from the scheduler, the list of tasks, i.e., the elements to start in that point in time. Once the computation of the scheduledTasks is finished, for each time instant, the player:

1) performs the tasks which must be executed at that particular time instant;
2) if an external event has occurred, checks if it affects the synchronization of other elements and, if so, it runs the scheduler engine again to calculate the new scheduledTasks;
3) searches for the next point in time in which a synchronization takes place and sets up a timer to notify the scheduler once that point has been reached;
4) suspends waiting.

Every time the scheduler starts an element, if its end time is available, it also creates a timer to stop it. Therefore many timer can apply to the same media once started, e.g., a timer to stop the media itself and a timer to start the transition effect at the end of the media playback.

The timer used to end elements does not affect the main scheduler, but are managed by the element itself, which acts like a sub-scheduler, in order to improve efficiency. This is particularly useful in case of the tag `smilText`, which can handle the timed text defined inside it, through its inner schedule, thus simplifying the solution of their synchronization constraints even in case of events. We follow the rule *"divide et impera"*: the tag `smilText` is considered like any other element by the main scheduler, and the inner schedule applies a set of simplified rules to its children to find out their start and end times.

Even events bound to user interaction are managed through the use of a secondary scheduler. Each time an event occurs, the player calculates the start and end times of the set of elements affected by that interaction, and creates a secondary scheduler which managed the tasks related to those elements. This choice has been made since a second interaction of the same kind (e.g., a user who clicks twice on the same image) causes a second start of the same set of elements. The use of a separated scheduler allows to easily stop the first instance of the scheduler, and analogously the n^{th} instance, and to create a new one with the new value of the time instant in which the event has occurred. All these operations are performed without affecting the main scheduler which can continue working, without delays.

For efficiency purpose, when the scheduledTask must be re-

[5]We must note here that while this operation is trivial in case of images, the situation is more complicated for continuous media file, especially for videos which can be streamed over the network. We plan to better consider this issue in future works.

[6]Consider the case in which a `seq`, or a `par` tag, ends together with its last child.

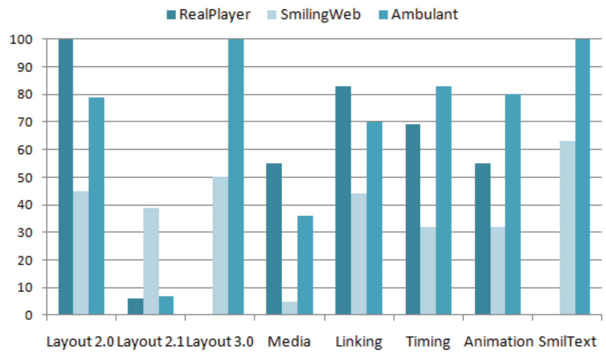


Fig. 1. Comparative analysis of the players with the SMIL 3.0 Testsuite

calculate due to a user interaction, the scheduler does not re-calculate the entire hash table, but only from the current time instant, i.e., the time instant in which the user interacts with the presentation. In this way, elements which are rendered in the past, no longer useful, are not considered again, preserving CPU resources.

IV. System Tests

SmilingWeb has been tested in two different ways. First of all, it was tested using the SMIL 3.0 Testsuite [14] with positive results. We test all the scripts of the testsuite with RealPlayer, Ambulant (both in the stand-alone application and in the plug-in version) and *SmilingWeb*. For each test, we rate the result with 0, 1 or 2 points depending on whether each player does not support tags or attributes contained in the script, experiments some problems, i.e., it provides only a partial support, or passes the test without errors. The result is shown in Figure 1, where the vertical axis represents the percentage of passed tests, and the horizontal axis divides tests according to the SMIL modules.

Figure 1 shows that, although RealPlayer's support to SMIL 2.0 features is very high, e.g. it passes 100% of tests on Layout 2.0, and 83% of tests on Linking module, its support to new features of the third version of the standard is very low or absent (e.g., it does not support SmilText). On the hand, *SmilingWeb* provides a good support to SmilText and Layout 3.0 Module (respectively 63% and 50% of passed tests). Also the support to the Timing, Linking and Animation Modules is rather good. We must note here, that, the score obtained by *SmilingWeb* is deeply influenced from how the scripts are designed. In particular, many tests regarding the Media Module used the attributes `clipBegin` and `clipEnd` which are not supported by our player. This means that *SmilingWeb* does not passed the tests, even if the particular element which the script wants to test is supported. Therefore the testsuite does not completely shows the level of support to SMIL of *SmilingWeb*.

Another consideration must be done: RealPlayer and Ambulant player are stand-alone applications, while *SmilingWeb* can be used to insert multimedia presentation into web pages. Although the authors of Ambulant player claim that a plug-in version of the player is provided, at the time of writing, it works only for Mozilla FireFox, therefore it is of no practical use on the web, where it is not possible to foresee the user browser.

Summing up, *SmilingWeb* passed about 40% of tests made. Despite this percentage is lower than other player, our player has been tested with Microsoft Internet Explorer (version 8 and 9), Mozilla Firefox (version 3.6 and 4), Google Chrome (version 10), and Opera (version 11), and works well with all browsers. All the performed tests are reported in [15].

A second series of tests consists in the execution of more complex multimedia presentations. In fact, the SMIL Testsuite allows to test what features of the standard are implemented, but each test is very simple and usually consists on few rows of code, without any nesting of tags. Our experience, made in more than 5 years of teaching and use of the standard SMIL, shows that, even if Ambulant Player obtained the higher score evaluating it with the testsuite, it is often unusable for complex multimedia presentations.

We test the player with five multimedia presentations, with different degrees of complexity. The presentations used as tests can be viewed at [15]. On average, the SMIL documents used contain about 200 lines of code (with a maximum of 443 lines) and the maximum level of nesting is 4. As an example, *"Focus on Elisa"* is a multimedia presentation about an italian singer. It begins with 36 colored blocks which move on the screen to compose the writing *"Focus on Elisa"* (see Figure 2), while the audio of a song of the singer plays in background. This intro was created with about 150 animations, 50 in parallel at the same time. Moreover, the documents also contain some menus or buttons with which the user can modify the normal behavior of the presentation. Other documents contain a karaoke with text animations synchronized with audio and images with transition effects. Another example describes the trip of the Fellowship of the Ring in the *'Lord of the Rings'* motion picture, through the use of text, images and animation of the path in a map.

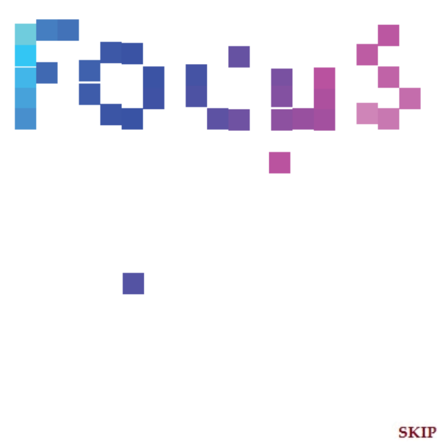


Fig. 2. Screenshot from the multimedia presentation *"Focus on Elisa"*

Even this second group of tests gives good results since the player is able to schedule and synchronize all the elements without any delay during rendering, even in case of user interactions. The user experiences a little delay before the presentation starts, but it is comparable to the loading time of common web pages. We test the presentations also with other available player, and we note that, when the presentation complexity increases, the Ambulant player begins to introduce delays or even errors in the synchronization of elements. As an example, the playback of *"Focus on Elisa"* described above, often experiences pauses and delays.

V. Conclusion

In this paper we have presented *SmilingWeb*, a JavaScript player which allows to reproduce SMIL scripts contained in web pages. The player has been tested with the SMIL Testsuite, provided by W3C and with a set of very complex multimedia presentations in order to check its support to the standard and its scalability.

All the tests reported positive results. However, we must note here that, even if *SmilingWeb* meets all the requirements in terms of scheduling of elements, in case of low network bandwidth, it is not always able to effectively calculate the correct interval of time needed to pre-load the elements. This is particulary true for continuous media like audio and video files. Therefore, in this case, it may happen that the user experiences some pauses during playback. This situation never happens for local document, or with an adequate network bandwidth. Therefore, the scheduling algorithm is correct, but we need to improve the analysis of the state of the network in order to calculate the correct time to start the presentation to have a good chance that all media items involved will be received on time. We plan to better analyze this issue in future works.

Another consideration must be done: although the other available players may have a more complete support to SMIL features, *SmilingWeb* is the first tentative to create a player for SMIL documents which can be used with all available browsers. Moreover, the technology used to implement the player have been chosen to ensure its compatibility also with future versions of the player, since we used web standards promoted by W3C.

The possibility to insert SMIL tags into web pages is particularly important, since it allows to apply synchronization to HTML tags, e.g., it allows to create a text moving around the screen, or to apply animations and transition effects without knowing JavaScript or Flash. Moreover, *SmilingWeb* allows to create accessible animations, or text description for audio and video files.

Concluding, we think that the possibility to use together XHTML and SMIL language allows the author to overcome some limitations typical of the two languages: in fact, SMIL allows the introduction of synchronization between media, but impose a fixed layout. XHTML and CSS on the other hand, define richer mechanism for layout definition, e.g., the definition of fluid layouts, i. e. web pages which adapt themselves to the size and resolution of the window on the user's screen.

References

[1] Bulterman et al, "Synchronized Multimedia Integration Language (SMIL) 3.0 Recommendation," December 2008. [Online]. Available: http://www.w3.org/TR/SMIL3/

[2] D. Bulterman and L. Rutledge, *SMIL 3.0, Flexible Multimedia for Web, Mobile Devices and Daisy Talking Books*, 2nd ed. Springer, 2009.

[3] D. Bulterman and L. Hardman, "Structured Multimedia Authoring," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 1, no. 1, pp. 89–109, 2005.

[4] Ambulant Open SMIL Player, "http://www.ambulantplayer.org/," 2009.

[5] RealNetworks, "RealPlayer 10.5," http://www.real.com/.

[6] P. Valente and P. Sampaio, "TLSA Player: A tool for presenting consistent SMIL 2.0 documents." in *Proc. of ICEIS2007*, Madeira, Portugal, June 2007.

[7] Microsoft Corporation, "Understanding SAMI 1.0," http://msdn.microsoft.com/en-us/library/ms971327.aspx, February 2003.

[8] SYMM Working Group, "Synchronized Multimedia - Players," http://www.w3.org/AudioVideo/♯SMIL.

[9] Ian Hickson, "HTML5, W3C Working Draft," http://www.w3.org/TR/html5/, April 2011.

[10] Faruk Ates, Paul Irish and Alex Sexton, "Modernizr," http://www.modernizr.com/.

[11] The jQuery Project, "jQuery," http://jquery.com/.

[12] Flowpalyer Ltd., "flowplayer," http://flowplayer.org/.

[13] A. Bossi and O. Gaggi, "Enriching SMIL with assertions for temporal validation," in *Proc. of ACM MM*, September 2007, pp. 107–116.

[14] W. Chang and T. Michel, "SMIL 3.0 Testsuite." [Online]. Available: http://www.w3.org/2007/SMIL30/testsuite/

[15] Ombretta Gaggi and Luca Danese, "SmilingWeb," http://docenti.math.unipd.it/gaggi/smilingweb/.

A Multimedia Tagging System to Index, Visualize and Retrieve Landscape Architecture Documents

Franck Favetta
National School of Higher Studies in Nature and Landscape Architecture (ENSNP)
Blois, France
franck.favetta@insa-lyon.fr

Robert Laurini
Research Center for Images and Intelligent Information Systems (LIRIS)
INSA of Lyon (Université de Lyon)
Lyon, France
robert.laurini@insa-lyon.fr

Abstract—**This article presents the outcome of an experimental process with the students of a Landscape Architecture university. We propose a multimedia geotagging system to help landscape architects during the analysis stage of a Landscape Architecture project. Our solution is specifically adapted to landscape architects requirements.**

Multimedia; geotagging; GIS; landscape visualization; Internet services; mobile application; GPS; collaborative application

I. Introduction

As computer science researchers working in collaboration with a Landscape Architecture university, our goal is to provide landscape architects (l.a.) with computer-aided solutions to help them in their work by fulfilling their specific requirements. This paper presents a proposed system to help l.a. during the analysis stage of a Landscape Architecture project. The Landscape Architecture University has been an ideal place for an experimental research process: we implemented a solution and tested it with our l.a. students, which provided us with useful feedback. We then used this feedback to improve our system. This article also presents feedback and improvements.

In the next sections, we will consider the specific requirements of l.a., then present a useful background and analyze a suitable solution for l.a.. We will then present our proposal, the use of our prototype by the students and our improvements.

II. Landscape Architects Requirements

This section presents two major requirements for l.a.

L.a. visit sites during the analysis and diagnostic stages of a project. Sites are often located far away from their office and the time spent on the premises is usually short, e.g., two or three days for a four-month project. During this short period, they need to collect as much information as they can. Equipment is light: a camera, a laptop and lately, often a smart-phone too. They take many pictures, record videos, draw sketches, obtain documents from local governments, interview the inhabitants, write texts and draw lists. Back at their office, l.a. need to memorize this large amount of collected files and quickly access specific ones.

L.a. need a way to capture the atmosphere of a place and the sensitive aspect of objects. Besides the use of 3D modeling or CAD applications, they also need to take pictures and to sketch landscapes generating geometrically distorted maps and views due to feelings and atmosphere. For this reason, l.a. often need to draw their own maps.

In a few words, l.a. need a way to order, visualize and retrieve a large amount of documents of different multimedia types and certain maps must be drawn by l.a. themselves.

III. Background and Analysis

Existing systems present notable features for our field of interest [1]. This section briefly describes them. 3D landscape visualization systems [2] offer the advantage of exploring existing sites and scenarios. Many GIS 3D extensions exist [3] as well as non-GIS-based systems.

We can classify landscape visualization systems into two kinds: those visualizing the landscape from the sky (orthographic or axonometric projections) and those visualizing it from the ground. We can also divide them into two other categories: those using 3D modeling and those capturing the real landscape thanks to pictures and videos. Some 3D modeling systems are textured with real landscape pictures and Augmented Reality fills the gap between both categories. Picture/Video-based systems better meet l.a. requirements as they better capture the atmosphere of places. 3D's perfect shapes usually provide l.a. with unreal atmosphere and feelings. L.a. appreciate axonometric projection like Bing Maps[i] and views from the ground.

Many 3D modeling systems exist. As an example, [4] presents a system to create ground, relief, vegetation and buildings, and [3] presents a survey of many GIS-based 3D landscape modeling products. Some systems use technologies like GeoVRML[ii] and CityGML[iii]. 3D modeling often requires a time-consuming process to create 3D landscapes that discourages l.a. from using it. Some systems automatically generate textured 3D models but they often require heavy equipment such as laser-range scanning systems [5] or cars mounted with cameras and video recorders [6].

Other kinds of systems based on the 1980's MovieMap system [7] and technologies like QuickTimeVr[iv] visualize landscapes from the ground with pictures. Pictures are taken with cameras or frames extracted from video recordings. In general, cameras and video recorders are mounted on vehicles to capture streets (often called "Google cars"). Some systems capture only the street sides and distort the captured pictures for a better geometry of selected points of view [8]. Most of them capture panoramas at different and close locations thanks to 360-degree capture devices. Most of them capture the streets of cities. The most well-known system is Google StreetView[v] but a growing number of rival systems already exist such as Streetside[vi] and iTowns[vii] or will soon come out, made by companies like TomTom[viii] and Nokia. Some of them capture the landscape away from the streets like MapJack[ix] and FlyAbout [9]. Some of them tag places with useful information (hotels, restaurants, etc), commercial offers or social networks messages ([10], UrbanDive[x]). Such tagging systems also exist for real-time process, e.g., Layar[xi] and WikiTude[xii]: useful indications around our own location augment the reality captured by a smart-phone (embedding GPS and Web-cam). StreetView-like systems are very useful for l.a. but they often fail to cover all the studied sites. Furthermore, as the picture is not taken by the l.a., it may not exactly correspond to some of the desired points of view, it can fail to show the precisely desired moment in the day (e.g., during the market in the morning), it can be a very old, out of date picture (e.g., there was no market on this square five years ago), or it can be insufficient to express the atmosphere of the captured place. The need for heavy equipment makes it difficult for l.a. to capture the studied site themselves with this method.

Another interesting solution requires nothing more than a camera. Some applications can overlap several pictures to create wider scenes and panoramas. For example, with PhotoSynth[xiii], the l.a. can visualize the whole landscape generated by the overlap of all the taken pictures. Another interesting feature of this system compared to StreetView-like ones is the navigation in this computer-generated landscape from one picture taken by the l.a. (with no distortion) to another. This solution requires no heavy equipment, but on the other hand, the number of pictures must be quite high and pictures must overlap.

With this background information and our analysis in mind, the challenge is to find a solution for l.a. to quickly and user-friendly capture a visited landscape themselves, visualize it, customize maps, and classify various multimedia documents.

IV. Proposed System

We propose a solution related to StreetView-like systems and the Photosync system but requiring only light equipment and no overlap of taken pictures. Pictures can be concentrated on one spot or spread over an area.

We propose to add graphical tags to the collected image files (Fig. 1). Tags can be placed on any background image that can be any map: drawn by l.a., obtain from local governments, downloaded from the Internet, exported from GIS, etc. It can be an orthographic or axonometric map. It can represent a place in the present or in the past. It can be touched up to visualize scenarios in the future. The background image does not have to be a map. It can also be a terrain section or just a collected picture with a chosen view from the ground. It can be a sketch or an interface designed like a Web page. Besides, tags are links between collected image files and other collected documents. L.a. can navigate from one background image to another, e.g., from one official map to a drawn one, then to a section, then to a picture, then to the textual description of an object in the picture. Linked documents may be images but also videos, sounds, simple texts, word processor documents, spreadsheets, and slide shows.

Graphical tags can be icons, areas, paths, or sections. A path can represent the video of a walking person. An arrow can be added to icons and paths to indicate the view direction of a picture or a video.

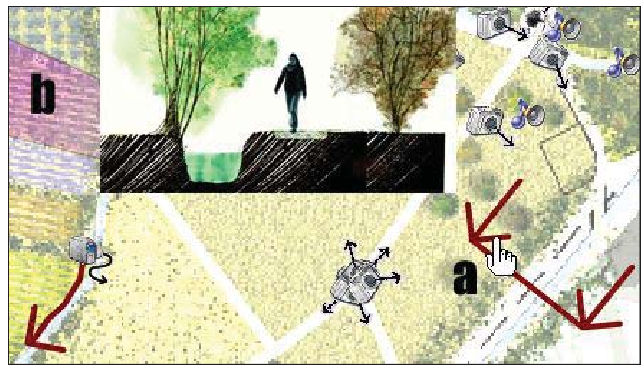


Figure 1. (a) A map section. As the mouse cursor touches it, its linked image is shown above; (b) An interactive area.

	Video and its trajectory		
	Oriented picture		Oriented sketch
	Text		Sound

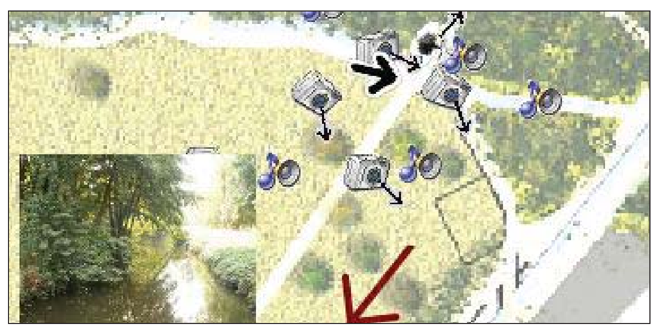


Figure 2. A *virtual walker* (the thick black arrow at the top) sees the picture in front of him and hears several surrounding sounds.

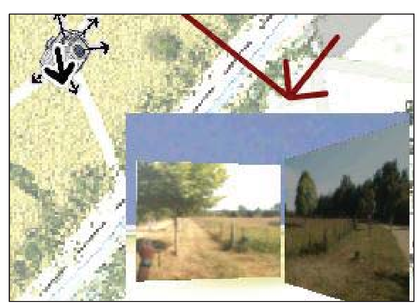


Figure 3. A 360-degree panorama in the *virtual walker 3D mode.*

A visualization of the linked documents is done on demand when l.a. move the mouse over a tag: images are viewed, videos are played, first pages of texts are shown, and sounds are played. We also introduce a virtual walker (*v.w.*) metaphor (Fig. 2) that simulates a person's walking on an image document map. When switching to *v.w.* mode, the mouse becomes the *v.w.*. A thick black arrow (see Fig. 2) indicates the direction of *v.w.* movement. The *v.w.* reacts to tagged documents: images are viewed and videos are played when approaching them on the map. The longer the distance is, the smaller the size of pictures and videos. Moreover, as the *v.w.* mode's goal is to simulate real movement, only oriented pictures and videos are displayed and only if their orientations approximately match the *v.w.* direction. In order to simulate the soundscape (sound landscape), the *v.w.* also hears every surrounding sound when approaching related sound icons. The closer the sound icon, the louder the sound is. Several surrounding sounds are mixed. This feature adds a real sound dimension to maps that usually only have a visual one.

The system also offers a *v.w.* 3D mode (Fig. 3). In 3D mode, the map becomes a floor and the *v.w.* can walk on it. In this 3D visualization, images are mapped on vertical planes like posters in an exhibition. They are oriented and only visible in the *v.w.* direction. The 3D model is automatically generated thanks to tags and no 3D modeling time is required for the user to create it. In *v.w.* 3D mode, l.a. can create 360-degree panoramas (Fig. 3) by placing pictures in a circle configuration around any chosen spot on the map. Realism can be lower than with other technologies, but the solution is quite fast and easy to use. The *v.w.* can explore the photographed area from one panorama to another like [9] and also on other spread spots showing only one or few pictures.

V. Real Use and Feedback

In order to communicate our solution to l.a. and to get their feedback, we have implemented a prototype. It has been used by the third-year students of a Landscape Architecture university for several real urban-planning contracts with several towns in France (Lille, Cluny, Montoire sur le Loir, Etel, Blois, Roubaix; about 190 students in all). For four of them, the University issued a DVD-ROM presenting different landscape analysis topics (hydrology, topography, vegetal species, networks, history, etc). Throughout these projects (one university semester for each), we followed the students' progress. They provided us with precious feedback. As they were free to use the prototype features, we could see which feature was useful for them or not and the way they actually used them. Different use cases emerged from this use. This section synthesizes them in the following illustrations. All of them are screenshots of real works done by the students with the prototype. We describe each use case in the captions.

As we mentioned above, we expected l.a. to use our system in order to structure their collected documents. We could actually observe that they rather used it to communicate their analysis to local authorities. We think this trend would change if pictures and videos could be tagged on site with a mobile version and located thanks to GPS and electronic compass.

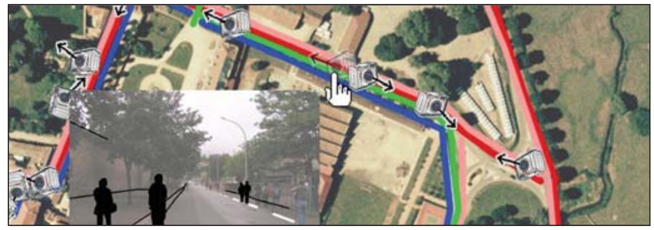


Figure 4. Views along a path - "Camera" icons indicate pictures taken along a path with different directions.

Figure 5. Creation of different sections of the site.

Figure 6. Inventories - "Camera" icons with a "flower" indicate macro-photos of the different plant species found on the site. Other inventories have been made for urban objects, shops, networks, etc.

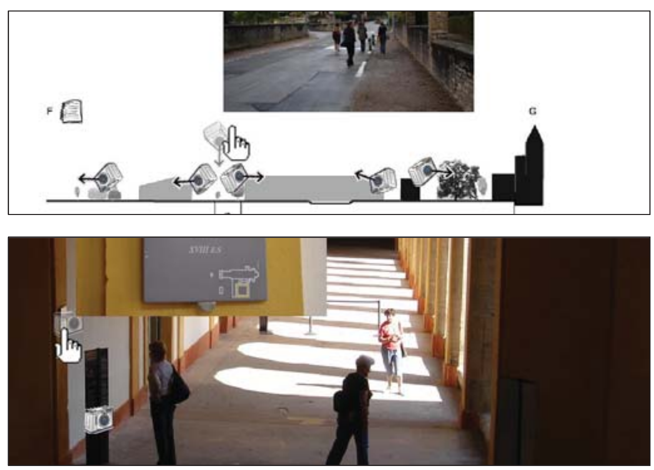


Figure 7. Background images that are not maps. Another interesting way to indicate a path: if icons had been placed at the end of the corridor, on the right and on the left, it could have offered a way to go to the next corridors.

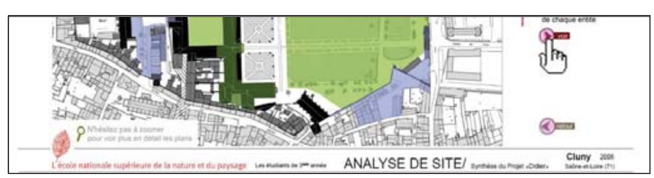


Figure 8. An interactive interface with buttons and menus.

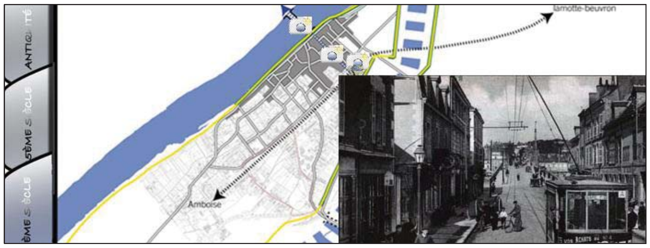


Figure 9. Tabs (on the left) to select maps at different times.

Figure 10. A *home page* for history and geography information.

L.a. asked for missing features that would improve the system. The icons represent the type of documents: images, videos, texts, etc. For each type, l.a. also need to represent different categories with different icons, e.g., a fixed video, a moving one, an interview, a slide show, etc. Besides, l.a. are very sensitive to icon design. L.a. found the *v.w.* solution interesting and asked for the possibility to record their virtual walks. They found the *v.w.* 3D mode interesting as well and expressed an interest for the addition of relief. L.a. appreciated being able to draw their own maps and to use any image such as pictures and sections as background. For certain situations, they also need a default map or a default empty background when they do not want to draw their own maps and backgrounds. L.a. often work in teams and have to communicate with many different participants (customers, local authorities, citizens, economic actors, contractors, etc). The participants have different interests and different rights of access to information. The system could be improved to help this communication and collaborative work.

VI. IMPROVEMENTS

Thanks to these real applications and the students' feedback we could improve our system with all their requirements.

A. Prototype Improvements

L.a. can use default icons and also customize their own icons. Digital Elevation Model and Triangulated Irregular Network (generated from topographic points) can be mapped with any image file generating relief in *v.w.* 3D mode. We also added a water level system in order to simulate different sea levels, river water levels and floods. Virtual walks and navigation through images can be recorded. We added the possibility to view several tagged documents at the same time, to customize the size of these visualizations and to fix them permanently onto the background. This visualization feature can also be recorded. We also had to compute a new image format in order deal with a high number of very large pictures.

We implemented these improvements and tested the prototype again with the students in the same conditions as before. We obtained a new interesting feedback from them. The following illustrations show different use cases we were able to observe. They are briefly described in the captions.

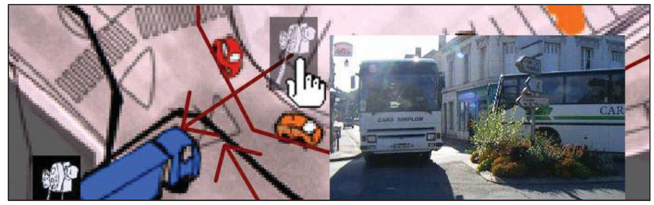


Figure 11. Video showing traffic and sound pollution at a junction.

Figure 12. Video showing a place at different hours of the day.

Figure 13. Landscape residents - Video of an interview.

Figure 14. Navigation in the streets like in the StreetView system but also away from the streets: the arrow on the left crosses the square.

Figure 15. The atmosphere of a theater with access to concerts, exhibitions and conventions information. While a music is playing, we can read the posters on the right wall and visit an exhibition on the left.

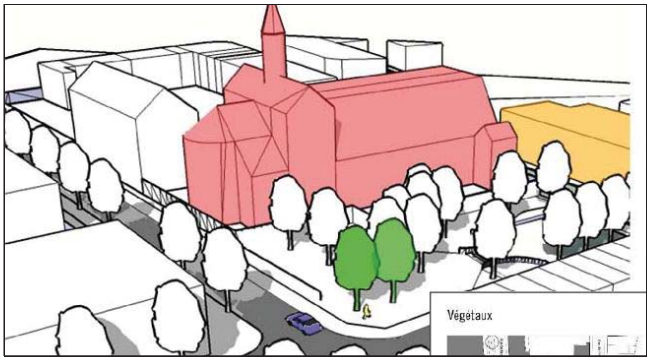


Figure 16. A visual portal. We can access to information about a church, trees, buildings and traffic when clicking on the related colored objects.

Figure 17. Relief of an estuary to give an idea of the field of view from a boat (relief does not need to be perfectly realistic).

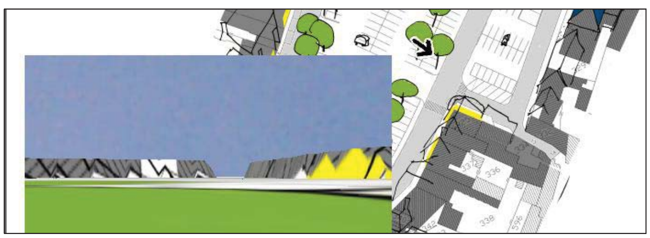


Figure 18. Coarse buildings to understand the field of view from a square.

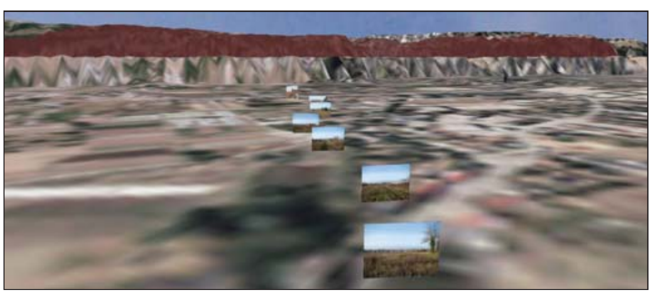


Figure 19. A path in *v.w.* 3D mode.

Figure 20. Simulation of the flood of the Loire river (France) showing part of the Town of Blois becoming an island. Different water levels on different areas can be moved up and down.

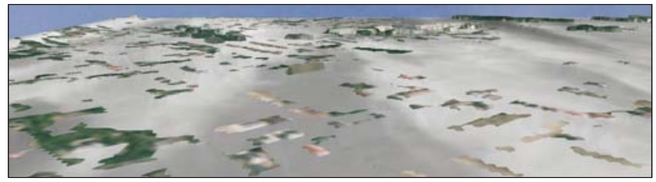


Figure 21. Water level is used as a spirit level to better understand the different heights of a landscape.

L.a. also used the prototype to realize multimedia presentations recording virtual walks, navigation through maps, visualizations of documents and oral comments.

As relief data are usually too expensive or, when they are free, offer a resolution too coarse to be useable, the students created them themselves thanks to the contour lines of maps. When little time is devoted to doing so, the result is imprecise but it is sufficient in many cases just to understand the geometry, the limits and the fields of view of the landscape. However, in certain cases, l.a. need a more precise relief and well textured 3D objects. In such cases, 3D modelers like SketchUp[xiv] are more appropriate.

B. Online, Mobile and Collaborative Version

We implemented a new version of our prototype that works online and on mobiles (Fig. 22). Thanks to a smart-phone, l.a. can capture and tag pictures and videos on site. They can also tag simple texts, word processor documents, spreadsheets, slide shows and Internet addresses. The prototype is based on five Internet Services: picture (Picasa), video (YouTube), office automation (Google Docs), database (Google Data) and map (Google Map) services. Captured pictures and videos must be uploaded to their related services to place them on the map (Fig. 23). The map service provides l.a. with a default map. The embedded GPS can geographically localize documents.

A project may be shared by several participants and each participant may handle different tag layers. Different l.a. teams can collaboratively work on the same project at the same time.

As a long time is required to implement it, this mobile version does not currently include previously proposed features such as different customizable backgrounds, icon customization, different graphical tags, *v.w.*, 3D. It has been written in javascript language, tested and optimized for Android Chrome Lite, FireFox and Internet Explorer but should work as well on any other browser, platform and device. All data are stored on the related services' servers (users need a Google account). Both prototypes (mobile and non-mobile) are available at http://franck.favetta.free.fr/landscape.

Illustrations are real prototype screenshots but the project example illustrated in this section was not made by the students. It is a demo made with a real project previously done by the students with the non-mobile prototype when they studied a square of *Montoire sur le Loir* town in France.

Figure 22. The online and mobile version of our prototype. In this example, two teams (market and circulation) tag multimedia documents at the same time on the same map of a square.

Figure 23. Selection of a picture and selection of a video.

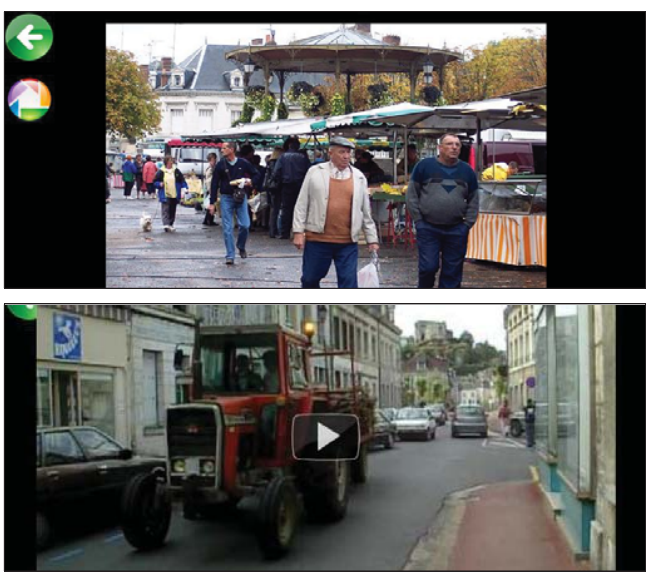


Figure 24. L.a. can view the images and play the videos.

VII. Conclusion

We presented a system to help l.a. to quickly and conveniently order, visualize and retrieve the documents collected during the analysis stage of a Landscape Architecture project in a way suited to their specific requirements. Our system takes advantage of existing solutions but requires no more equipment than a camera, a laptop or a smart-phone. The collected pictures do not need to overlap and no 3D modeling time is required. The outcome of this experimental work was reached thanks to the feedback we got from the use of our prototype by the students during real projects. We can now test our last online, mobile and collaborative system. We intend to transform our system into a service that could be embedded in a Web page. Our solution for l.a. collaboration could be improved to integrate communication with the numerous other participants in a Landscape Architecture project, in a global groupware solution.

References

[1] R. Laurini, "Information Systems for Urban Planning: “A Hypermedia Cooperative Approach,” Taylor and Francis, London, New-York, 308 p., February 2001.

[2] S. M. Ervin, H. H. Hasbrouck, “Digital Techniques for Landscape Visualization,” McGraw-Hill Professional Publishing, 2001.

[3] U.S. Army, “Survey of Terrain Visualization Software", U.S. Army Topographic Engineering Center, Topography,” Imagery and Geospatial Research Division, Data Representation Branch, 28 November 2005. http://www.agc.army.mil/research/software/survey/

[4] E. Lim, D. Kawashima, K. Umeki, T. Honjo, “A Visualization System for Urban Planning and Design,” Proc. Second International Symposium on Plant Growth Modeling and Applications, 2006, pp. 302 – 306.

[5] L. Rui, D. Burschka, G. Hirzinger, “Real time landscape modelling and visualization,” Proc. IEEE IGARSS 2007, pp. 1820 – 1823

[6] N. Cornelis, B. Leibe, K. Cornelis, L. Van Gool, “3D city modeling integrating recognition and reconstruction,” 8th conference on optical 3-D measurement techniques, July 9-12, 2007, Zürich, Switzerland, https://securewww.esat.kuleuven.be/psi/visics/home/visics

[7] A. Lippman, “Movie-Maps: An Application of the Optical Videodisk to Computer Graphics,” Computer Graphics 14(3), July 1980.

[8] A. Roman, G. Garg, M. Levoy, “Interactive design of multi-perspective images for visualizing urban landscapes,” Proc. IEEE Visualization 2004, pp. 537 – 544.

[9] D. Kimber and J. Foote, “FlyAbout, Spatially Indexed Panoramic Video,” Proc. ACM MM'2001, 2001, pp. 339 – 347.

[10] R. Karam, R. Kilany Chamoun, F. Favetta, R. Laurini, “Location and Cartographic Integration for Multi-Providers Location Based Services,” 25th International Cartographic Conference, Springer book ed. Palais des Congres, Paris, France, 2011.

---

i http://www.bing.com/maps
ii http://www.ai.sri.com/geovrml
iii http://www.citygml.org
iv http://www.quicktimevirtualreality.com
v http://maps.google.com/intl/fr/help/maps/streetview
vi http://www.bing.com/maps/explore
vii http://www.itowns.fr
viii http://www.tomtom.com
ix http://www.mapjack.com
x http://www.urbandive.com
xi http://www.layar.com
xii http://www.wikitude.org
xiii http://photosynth.net
xiv http://sketchup.google.com

Evaluating P2P Live Streaming Systems: the CNG Case*

Shakeel Ahmad[1] Christos Bouras[2] Eliya Buyukkaya[3] Raouf Hamzaoui[1]
Vaggelis Kapoulas[2] Andreas Papazois[2] Alex Shani[4] Gwendal Simon[3]

[1]Department of Engineering, De Montfort University, Leicester, UK
sahmad@dmu.ac.uk, rhamzaoui@dmu.ac.uk
[2]Computer Technology Institute & Press "Diophantus", Patras, Greece
bouras@cti.gr, kapoulas@cti.gr, papazois@ceid.upatras.gr
[3]Institut Telecom, Telecom Bretagne, France
eliya.buyukkaya@telecom-bretagne.eu, gwendal.simon@telecom-bretagne.eu
[4]Exent Technologies, Petach-Tikva, Israel
ashani@exent.com

Abstract

Many peer-to-peer (P2P) systems have been proposed for the provision of scalable live video streaming services over the Internet. While the literature contains surveys of the architectures of these systems, there is a lack of work on methodologies for their evaluation. We identify the main issues in the evaluation of P2P live streaming systems and use the Community Network Game (CNG) project as an example to illustrate them. The evaluation of the P2P system consists of two phases: a laboratory one using the ns-2 network simulator and an online field test with Massively Multiplayer Online Games (MMOG) players.

1. Introduction

Traditional client-server video streaming systems have critical issues of high cost and poor scalability. P2P networking exploits the upload bandwidth, computing power and storage space of the end users to reduce the burden on the servers and has been shown to be cost effective and easy to deploy. We identify the main issues in the evaluation of P2P live video streaming systems and use the CNG project as an example to illustrate them. The CNG project (`http://www.cng-project.eu/`) is an EU-funded research project that is focused on applying new network technologies to support community activities over highly interactive centrally managed MMOGs. CNG enhances collaborative activities between online gamers and develops new tools for the generation, distribution and insertion of User Generated Content (UGC) into existing MMOGs. It allows the addition of new engaging community services without changing the game code and without adding new processing or network loads to the MMOG servers. In particular, CNG proposes a P2P live video system to stream screen-captured video of MMOGs.

This paper presents the procedure that will be followed to evaluate the P2P live streaming system. Two phases are planned: a laboratory "offline" one using simulation software and an online one based on a real deployment. The online evaluation will be done by real gamers who will provide feedback through questionnaires. The performance of the system will be monitored by software, which will collect and provide useful information for further analysis.

The remainder of the paper is as follows. Section 2 provides an overview of the CNG P2P live streaming system. The plans for CNG laboratory experiments and online evaluation are presented in Section 3 and 4 respectively.

2. CNG P2P Live System

Allowing MMOG players to share their live game play with other players can have many useful applications. For example, skilled players can showcase their game to a large audience. Currently, the only platform that offers this service is Xfire (`http://www.xfire.com/`). Xfire captures the video of the game from the screen and sends it to a central server which broadcasts it live. However, this solution, which relies on central servers is expensive due to bandwidth and maintenance costs. To address these limitations, the CNG project proposes to use a P2P system. While many P2P live video systems have been developed, none of

*The research leading to these results has received funding from the European Commission's Seventh Framework Programme (FP7, 2007-2013) under the grant agreement no. ICT-248175 (CNG project).

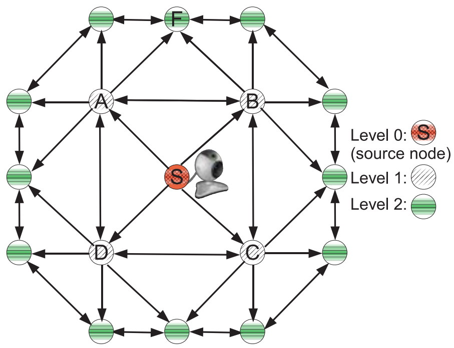


Figure 1. P2P topology.

them is suitable for the specific requirements of MMOGs:

- MMOG friendliness. The game experience should not be harmed by the P2P streaming. Thus, data communication with the MMOG game server must be given higher priority.
- Efficient management of multiple P2P overlays. Many MMOG players may simultaneously emit live streams, so the P2P overlay consists of many concurrent P2P overlays. A peer cannot participate in all P2P overlays because some of its resources will be used in every overlay it belongs to. The challenge for a user is to adequately allocate its physical resources, including upload and download bandwidth. These resources are limited, so they have to be shared carefully.
- Live video should be delivered at about the same time for all peers at the same "level". Moreover, peers in a higher level should in general be able to watch the video before those in a lower level. A level can be a priority class in a multi-tiered premium service. Alternatively, a level can be defined as the set of MMOG players that are in the same region of the virtual world.

In the following, we describe the CNG P2P system. The video of the game is captured in real time from the computer screen of the source and compressed. The resulting bitstream is partitioned into a sequence of source blocks, each of which corresponds to one Group of Pictures (GOP).

A mesh topology is used for the P2P network. This mesh is a directed graph $G = (V, E)$ where V is the set of peers, and (x, y) is in E if x may directly send packets to y. Peers are organized in levels. Level 0 contains the source. Level 1 consists of all peers that are direct successors of the source. In general, level k consists of all peers that are direct successors of level $k-1$ peers but are not in level $k-1$.

The UDP protocol is used as the transport protocol. The source applies rateless coding on each source block and sends the resulting encoded symbols in successive packets to level-1 peers until it receives an acknowledgment or a timeout occurs.

Packets are sent according to a scheduling strategy. The strategy specifies the maximum number of encoded packets n that can be sent by the source, the time at which a packet is sent, and a hierarchical forwarding scheme.

An example of a scheduling strategy for the P2P network of Fig. 1 is as follows. Packet 1 should be sent at time t_1 to A, which should forward it to B and D. Packet 2 should be sent at time t_2 to B, which should forward it to C. Packet 3 should be sent at time t_3 to C, which should forward it to B and D. Peer D should forward it further to A. Packet 4 should be sent at time t_4 to D, which should forward it to A and C. Peer C should forward it further to B.

Level-1 peers forward packets that are directly received from the source to their adjacent level-2 peers.

When a level-1 peer completes the decoding of a source block, it sends an acknowledgment to the source. Then it applies rateless coding on the decoded source block and starts acting as a source for level-2 peers that are its direct successors. The same procedure applies to peers at the next levels.

3. Laboratory Experiments

In the laboratory experiments, the CNG network solution with be evaluated with simulation software.

3.1. Metrics

To evaluate the performance of P2P video streaming systems, the following metrics were used in the literature: (1) Start-up delay: delay between the time a user joins a P2P system and the time it starts playing back the video [1] (2) Playback lag: time difference between the playback position of the source and that of the receiving peer [2] (3) Failure rate: probability that a user is rejected when it tries to join the system [1] (4) Continuity index: ratio of the number of video blocks that are available at their due playback time to the number of blocks that should have been played back by that time [1] (5) Peak Signal to Noise Ratio (PSNR) [3]. The PSNR is a standard video quality metric computed as: $PSNR(dB) = 10 \log_{10} \frac{255^2}{MSE}$ where MSE is the mean squared error between the original frame and the reconstructed frame (6) Percentage Degraded Video Duration (PDVD) [4]. The PDVD is the percentage of received frames whose PSNR is more than 2 dB worse than the PSNR of the corresponding encoded frames.

Both the PSNR and PDVD can easily be computed in the lab experiments as all required videos (original, encoded, received) are available.

In addition to measuring mean values of these metrics, we plan to divide CNG users into classes according to bandwidth, and measure minimum, maximum and variance values of the metrics for each class.

3.2. Examined Aspects

Scalability: Scalable systems are characterized by the property that the usage of resources is independent of the size of the system. Simulations for P2P live video streaming systems rarely consider more than several thousand simultaneous peers. Indeed, it is assumed that if the system scales at this stage, it is highly likely that it will scale to more peers. For the simulation of the CNG P2P video system, we plan to consider at least 2,000 peers.

Heterogeneity: Measurements show that P2P systems are characterized by a very high diversity of participating peers [5]. The variability of the *upload capacity* is considered as the major challenge in terms of heterogeneity, because it requires specific strategies in order to leverage the high capacity on some peers, and serve the peers with low capacity [1]. In [6], the average upload capacity of peers using BitTorrent has been found to be 180 kbps, while [7] shows an average upload capacity of 150 kbps. Two empirical distributions of peer upload capacities, one of broadband hosts (`http://www.dslreports.com/archive`) and one of BitTorrent hosts [8] are similar and are well modelled by a log-normal distribution [9]. Taking into account the above studies, the upload capacity in our experiments will follow a log-normal distribution with parameters μ in $\{150, 500, 1000\}$ and σ in $\{0.2, 0.9, 1.4\}$.

Churn: Based on the results from [10] and [7], we propose to use an exponential distribution of Time to Live (TTL) to model the churn rate. The parameter of this exponential distribution should result in approximately one tenth of the peer population refreshed at every simulation unit (which can be fixed at 1 hour). This setting, which is higher than what has been measured in [10], represents a system with a high churn. We will also evaluate our system under lower churn rates.

3.3. Simulation Tools

We considered several simulators including: ns-2, ns-3, P2PSim, Overlay Weaver, PeerSim, PlanetSim, Neurogrid, Query-Cycle, and Narses. We have assessed the available simulators based on several criteria, like simulator architecture, usability, scalability, statistics, underlying network simulation, and system limitations. We also considered the use of a testbed, like Planetlab. However, testbeds do not permit complete control of the system (e.g., for churn purposes) and, additionally, the scale of the experiments is severely limited. Taking into account the above, we decided to use the ns-2 simulator for CNG laboratory experiments.

3.4. Simulation Environment & Settings

A network game model for Counter-Strike is proposed in [11]. It is reported that 3–4% of all packets in a backbone could be associated with only 6 popular games [12]. One major concern is the upstream bandwidth. The peak percentage of traffic contributed by clients playing Counter-Strike to the total UDP traffic in the upstream direction can go up to 12% [13]. The work in [11] provides a simple traffic model for fast action multiplayer games. The game traffic model consists of only two independent modules, the client traffic model and the server traffic model with a burst size equal to the number of clients participating. In [14], the Extreme Value distribution has been identified to fit best for Quake traffic and other measurements have shown that newer MMOGs have bandwidth requirements that surpass those of older games [15].

It is important to identify user behavior with respect to network gaming. The study in [13] shows the percentages of subscribers playing Counter-Strike in various markets and identifies the trends in user behavior versus the day of the week. The common trend for all the markets is that this proportion increases as the weekend approaches, peaking on Friday, Saturday or Sunday. Also, there is a period of time in a day when there are very few to no users playing games. The exact hours vary depending on the market, indicating varying user behavior. The authors of [16] have filtered and analyzed MMOG traffic of the three-day long passive measurement, which contained about 200 World of Warcraft (WoW) flows and 100 other MMOG flows. Finally, [17] analyses a 1,356-million-packet trace from a sizable MMOG called ShenZhou Online.

3.5. Comparison to Similar Systems

Comparing the CNG system to state of the art commercial systems such as PPlive (`http://www.pptv.com/`) would present a major challenge as an implementation of these systems in ns-2 is not available. Moreover, the existing systems have not been designed for the CNG envisioned application. As an alternative, we propose to compare the CNG system to two related systems [18, 19] which although not designed for an MMOG environment can be implemented in reasonable time by making appropriate changes in the CNG system implementation.

4. Online Evaluation

In the online evaluation, MMOG players will be asked to answer an online questionnaire to describe their experience. In addition, the performance of the P2P network will be monitored using online tools.

Evaluation tools will be used for monitoring the activities and the traffic generated by the CNG system. Packet analyzers like Wireshark (`http://www.wireshark.org/`) are commonly used for traffic monitoring. In [20], Wireshark is used for analyzing and modeling the traffic generated by WoW. In [15], Wireshark was running on a computer along with Second Life, capturing game packets while filtering out irrelevant packets. On the other hand, built-in packet sniffers are useful for capturing the traffic and investigating protocols or applications in a non-intrusive way. When the source code of the application is available, traffic analysis can be done by adding proper logging and profiling functions to the source code [21].

In the online evaluation of the CNG network solution, Wireshark will be used to monitor the MMOG and other background traffic. On the other hand, the P2P network will be monitored through modules that will be developed in order to track user activities as well as the generated traffic.

For the online evaluation, a sample of 100-200 players will be recruited through forum posts and direct invitations to MMOG communities. An online user guide will be made available to help the players install and use the CNG tool. A number of evaluation measures will be used before, during and after the evaluation trial. The evaluations will be conducted online, allowing the project to obtain fast feedback. Participants will then be requested to continue their MMOG play and to augment it where they wish to with the CNG tools available to them. The usage period will be in the region of 4 weeks. Online players will be able to provide their feedback through a questionnaire where they will describe their experience with the CNG system, indicate any technical problem they encountered and suggest any further enhancement for the system.

5. Conclusion & Future Work

We have presented the plannning for the evaluation of the CNG P2P live streaming system. The process consists of two phases: laboratory experiments and online evaluation of the integrated system. The paper also includes a survey on the state of art for the evaluation of similar systems.

The next step of this work is the execution of the above plans. Our goal is that the execution of CNG evaluation will not only lead to significant results on the CNG system performance, but it can also introduce innovative ways of working for the evaluation of similar systems.

References

[1] B. Li, S. Xie, Y. Qu, G. Keung, C. Lin, J. Liu, and X. Zhang. Inside the new coolstreaming: Principles, measurements and performance implications. In *Proc. INFOCOM 2008*, pages 1031 –1039, Apr. 2008.

[2] D. Purandare and R. Guha. BEAM: An efficient peer to peer media streaming framework. In *Proc. Local Computer Networks '06*, pages 513–514, 2006.

[3] V. Padmanabhan, H. Wang, and P. Chou. Resilient peer-to-peer streaming. In *Proc. 11th IEEE International Conference on Network Protocols*, pages 16 – 27, Nov. 2003.

[4] 3GPP. Technical specification group services and system aspects; video codec performance (rel. 9), TR 26.902 v9.0.0. Technical report, 2009.

[5] J. Sacha. *Exploiting heterogeneity in peer-to-peer systems using gradient topologies*. PhD thesis, University of Dublin, Trinity College, 2009.

[6] O. Luzzatti. P2P live streaming for the masses - deployment and results. In *Proc. P2P-TV '07*, 2009.

[7] S. Khirman and V. K. Gurbani. Swarm analysis: 7 days of two popular swarms. In *Proc. P2PRG*, pages 26–31, Jul. 2009.

[8] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani. Do Incentives Build Robustness in BitTorrent. In *Proc. NSDI'07*, 2007.

[9] A. Bharambe, J. R. Douceur, J. R. Lorch, T. Moscibroda, J. Pang, S. Seshan, and X. Zhuang. Donnybrook: enabling large-scale, high-speed, peer-to-peer games. In *Proc. SIGCOMM '08*, pages 389–400, 2008.

[10] H. Liang, R. N. Silva, W. T. Ooi, and M. Motani. Avatar mobility in user-created networked virtual worlds: measurements, analysis, and implications. *Multimedia Tools Appl.*, 45:163–190, Oct. 2009.

[11] J. Färber. Network game traffic modelling. In *Proc. NetGames '02*, pages 53–57, 2002.

[12] S. McCreary and K. Claffy. Trends in wide area ip traffic patterns - a view from ames internet exchange. In *ITC Specialist Seminar on IP Traffic Modeling, Measurement and Management*, pages 1–11, 2000.

[13] A. Sinha, K. Mitchell, and D. Medhi. Network game traffic: A broadband access perspective. *Comput. Netw.*, 49:71–83, Sep. 2005.

[14] M. S. Borella. Source models of network game traffic. *Computer Communications*, 23(4):403 – 410, 2000.

[15] J. Kinicki and M. Claypool. Traffic analysis of avatars in second life. In *Proc. NOSSDAV '08*, pages 69–74, 2008.

[16] G. Szabó, A. Veres, and S. Molnár. Effects of user behavior on mmorpg traffic. In *Proc. ICC'09*, pages 1531–1536, 2009.

[17] K.-T. Chen, P. Huang, C.-Y. Huang, and C.-L. Lei. Game traffic analysis: an MMORPG perspective. In *Proc. NOSSDAV '05*, pages 19–24, 2005.

[18] M. Grangetto, R. Gaeta, and M. Sereno. Reducing content distribution time in p2p based multicast using rateless codes. In *Proc. Italian Networking Workshop*, pages 1–12, Jan. 2009.

[19] C. Wu and B. Li. rStream: Resilient and optimal peer-to-peer streaming with rateless codes. *IEEE Trans. Parallel Distrib. Syst.*, 19:77–92, Jan. 2008.

[20] M. Suznjevic, M. Matijasevic, and O. Dobrijevic. Action specific massive multiplayer online role playing games traffic analysis: case study of world of warcraft. In *Proc. NetGames '08*, pages 106–107, 2008.

[21] W. Acosta and S. Chandra. Trace driven analysis of the long term evolution of gnutella peer-to-peer traffic. In *Proc. PAM'07*, pages 42–51, 2007.

CAPTAIN: A Context-Aware system based on Personal TrAckINg

Reinaldo Bezerra Braga
Joseph Fourier University (UJF)
Grenoble Informatics Laboratory (LIG)
braga@imag.fr

Hervé Martin
Joseph Fourier University (UJF)
Grenoble Informatics Laboratory (LIG)
herve.martin@imag.fr

***Abstract*—We propose the CAPTAIN: A Context-Aware system based on Personal TrAckINg. The motivating concept is to map the trajectory of a mobile user, adding contextual information related to each user position. The CAPTAIN is divided in three parts, the mobile application, the desktop application and the server. The mobile application performs the yacht tracking, associating contextual information with each position. The second part is a desktop application, which process data in order to export both trajectory and user's annotations to a website. In this paper, we relate this experiment and we analyze it according to user and programmer point of view. The results presented can be a valuable instrument for software designers in planning context-aware system based on personal tracking deployment.**

I. Introduction

The context of a specific application might take into account information about the user location, identity of nearby objects and people, date, season and temperature, physical and conceptual statuses of interest to the user, or any information relevant to a user interaction with an application [1]. However, it is important to observe that a large number of location-based services are offered in distinct applications. Therefore, there is still a lack related to adaptation and association among the information and data obtained by these applications. Besides that, several proposals have been presented to overcome the obstacle of automatically organizing a personal geo-referenced data collection, in order to achieve an efficient browsing for photos, videos and audios jointly with annotations.

Based on the need to collect, treat and publish contextual information about the user position, weather, velocity and other data stemmed from sensors, we developed the CAPTAIN:A Context-Aware system based on Personal TrAckINg. This system is composed by three applications: a mobile application to collect and register the contextual information related to every geographical position; a desktop application to get each data collected by mobile device in order to add and infer about a new information; and a web application, in which receives the data sent by the desktop application and publish it on the web. For the purpose of validating our proposal, we tested the CAPTAIN as an electronic logbook in the ZeroCO2 project. The main objective of the ZeroCO2 project is to sail around the Mediterranean using a yacht powered by a clean carbon-free auxiliary motor to replace petrol motors, which are commonly used in yachts [2].

The results presented in this paper touches several research aspects according to the programmer point of view. Foremost, the spatial and temporal metadata-based access to content is been exploited by the CAPTAIN system. For instance, our system supports spatio-temporal exploration of data and contextual information by means of interactive maps and displayed relevant content. In a similar fashion, it allows an alternative technique to explore previous spatio-temporal data in order to obtain new information, using a HTML parser. Besides that, we use our system to perform the essential content annotation by way of the desktop application, i.e., relating metadata to content in terms of user's requirements. Therefore, we offer a collection of procedures that could be reused for storing contextual information and data, annotating text, and inferring about new information. In relation to a general text mining mechanism, we can use our approach for geocoding content, recognizing specific key words and relating them to coordinates.

In additional, the results show that our system collaborates with the user to collect, treat and publish all yacht trajectory, adding contextual information, such as position, date, temperature, speed, course, annotations and photos. Consequently, the user demonstrated that he is comfortable using the CAPTAIN system and found it useful as an excellent tool to publish the contextual information based on the geographical position. Besides that, we conclude that our proposal can be used in several types of scenarios, for example: to track an excursion in forests and mountains; to study the behavior pattern of a vehicle based on its speed, course, position and contextual information; and to map the course of runners and other athletes.

II. Related Work

The information available from a data can be usually organized into two groups: content-based and context-based. Content-based information about a photo is composed of low-level properties, such as texture, color, shape, etc. and high-level properties that are derived from the pixel intensity. Context-based metadata makes reference to the information about the photo. Location and time mainly provide spatio-temporal information [3]. Albeit they provide indications about the scene (for example, the location can indicate indoor or outdoor and the time can indicate day or night) the association

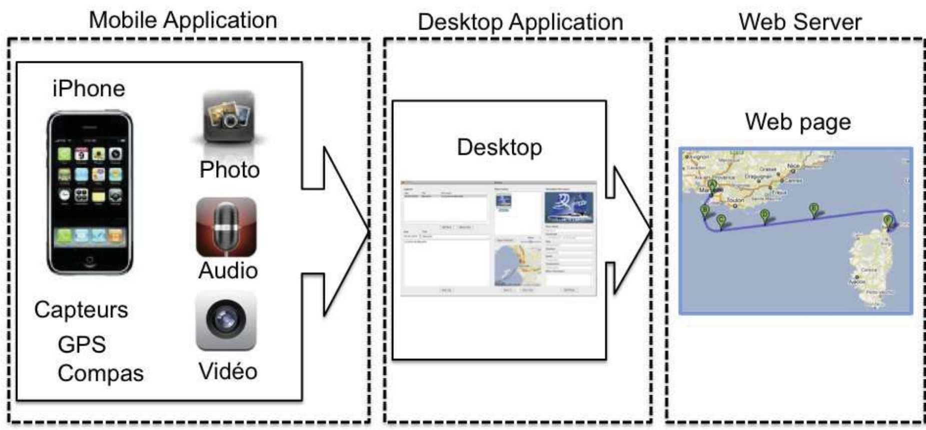


Figure 1. The CAPTAIN infrastructure.

among these properties is still limited. In general, an event is defined as a group of data and information collected in a relative time interval. In [4], the authors use attributes such as visual appearance, GPS locations, user-assigned tags, and dates to filter and group photos as a solution to support the users in serendipitously finding multiple relevant photos. On the other hand, Google Picasa [5] organizes the photos only by the date information. According to the proposals [6] and [7], it is important to study personal and social factors to use and organize personal data, such as photos, videos and audios. Bentley et al. present a solution based on how people search and browse data through collections of personal photographs and music collections [6]. Frequently leading to photo sharing, Kirk et al. observe the user activities around photos after being taken, in preparation to end-use [7]. Based on the idea of offering a solution for annotating personal photos, the PhotoMap was presented in [8]. The main objective of this work is related to offer a solution for personal photo annotations using OWL-DL ontology. It is presented an ontology called ContextPhoto and a contextual photo annotation approach as a way to improve the development of personal image management tools.

III. CAPTAIN

In spite of the large number of location-based services, we observe that the information of the navigators, tracking mechanisms and recommendation systems are available through specific application and there is a strong necessity to gather all this information in a single application. Based on the necessity to offer a system able to adapt, infer and publish the information registered by these applications we propose the CAPTAIN: A Context-Aware system based on Personal TrAckINg. Figure 1 presents the infrastructure of our system.

As indicated by Figure 1, the CAPTAIN uses the Mobile Application to capture the user position jointly with date, course and velocity, using a tracking mechanism provided by the smart device. Besides that, it can include other types of data, such as photos, audios and videos according to user's requirements. These information and captured data are sent to the Desktop Application, which is capable to infer new information, such as temperature, humidity, wind speed and others. Such information is used to help the user to annotate his own documents like a photo or a video captured using his smart phone. Finally, the annotations are sent to a website in order to be published on the web.

We created a mobile tracking application by registering the user trajectory. We designed a mechanism to capture photos, audios and videos and register contextual information in order to associate them with the current position. In the first stage of this work, we limit the scope of CAPTAIN to digital photo but it could naturally apply to other media such as audio and video.

The mobile application stores all information and data related to the trajectory in a predefined folder. When the mobile device is connected to the desktop computer, it is possible to run a synchronization process to make an update. Then, the desktop application proposes an interface to treat and to add information in accordance with data collected by the mobile phone. Annotations and data are stored in the desktop during the time interval without Internet connection. Since the connection is up, the user can obtain contextual information and export all annotation to the server. The Desktop Application was developed in the Objective C programming language under Mac OS X version 10.6.4.

The Server Application is responsible for registering the annotations and data sent by the Desktop Application. This information is used to put on public view the trajectories, annotations and photos of the mobile user on the web page. All information are stored in a MySQL database and requested by the Java and PHP scripts.

In addition, we offer a map interface for navigation through small clouds to present the annotations, photos and contextual information related to the geographical position. We created the CAPTAIN web page making use of map-based interfaces taking into account the usability studies presented in the literature [11] and [12]. These works show that map interfaces

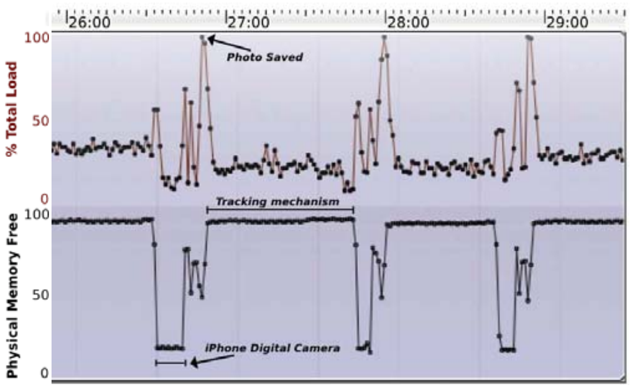


Figure 2. Physical Memory Free and Total Load in the iPhone.

demonstrate more interactivity advantages than browsing information with hierarchical links. Moreover, with a map-based interface, we could easily illustrate the trajectories generated by the mobile users jointly with the contextual information.

Besides that, the CAPTAIN server application performs also an indexation process to improve the browsing and interaction procedures. The amount of collection containing data and contextual information increments quickly in our system. Then, in order to avoid future performance difficulties related to the large number of access, spatial and temporal indexes are associated with each annotation in the MySQL database.

IV. Evaluation Results

For the purpose of evaluating our system, we evaluated the CAPTAIN in the ZeroCO2 project environment. The evaluation was performed during a travel from Marseille to Ajaccio (Corsica Island). Figure 2 shows the performance evaluation of the CAPTAIN in the mobile phone during the interval from 26 to 29 minutes. The evaluation was performed during the first tests, using the XCode Instruments [13] version 2.7. We observed that the *Total Load* (i.e. System and User) and the *Physical Memory Free* followed the same behavior while the mobile application functions were in operation. According to the results, the tracking mechanism requires approximately 10% of the memory and 25% of the processing to capture and register the positions. Likewise, while the iPhone digital camera is working, the memory used is approximately equal to 80% and the total load did not change. After taking the photo, the function *Save Photo* can be selected. When the *Save Photo* function is activated, the maximum load is used in order to associate and register all data and contextual information in the hard disk. Finally, the memory is cleaned when all data and information are associated and saved and the total load returns to follow the tracking mechanism. These results were important to guarantee that the user can utilize the application for long time without stopping it due to memory or processing overhead problems.

Two important contributions are related to the desktop application. The first one is the *Weather* function, in which the user is able to obtain previous weather information that are published on the webpage *Daily History* of the Weather Underground database. This is an important contribution since the commercial weather services provide only real-time or future weather information via Internet [14]. The second contribution is associated to the photo management tool in order to provide photo organization functionalities, which has been an active subject over the last years [15] [16].

Figure 3(a) presents the blog in the tab *Le Bateau en images* with the annotations and photos registered and exported by a yacht crew member. As we can observe, there is an area with the annotations and photos in the left side and another area containing a map in the bottom right side. This map shows a summary of the yacht trajectory, using the trajectory registered by the mobile phone. The photo is associated with the annotation in the blog and with the map presented in Figure 3(b). It means that each photo jointly with its contextual information will be associated and presented in a placemark on the map. The CAPTAIN blog can be accessed on the website: http://zeroco2sailing.com/blog/.

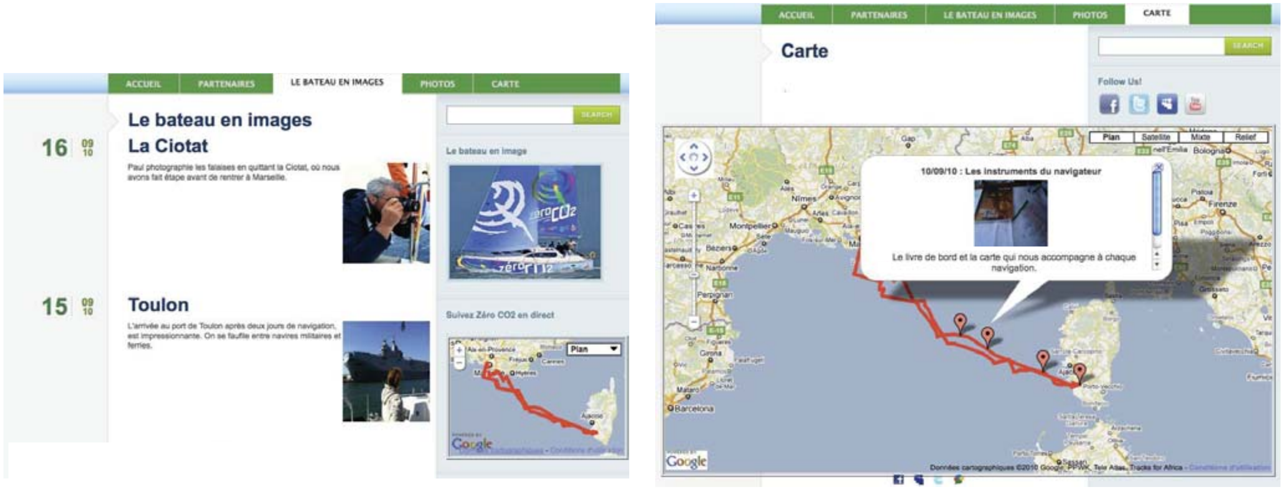
Figure 3(b) presents the map-based interface, which can be accessed in the tab *Carte*. This map is generated using the contextual information captured in the moment that the photo was taken and the user annotations. As described before, the photo and its contextual information are associated with the content that is exhibited by placemarks on the map. Each placemark represents the photo's position obtained. In addition, the user can add other personal information about this moment, such as the passage of animals or other yachts.

Other important results are related to the mobile phone battery consumption during the trajectory registration. In the first test, when the distance filter had been configured to register each movement of the user, the iPhone battery level was down to 10% after 2 hours. After configuration of the distance filter with fifty meters, the iPhone battery level was down to 10% after 3 hours. Another factor that can affect this result is related to the frequency that photos are captured. However, the yacht was equipped with batteries to maintain the iPhone charged during all trajectory registration, as well the laptop to manipulate the desktop application.

Currently, the blog and the map are being utilized to publish the details about the travels. Hence, the on-line users could follow the ZeroCO2 project using the blog and publish the annotations and photos in their social networks profiles, such as facebook and twitter.

V. Conclusions

In this work, we presented the results of a practical experiment of the CAPTAIN. The main motivating concept was to map the trajectory of a yacht in the Mediterranean Sea, adding contextual information related to each user position. The tests were performed in collaboration with ZeroCO2 project, which has the objective to use a yacht powered by hydrogen and renewable energies to demonstrate the efficiency of the combined energy types and analyze the pollution level with data obtained from the air in the sea. The evaluation



(a) Logbook with photos and annotations.

(b) Detailed map.

Figure 3. Logbook and detailed map

results demonstrated that the users were comfortable using the CAPTAIN and found it useful and an excellent tool to accurately publish contextual information according to the geographical position.

As future work, we will focus on researching other types of navigation, such as FOAF-based navigation. Besides that, we plan to study the content combination of different users in order to create reference among their trajectories, personal data and contextual information. Finally, we have the intention to offer our system for personal utilization, for example: to track an excursion in forests and mountains; to study the behavior pattern of a vehicle based on its speed, course, position and contextual information; and to map the course of runners and other athletes.

Acknowledgments

The authors would like to thank the sponsors of the Zero CO2 project and the Ministère de l'Enseignement supérieur et de la Recherche de France.

References

[1] M. Munoz-Organero, G. A. Ramiez-Gonzalez, P. J. Munoz-Merino, and C. D. Kloos, "A collaborative recommender system based on space-time similarities," *IEEE Pervasive Computing*, vol. 9, pp. 81–87, 2010.

[2] Joseph Fourier University, Atomic Energie Commission (CEA), Floraris and RM Fora Marine. (2010, Sep.) Zeroco2 project. [Online]. Available: http://zeroco2sailing.com

[3] K. Chandramouli and E. Izquierdo, "Semantic structuring and retrieval of event chapters in social photo collections," in *MIR '10: Proceedings of the international conference on Multimedia information retrieval*. New York, NY, USA: ACM, 2010, pp. 507–516.

[4] A. Girgensohn, F. Shipman, T. Turner, and L. Wilcox, "Flexible access to photo libraries via time, place, tags, and visual features," in *JCDL '10: Proceedings of the 10th annual joint conference on Digital libraries*. New York, NY, USA: ACM, 2010, pp. 187–196.

[5] Google Inc. (2010, Sep.) The Google Picasa. [Online]. Available: http://picasa.google.com

[6] F. Bentley, C. Metcalf, and G. Harboe, "Personal vs. commercial content: the similarities between consumer use of photos and music," in *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*. New York, NY, USA: ACM, 2006, pp. 667–676.

[7] D. Kirk, A. Sellen, C. Rother, and K. Wood, "Understanding photowork," in *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*. New York, NY, USA: ACM, 2006, pp. 761–770.

[8] W. Viana, J. B. Filho, J. Gensel, M. V. Oliver, and H. Martin, "Photomap - automatic spatiotemporal annotation for mobile photos," in *W2GIS'07: Proceedings of the 7th international conference on Web and wireless geographical information systems*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 187–201.

[9] Masashi Inoue, "Image retrieval: Research and use in the information explosion," *Progress in Informatics*, no. 6, pp. 3–14, March 2009.

[10] T. Kliegr, K. Chandramouli, J. Nemrava, V. Svatek, and E. Izquierdo, "Combining image captions and visual analysis for image concept classification," in *MDM '08: Proceedings of the 9th International Workshop on Multimedia Data Mining*. New York, NY, USA: ACM, 2008, pp. 8–17.

[11] D.-S. Ryu, W.-K. Chung, and H.-G. Cho, "Photoland: a new image layout system using spatio-temporal information in digital photos," in *SAC '10: Proceedings of the 2010 ACM Symposium on Applied Computing*. New York, NY, USA: ACM, 2010, pp. 1884–1891.

[12] N. O'Hare and A. F. Smeaton, "Context-aware person identification in personal photo collections," *Trans. Multi.*, vol. 11, no. 2, pp. 220–228, 2009.

[13] Apple Inc. (2010, Sep.) Instruments for Performance Analysis - Version 2.7. [Online]. Available: http://developer.apple.com/technologies/tools/

[14] Weather Underground Inc. (2010, Sep.) Weather Underground Organization (WUO). [Online]. Available: http://www.wunderground.com/

[15] B. Epshtein, E. Ofek, Y. Wexler, and P. Zhang, "Hierarchical photo organization using geo-relevance," in *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*. New York, NY, USA: ACM, 2007, pp. 1–7.

[16] A. Girgensohn, F. Shipman, L. Wilcox, T. Turner, and M. Cooper, "Mediaglow: organizing photos in a graph-based workspace," in *IUI '09: Proceedings of the 13th international conference on Intelligent user interfaces*. New York, NY, USA: ACM, 2009, pp. 419–424.

Dialogue-driven Search in Surveillance Videos

Vincenzo Deufemia, Massimiliano Giordano, Giuseppe Polese, Genoveffa Tortora
Facoltà di Scienze MM.FF.NN., Università di Salerno
Via Ponte don Melillo, 84084 Fisciano(SA), Italy
{deufemia, mgiordano, gpolese, tortora}@unisa.it

Abstract

Video surveillance is a key tool for enabling security personnel to monitor complex and dangerous environments, and the access to the semantic content of surveillance video has become a challenging research area. In this paper we present a query-processing system for investigating human activities in surveillance videos. The interaction between the user and the system happens through a gesture-based dialogue, which implements a novel question-answering model to support users during the information-seeking process. We also present a visual interface devised to facilitate the query-specification process.

1. Introduction

With the increasing need of security in today's society surveillance systems have become of fundamental importance. Video cameras and monitors pervade buildings, factories, streets, and offices. Thus, video surveillance is a key tool for enabling security personnel to safely monitor complex and dangerous environments. However, even in simple environments, a video surveillance operator may face an enormous information overload. It is nearly impossible to monitor individual objects scattered across multiple views of the environment. It thus becomes vital to develop interfaces making the investigation process on the overwhelming quantity of videos more intuitive and effective.

In the recent years intelligent user interfaces (IUIs) have been investigated for multimedia applications, aiming to improve efficiency, effectiveness, and naturalness of human-machine interaction by representing, reasoning, and acting on models of the user, domain, task, discourse, and media [12]. IUIs have to make the dialogue between the user and the system possible. Real interaction occurs when there is a need to ask for information during a computation. This need actually arises during the computation and cannot be shifted to the starting point of the computation process. This kind of interaction affects the computation and only the interfaces able to realize it can be considered intelligent and able to manage the interaction between system and user.

A problem arising in the application of this view is the need of a powerful language, like it happens among people using natural language for dialoguing. Natural language interfaces are difficult to realize as they yield difficult problems related to natural language processing. The Question-Answering (Q/A) paradigm [17] is a suitable mean to interact with video surveillance systems. Indeed, Q/A implements the investigative dialogue and supports the guided investigation by foreseeing the user actions. On the other hand, it is essential for interfaces to have human-like perception and interaction capabilities that can be utilized for effective human–computer interaction (HCI).

In this paper we present a query-processing system for investigating human activities in surveillance videos. The interface exploits the information computed by the recognition system to support users (security operators) in the investigation process. The interaction dialogue provides a query language based on gesture that enables users to easily specify various kinds of questions about both actions and states, and the nature of responses one wishes. In particular, the user makes use of hand-drawn symbols for describing the situations s/he needs to investigate, whereas the system uses visual symbols to represent the questions to the user. The contribution of this research is twofold: (i) proposing an intuitive interaction mechanism for surveillance video investigation, and (ii) proposing a question-answering model to support users during the information-seeking process.

The paper is organized as follows. Section 2 surveys related work. Section 3 presents the components of the video understanding system. Section 4 introduces the main notions about the question-answering system, while the proposed user interface for investigating human activities is described in Section 5. Finally, Section 6 provides conclusions and outlines future research directions.

2. Related Work

In the domain of video surveillance much attention has been devoted to the problem of using visualization techniques for clustering and anomaly detection [3][15]. Little work has been devoted to the development of interfaces and interaction paradigms to support users in the investigation process.

The set of interface components presented in [7] aims to improve the ability of security personnel to locate and follow important activities within security videos. The components include the recognition and visualization of localized activities in a video feed and provide activity-highlighting video summaries in the form of enhanced keyframes, timelines, and storyboards to give users quick access to interesting events in recorded videos.

In the recent years many video retrieval frameworks for visual surveillance have been proposed [9]. They support various query mechanisms, because queries by keywords have a limited expressive power. In particular, query-by-sketch mechanisms have been adopted to express queries such as "a vehicle moved in this way". A drawing interface allows users to draw motion trajectories which approximately represent the activities that users expect to query.

An approach similar to the one presented in this paper has been developed by Katz et al. [10]. They integrate video and speech analysis to support question-answering about moving objects appearing within surveillance videos. Their prototype system, called Spot, analyzes objects and trajectories from surveillance footage and is able to interpret natural language queries such as "Show me all cars leaving the garage". Spot replies to such a query with a video clip showing only cars exiting the garage.

In the recent years several video retrieval systems have been developed to assist the user in searching and finding video scenes. In particular, interactive video retrieval systems are becoming popular. They try to reduce the effect of the semantic gap, i.e., the difference between the low-level data representation of videos and the higher-level concepts a user associates with videos. An important strategy to improve retrieval results is the query reformulation, whereas strategies to identify relevant results are based on relevance feedback and interaction with the system. The system proposed in [1] combines relevance feedbacks and storyboard interfaces for shot-based video retrieval.

The interaction dialogue proposed in our approach is a generalization of relevance feedback. Indeed, the relevance questions are asked to catch the user's information need, whereas the question-answering process we propose implement a real dialogue between the user and the system making the investigation process more effective.

3. A Video Understanding System Based on Conceptual Dependency

Video understanding aims to automatically recognize activities occurring in a complex environment observed through video cameras [6]. The goals of the proposed activity recognition system are to detect predefined violations observed in the input video streams and to answer specific queries about events that have already occurred in the archived video [4].

We exploit Artificial Intelligence techniques to enable the system "understand" events captured by cameras. In particular, our approach is based on the Schank's theory [14], a "non-logical" approach that has been widely used in natural language processing. Two main reasons led us to use this theory in video surveillance systems. First, the presence of well-studied primitives to represent details about the actions; second, the possibility to use highly structured representations like scripts, which are a natural way to manage prototypical knowledge. Thus, we are able to associate different levels of meanings to a situation: conceptualization, scene, and script level, which allow us to deeply understand the current situations and to detect anomalies at different levels. Moreover, the structured information makes the design and reasoning process easier.

In order to detect anomalies and to raise alert messages, the system tries to interpret a scene based on its knowledge about "normal" situations, using the conceptual dependencies to describe single events and scripts for complex situations. Therefore, the proposed video-surveillance system is an intelligent system associating semantical representations to images.

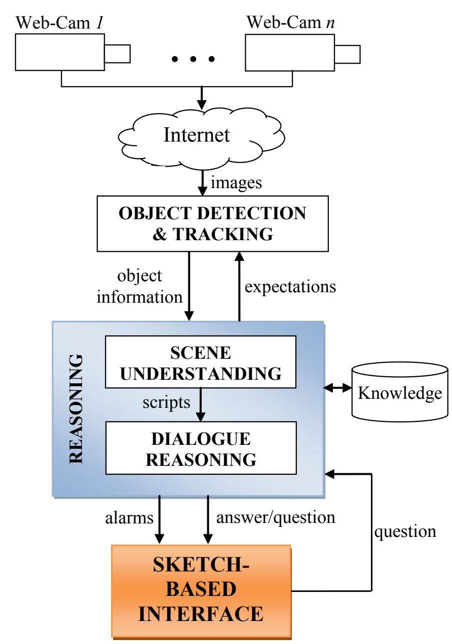


Figure 1. Overview of our system.

Fig. 1 gives an overview of our video understanding system which is composed of three main modules: detection and tracking of multiple objects, scene understanding, and reasoning. The module for tracking multiple objects is implemented by use the codebook based adaptive background subtraction algorithm proposed in

[11]. We are concerned with tracking three kinds of objects - human, vehicles and packages.

The reasoning module has twofold functions: understanding the situations that happen and managing the dialogue with the interface. The first task is accomplished by the scene understanding module, whose aim is of associating a semantic representation to the content of the scenes. This module recognizes events and actions using the knowledge about the standard events and situations stored in its knowledge base. In particular, the information on the tracked objects, i.e., trajectories and features (such as color, size, etc.), are synthesized by constructing conceptualizations, which are given in output to the next module. As an example, the following conceptualization expresses the fact that a given car moves from the garage entry to a car place.

$$car \Leftrightarrow PTRANS \xleftarrow{o} car \xleftarrow{R} \begin{cases} \to CarPlace \\ < Entry \end{cases}$$

The scene understanding module also activates pertinent scripts and appropriate scenes from the script produced by the tracking module in order to identify possible anomalies. In particular, when a script is activated, the conceptualizations belonging to the scenes that might occur are sent to the tracking module to work in a predictive mode. To correctly understand the scene structure, we label various areas of the background, such as doors, elevators, ATM, and so on. The conceptualizations are generated based on object properties and their interactions with these labeled background regions. The output of the understanding module is the scripts describing the occurred situations.

The reasoning tasks of the scene understanding module are:

1. *to understand events*: The task of representing current events using the stored knowledge is accomplished both reducing events to simple ones and instantiating the objects in the conceptualization with actual data;
2. *to reason about events*: Once an event has been interpreted using the existing knowledge it is possible to make inferences and to supply the lack of information in occurred events.

The object detection & tracking module and the scene understanding module exchange information themselves since the first one passes information to be conceptualized to the second one, and, in turn, the latter passes expectations to the first one. Expectations are events or actions which typically follow the last recognized event, making easier the low level recognizing of events.

The task of managing the dialogue with the interface is realized by the dialogue reasoning module, whose main task is to answer questions about occurred events. In order to do so, the module tries to start a real dialogue, constituted by a follow up of questions and answers, with the interface. The function of this module will be treated in deepen in the next section.

The sketch-based user interface allows users to interact with the videos through a language which is natural and intuitive for the user and complete to support a dialogue between the user and the system.

4. Dialogues for Investigation

The classical investigation process in multimedia retrieval is accomplished through the interactive search process shown in Fig. 2. The user repetitively submits a query to the system based on the topic under investigation, the previous queries, and the results obtained so far. The introduction of relevance feedback has allowed to improvement this process. Relevance feedback is the method of reformulation and improvement of the original search request, based on information from the user about the relevance of the retrieved data [13]. However, this method suffers from several limitations such as the constraint to browse the results to give feedback to the system.

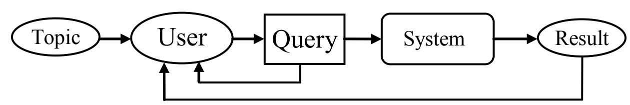


Figure 2. Interactive search: human (re)formulates query, based on topic, query, and/or results.

If we think of both the user and the system as interrogative reasoners, the interface can be interpreted as an oracle for both user and system. In fact, the system does not know who the user is, but it is sure that s/he tells the truth; analogously the user trusts the system, thinking that it tells the truth. Therefore, the interface represents the system for the user and it represents, in turn, the user for the system.

4.1 Metaquestioning

Question-answering systems (henceforth Q/A) have the goal of finding and presenting answers to questions that the user makes. In these systems, the interface has the role of managing a common language between user and system in order to enable the former to make questions and the latter to provide answers. However, Q/A systems do not realize a real dialogue, because the user can only ask questions and the system can only answer. As observed by Driver in [5], it is possible and often desirable that a question be followed by another one, as in the following examples:

q1 What happened yesterday?
q2 Would you like a short or a long response?

Question *q2* in the previous examples are called *metaquestions*. They occur between an inquirer (questioner) asking a first order question and a responder (answerer/metaquestioner) answering through a metaquestion.

The importance of metaquestions in the context of Q/A systems is due to the fact that they can be used to overcome obstacles to answering the first order questions and, hence, they have an active role in the Q/A process itself. In a sense, metaquestions can be seen as a generalization of feedbacks.

4.2 MetaReasoning

Metaquestioning involves many features deriving from the fact that it is related to different research areas, like dialogue theory, problem solving, and metareasoning. From the point of view of dialogue theory [16], metaquestioning can be seen as the general process underlying the *information-seeking* type of dialogue, whose goal is to *exchange information* and where the goals of user and system are to *acquire information* and to *give information*, respectively. According to this view, it turns out that metaquestioning is a version of the analytic method involving two reasoners: the user and the system. This observation tells us that metaquestioning is a *heuristic process* and that, as a consequence, it is characterized by certain but incomplete rules, like, for instance, the problem specialization or generalization.

In our model, we suppose that there are the user *U* (questioner) and the answerer/metaquestioner *A*/*M* such that *U* asks a question q_1 to *A*/*M* and *A*/*M* responds with q_2. Questions are represented according to the logic programming style. As an example, the question "Who entered from the main entrance between 10 and 11 p.m.?" is represented by the clause *act(X, enter, MainEntry, [10-11 p.m.])*, whereas the question "What are the paths of John between 10 and 11 p.m.?" is represented by the clause *act(John, path, X,* [*10-11 p.m.*]*)*. Thus, a question has an unknown (indicated with an *X*) and a body (represented by the action *enter* in the previous examples).

5. Sketch-based Dialogues for Human Activity Investigation

A problem arising in the development of dialogue interfaces is the need of a powerful language, like the natural languages people use for dialoguing. Natural language interfaces are very difficult to realize, since Natural Language Processing has already proven to raise many problems.

We propose the use of hand-drawn gestures as a dialogue language for investigating human activities in surveillance videos. The language is not as versatile as natural languages, but it allows users to query the system in a natural way.

5.1 The Sketch Language for Activity Investigation

A sketch language is formed by a set of sketch sentences over a set of shapes from a domain-specific alphabet. To support the question-answering process the sketch language should allow users specify:

1. the kind of object to be retrieved (the unknown) and the constraints on it, (e.g., a person in an angle of the room);
2. actions and states involving the scene objects, (e.g., a person opening a door, a person waiting the lift);
3. temporal information on the events to be investigated, (e.g., the time interval of a theft);
4. elements of the metaknowledge, (e.g., some properties of the response).

In the following, we describe the symbols composing the dialogues between users and the system.

Language Symbols. The user can associate a sketched symbol to each kind of object identified by the Object Detection algorithm, and will use it to refer to the objects in the context of questions. As an example, if the detection algorithm is able to categorize the mobile objects in people and packages, then during the specification of the questions the user can refer to them by using the sketched symbols in Fig. 3.

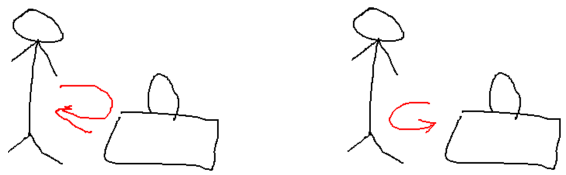
(a) **(b)**

Figure 3. Sketched symbols of the objects detected in the video scene.

In case the objects involved in the question are part of the scene, the user can select them by hooping them with a hand-drawn circle. As an example, Fig. 4 shows a question where a person is already in the scene and it is included in the question by the red ellipse stroke.

Figure 4. A question involving a detected object and a path as the unknown.

Actions and States. Useful information during the investigation process is in the actions involving the detected objects, and the states in which they could be. As for relationships, they depend on the actions that the algorithm used to generate the facts that is able to infer. As an example, the action "a person picks up a package" is described by the sketch in Fig. 5(a), while the sketch in Fig. 5(b) describes the action of leaving a package.



(a) (b)

Figure 5. The sketched symbols representing the actions of a person (a) picking up and (b) leaving a package.

Temporal Information. Questions information regarding time intervals is specified by drawing a circle on the timeline at the bottom of window.

Metaquestions. Sketches are also used by the system to represent questions for the user. We have defined the sketch symbols for a set of metaquestions, such as "how long should the response be?", "which is the path followed by the person?". The example in the next section shows some metaquestion symbols.

Unknown. The unknown of the question is indicated with a question mark on top of the sketched symbol. As an example, the unknown of the question in Fig. 4 is the walk path of the person enclosed with a circle.

5.2 The Sketch-based Interface

We have built a prototype video surveillance system with a sketch-based interface answering interesting questions about video surveillance footage taken in university offices, corridors, and halls. The scenes contain both persons and packages. A typical segment of the video footage shows persons leaving and entering offices, persons discussing in the hall, and persons putting down and picking up packages in the offices and corridors.

Fig. 6 shows the system interface. The main window contains the (background) image of the selected camera, on which the user can draw the sketch representing a question, and the system can reply with another question. As said above, timeline at the bottom of the image is used to specify the temporal information of the question as we show in the following example. The frame at the bottom of the interface (Fig. 6(b)) contains the images obtained from the user question. A storyboard containing the previous investigations is on the right of the interface (see Fig. 6(c)).

The sketches drawn by the user are recognized by using the approach proposed in [2]. The recognition algorithm employs visual language parsing techniques and exploits contextual information to solve ambiguities in the sketches. On average the algorithm gets a recognition rate of 96% since the symbols of the language are not complex and the number of overlapping symbols is very low.

The interface also includes a *Question Visualizer* module, which receives the metaquestions from the Dialogue Reasoning module, and adapts their layout to the user input. In particular, it adapts the shape of the metaquestion symbols to the sketch question drawn by the user.

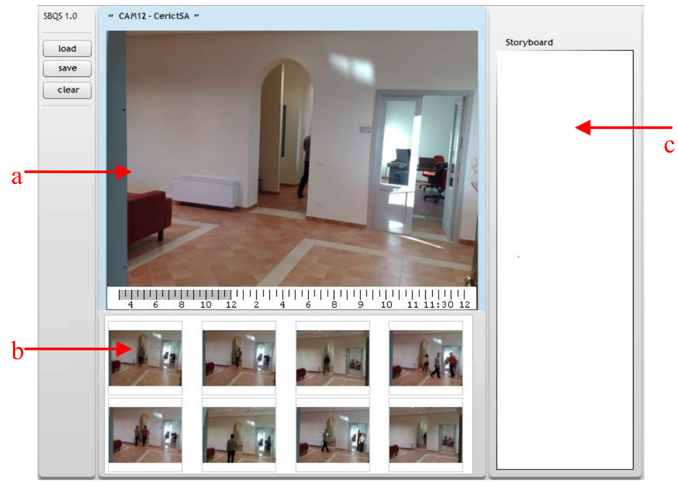


Figure 6. Sketch-based interface showing (a) the (background) image of the selected camera, (b) the images resulting from a previous query, (c) the storyboard of the previous investigations.

Fig. 7 shows an example of dialogue between a user and the proposed system. Notice that user sketches are orange, whereas the metaquestions of the system are red. In Fig. 7(a) the user submit the question "who exited from the corridor between 9 and 11?". Since the result to the question contains a lot of responses, the system question to the user "do you want a long or short response?". This question is represented by the red sketch in Fig. 7(b). The user replies by drawing a circle on the sketched rectangle representing the list of responses (see Fig. 7(c)). To reduce the number of responses the system asks to the user "which path the person followed?" through the red sketch in Fig. 7(d). Notice that the unknown is the path and not the person. The user draws a sketched path indicating that he is interested in the person that entered the room in front of the camera, and the system visualizes two results.

6. Conclusions and Future Work

We have presented a novel interface for investigating human activities in surveillance videos. The interaction with the surveillance system consists of sketch dialogues allowing users to easily specify various kinds of questions about occurred events. The presented proposal contains two innovative solutions: the use of sketching for representing the dialogue between the user and the system, and the use of question-answering to support users in the investigation process.

We have also presented a system prototype that implements several dialogues in a scenario of multi-camera university activity surveillance. In the future we plan to extend the sketch language both for questions and metaquestions. This will allow the system to support dialogues that are more complete and concrete for the user.

We also plan to perform thorough usability experiments. Initial experimental uses of the system have

(a) (b) (c) (d) (e)

Figure 7. A sketch-based dialogue with the proposed system.

already shown a potential enhancement in the user support during the investigation process. However, in some cases the system fails to understand the sketch question due to presence of ambiguities in the positioning of the sketched objects. In particular, this happens because in the same position of the image there are different places where the objects can appear. As an example, when a sketch symbol representing a person is drawn on a space that cover more than one street.

Finally, we intend to extend the interface for supporting the specification of questions involving more than one camera.

7. References

[1] Christel, M. G., and Yan, R. Merging Storyboard Strategies and Automatic Retrieval for Improving Interactive Video Search. In *Proceedings of CIVR 2007*, pp. 486-493.

[2] Costagliola, G., Deufemia, V., Risi, M. A Multi-layer Parsing Strategy for On-line Recognition of Hand-drawn Diagrams. *In Proceedings of VL/HCC'06*, 103-110.

[3] Davidson, I. ,and Ward, M. A Particle Visualization Framework for Clustering and Anomaly Detection. *In Proceedings of KDD Workshop on Visual Data Mining*, Sept. 2001.

[4] Deufemia, V., Giordano, M., Polese, G., Vacca, M. A Conceptual Approach for Active Surveillance of Indoor Environments. In *Proceedings of DMS'07*, 45-50.

[5] Driver, J. L. Metaquestions. *Nous*, *18*, (May 1984), 299-309.

[6] Fusier, F., Valentin, V., Brémond, F., Thonnat, M., Borg, M., Thirde, D., and Ferryman, J. Video Understanding for Complex Activity Recognition. In *Journal of Machine Vision and Applications*, *18*, (2007), 167-188.

[7] Girgensohn, A., Shipman, F., Turner, T., Wilcox, L. Interface Components for Monitoring Security Video. In *Proceedings of UIST'06 Posters*, (Montreux, Switzerland, October 15-18, 2006).

[8] Hintikka, J., Halonen, I., Mutanen, A. Interrogative Logic as a General Theory of Reasoning. *Handbook of Practical Reasoning*, 2002, Kluwer Academic.

[9] Hu, W., Xie, D., Fu, Z., Zeng, W. and Maybank, S. Semantic-Based Surveillance Video Retrieval. *IEEE Trans. on Image Processing*, *16*, 4 (April 2007), 1168-1181.

[10] Katz, B., Lin, J. J., Stauffer, C., Grimson, W. E. L. Answering Questions about Moving Objects in Surveillance Videos. In *Proceedings of AAAI Spring Symposium on New Directions in Question Answering*, 2003, 145-152.

[11] Kim, K. Chalidabhongse, T. H. Harwood, D. and Davis, L. Background Modeling and Subtraction by Codebook Construction. In *Proceedings of IEEE International Conference on Image Processing*, 2004.

[12] Maybury, M. T. and Wahlster, W. Intelligent User Interface: An Introduction. *Readings in Intelligent User Interfaces*, 1998, 1-14, Morgan Kaufmann Press.

[13] Salton, G., Fox, E. A., and Voorhees, E. Advanced Feedback Methods in Information Retrieval. *Journal of the American Society for Information Science*, *36*, 3 (1985), 200-210.

[14] Schank, R. C. and Abelson, R. *Script Plans Goals and Understanding*, 1977, Lawrence Erlbaum Associates.

[15] Seo, J. and Shneiderman, B. Interactively Exploring Hierarchical Clustering Results. *IEEE Computer*, *35,* 7 (2002), 80-86.

[16] Walton, D. The Place of Dialogue Theory in Logic, *Computer Science and Communication Studies Synthese,* Vol. 123, 2000. pp. 327-346.

[17] Wisniewski, A. The Posing of Questions. *Logical Foundations of Erotetic Inferences*. Kluwer, 1995.

Analyzing video produced by a stationary surveillance camera

Paolo Buono
Dipartimento di Informatica
Università degli Studi di Bari "Aldo Moro"
via Orabona, 4, Bari, Italy
buono@di.uniba.it

Abstract — **Today surveillance systems are everywhere. Human observers watching live videos of specific areas are not efficient due to the likely loss of attention. On the other side, unattended surveillance systems require that people analyze hours of recordings when they have to search for some specific events, e.g. identify people responsible of violence, theft or other offences. In many cases a specific search in the video has to be accomplished in the shortest amount of time. This paper presents MotionFinder, a tool that performs video analysis by computing an interactive summarization of the movements in a scene. Once the summarization process is complete, the tool responds in real time to inquires. For example, human investigators may search for specific areas in the video that show high levels of activity or where they know that something occurred (e.g.: property damaged or stolen). The tool responds by showing only the scenes in which some activity occurred for that specific area of the video.**

Video summarization, video analysis, visual analytics, stationary surveillance cameras

1 Introduction

One morning, John enters into his work building and he notices that the director, the responsible of security, the guardian and a technical employee are looking at the guardian's monitor in the guardian's room. After a couple of hours, John passes by the guardian's room again and the four people are still there, so John asks what is going on. It took about four hours of the four people to get to the exact moment in the video in which a thief stole a laptop from an office. The four people moved back and forth in the video and also used the fast-forward feature without noticing any change in the office scene, since only a few frames within a video recording of about 24 hours referred to the very quick action of the thief getting the laptop from the office.

The previous scenario describe a typical situation in the video surveillance context. Today video surveillance stationary cameras have built-in software able to recognize movements in the scene, i.e. to capture scenes in which something moves. Motion detection doesn't always helps, in particular for heavy traffic areas, where almost always something moves.

Having a human guardian watching surveilled areas could be not optimal, particularly when there are many cameras. Human attention span can drop below acceptable levels after only 20 minutes, even in trained observers [4]. If we add that humans can handle a number of items of seven, plus or minus two, the use of observers is very ineffective when more than a tenth of video surveillance cameras are installed. Nevertheless, a surveilled area is a deterrent for people and reduces the number of causalities or damages. Entire cities have networks of surveillance cameras in order to cover specific locations for detecting and identifying potential threats or suspicious events. These systems often adopt real-time algorithms for detecting anomalies, identify objects and track them.

This paper describes a novel technique we have developed for video analysis and a sofware tool, called MotionFinder, which implements this technique. A first proposal of the technique was presented in [2]. MotionFinder allows a human investigator to speed up its search for anomalies by quickly selecting excerpts of the video in which an event occurred. The tool is intended for post-processing activities, not for providing real-time alerts. Nevertheless, it is possible to work in real-time, since the adopted summarization technique is very fast.

Next section provides an overview of related work. In Section 3, the summarization technique is presented. MotionFinder is illustrated in Section 4. Section 5 describes an interaction session in order to show how the tool works. Finally, conclusions and possible future research directions are reported.

2 Approaches in video surveillance

This section reports the most relevant work about the current video analysis systems. Most of the deveoped systems try to solve the problem of video analysis by creating fully automated applications. In video surveillance, events and requirements are often unpredictable, making useless the software used for the analysis. In effective video analysis it is important to combine the perception, flexibility, creativity and general knowledge of the human mind with the enormous storage capacity and computational power of computers [8]. Because analysing video is time-consuming, researchers try to to support the analysts by allowing them to focus their efforts on relevant parts of the video, avoiding wasting precious time on irrelevant segments.

A significant research has been done in building fully automated systems, which should leverage human investigators in analyzing the videos. A survey on the sate of the art about automated visual surveillance technologies is [13]. The survey focuses on intelligent surveillance systems, which use techniques and methods for recognizing objects and humans with the aim to describe their actions and their interactions. Techniques and tools are presented focusing on aspects like object recognition, behavioural analysis, and surveillance systems architectures. Instead of focusing on automated techniques, our goal is to provide human investigator with tools that empower their abilities.

In [5], beside a discussion about the state of the art for object recognition, particular attention is given to the architecture and the user interface, which should allow users to easily operate with the system. The authors build a "surveillance index browser", which is part of a larger architecture that includes a face-recognizer module. In the user interface, a timeline represents the overview of all the events detected by the system in a particular time-frame. A second timeline provides a zoomed version of any video segment chosen by the user. A window displays the output of the tracker camera, while a second window displays a zoomed-in video about the moving object. Moving objects can be shown as color coded traces; according to time, older events are drawn in white, which gradually turn in red as time passes. In order to limit screen cluttering, the user can also apply filters. Also our tool uses a timeline, a window for viewing the video, and the color coded traces.

Performing queries in video databases is another requirement of video analysis. The system of Huston et al allows the human investigator to perform queries on unstructured data in a large number of distributed camera located in a city area [6]. The goal is to take advantage of the computer resources for searching video data, allowing the user to focus the investigator attention on interpreting the results in the hope of gaining insights that might help to accomplish the task the user is undertaking. A query is triggered by selecting a portion of an image; the system starts the search using a brute-force algorithm, which is the only viable solution, because the user is free to select any region of any frame of the video, and generally it is not expected to know what to search for, making indexing useless. Moreover, the authors point out that the rate at which new data is generated far exceeds the rate at which it can be analysed, and most of it will never be searched before being erased; as a consequence, pre-processing can be seen as a waste of resources. The human investigator can select different search parameters, like color, shape, texture and object detection. In order to avoid long waiting, the system shows results as soon as they are found and allows users to perform concurrent searches. In the tool proposed in this paper, the analysis performed by the user is not known in advance, but indexing is performed because: 1) the produced image can be used as a visual search for videos, since it is considered as a preview of the video; 2) indexing process is faster than the production of the video itself; 3) index files are very small.

Some thoughts about video occupancy space are provided by Romero et al. [10]. They address a system that tracks over 7500 hours per month, which occupy about 7500 GB space (240 GB per day). This means that it is unfeasible to use a typical laptop for the analysis. In order to be able to keep in a laptop one month of recorded videos, they exclude all frames in which no activity is detected; in this way, the needed space for one month video recordings drops to 120 GB. As will be presented in Section 3, the technique proposed in this paper allows to show to the user a set of previews. A single preview is an image representative of a whole video, so the compression is very high.

Real settings today are composed by many cameras. An interesting work that allows to monitor 20+ areas is DOTS [3], where the experience of a one-year installation of an indoor multi-camera surveillance system for use in an office setting is described. The user interface displays thumbnails of the cameras set in a building. The system tracks people movements and shows people positions in a 2D or a 3D model of the building. People are recognised through face-detection software. When the investigator is interested to one camera or one person, a preview area shows images coming from the selected camera. The system works by providing 2D and 3D models of camera settings, in this way the system can track people across multiple cameras. In the proposed version, MotionFinder addresses only one video at a time. DOTS could be inspiration for the next version of the tool.

A frequent requirement is to count people entering and exiting from a place. This problem has been faceted for various scenarios, like people going in and out a building, or a bus or a train [1]. Sidla et al. adopt a vision-based pedestrian detection and tracking system, able to count people in very crowded situations, like escalator entrances in underground stations [11]. Our tool provides a simple counter that counts the different time intervals in which the movements have been occurred.

Current technologies allow people to track human gestures, and eye gaze, to help in human-computer interaction. Vural et al. uses eye-gaze analysis to capture overlooked actions in order to prepare a summarized video for a subsequent analysis [14]. This is particularly useful if there are many operators; the supervisor can quickly check what the operators observed.

Video summarization is an approach adopted in several systems. According to [9], each object moves in a “tube”. By removing the constraint of the time in which an event occurred, it is possible to visualize several tubes in the same moment. The user perceives that many objects are doing something in the scene, this approach saves the human investigator time because, instead of watching moving things sequentially, the analyst can watch them in parallel. This approach is useful for videos of almost desert locations, but does not scales for crowded scenes.

A simple and effective technique is proposed by Tang et al. where the user draws small segments on the screen, the system shows the pixels beneath the segments in a timeline [12]. The idea behind this technique comes from how the scanner works. Only the parts of the video that “pass” through one or more segments drawn by the user are “scanned”. For example, if such a line is drawn across a road, the resulting timeline will show the shapes of the car passing along it. These shapes will be more or less elongated depending on the speed. If nothing is happening, the line will keep writing the same pixels. So any object that passes over that line will be easily identifiable because background pixels, will be uniform and foreground objects will stand out. Our approach uses a similar idea but, instead of drawing a line, the user selects an area, and the tool shows videos, instead of still images.

3 The summarization technique

The approach presented in this paper relies on the creation of an image that summarizes the activity recorded by a stationary camera. The goal is to display traces of movements across the scene. The tool is very useful when it is known what happened, but the precise details about how and when are unknown (i.e. a laptop is stolen but it is not known when and by who). Traces of movements that took place in a given scene are shown using a color scale which ranges from yellow to red: yellow being used for previous activities, red used for more recent activities.

The video analysis process starts with the creation of an N x M matrix, called sequences matrix. N and M are the rows and columns of the original video frames respectively, so the matrix has the dimensions of the video resolution, expressed in pixels. Each cell of the matrix represent a corresponding pixel position in the video and contains a counter and a list of integers representing the frame number in which a movement was registered in the corresponding pixel; this list is called sequence. The counter is an integer indicating the number of frames in the sequence, see Figure 1; it represents the amount of activity for the associated pixel. The information about the amount of activity in a given pixel can be used t highlight “most active” pixels or filter out areas with a low activity or vice versa. Another use of the counter is that to quickly jump to the last recorded frame in order to get “the age” of the last movement in that pixel.

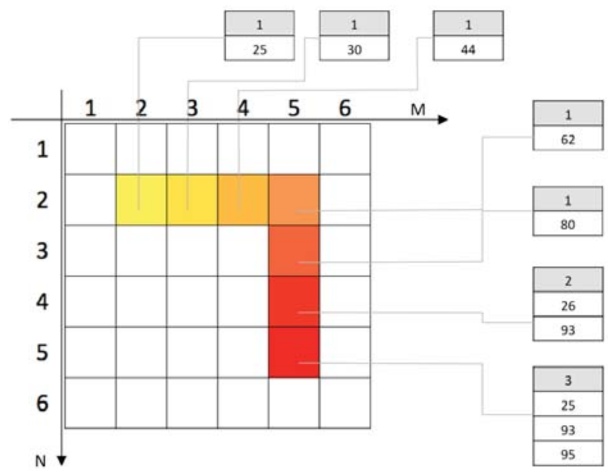


Figure 1. Simple example of generating the heat map from the sequence matrix

The proposed technique creates a summarization image whose goal is to show the activities occurred in the pixels of a video scene. The image is generated fom the matrix by assigning to each pixel a color using the yellow-red scale: a pixel showing a yellow color means that the last activity in that pixel occurred further back in time, while a pixel with a red color means that the last activity occurred recently. The resulting image shows all the movements that occurred in that certain scene together with a rough indication of how recent are the movements. If no movement is detected in the whole video in a specific pixel, the corresponding sequence will be empty and the pixel color in the visualized heat map will be transparent.

In order to better explain how the heat map is generated from the sequence matrix, Figure 1 shows a very simple example. In this exampe, the video has a resolution of 6 x 6 pixels, so the matrix will have 6 x 6 cells. Only in seven of the 36 cells something has moved, and the corresponding

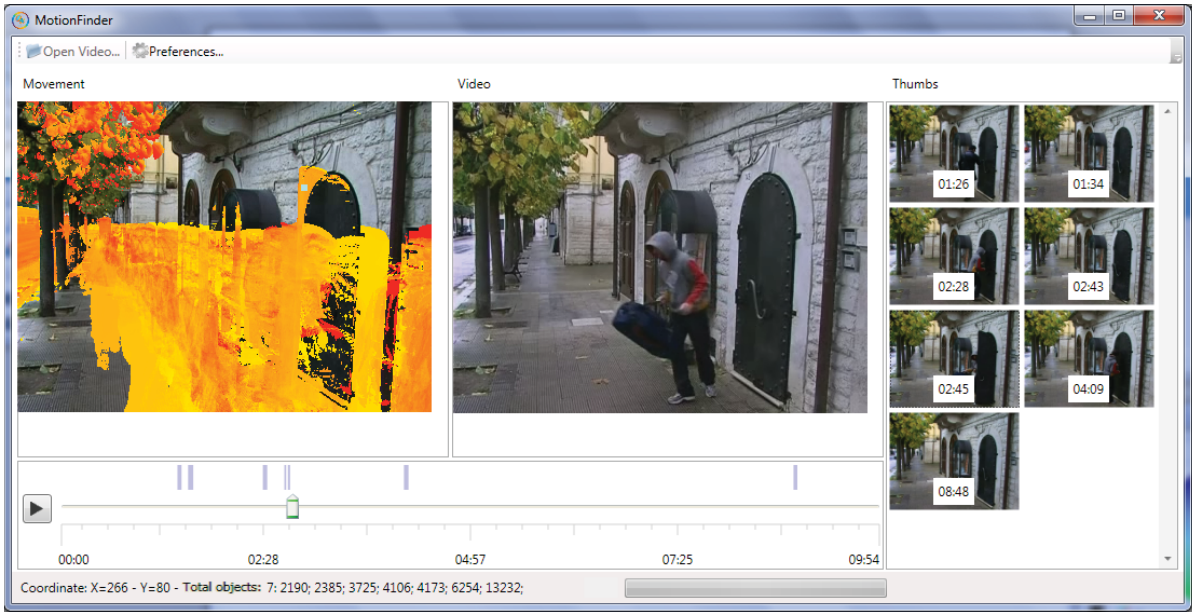


Figure 2. A screenshot taken from the analysis of the camera theft scenario

sequences are displayed. In this example, movements have been detected starting from the frame 25 until the frame 95. The cell (2,2) shows that the event occurred very early (in frame 25), while in cells (4,5) and (5,5) last activity occurred late in the video (in frames 93 and 95).

The video is analysed using the frame difference algorithm, which detects the movement of objects across two (single difference) or three frames (double difference [7]). Since the video sequences can be very long, a logarithmic scale can be applied in order to make visible movement that take place in short intervals of time.

4 MotionFinder

This section describes MotionFinder, a tool which performs video analysis by implementing the summarization technique we have illustrated in the previous section. MotionFinder is implemented in C#, and includes the OpenCV library for the image processing and object tracking. Since OpenCV is created in C/C++, the wrapping with C# is performed using the EmuCV platform.

The user interface of MotionFinder is shown in Figure 2. The image on the left is called *Movement* and shows the heat map produced from the sequence matrix as described in the previous section. The heat map represents, in a scale from yellow to red, traces of movements that took place in the scene; it is displayed overlayed on a frame in which no movement occurred. Yellow and red have been chosen for the heat map because, in outdoor environments, yellow and red are the least likely colors to be found; moreover, most camera equipments record video in grey scale. The heat map in Figure 2 that uses these two colors is well visible. Furthermore the yellow-red gradient carries some implicit meanings, like in thermography imaging, in which yellow is used for colder temperatures while red is used for hot temperatures. We consider an analogy to a heating process, in which hot temperature are reached during time; similarly, the map representing scene movements shows in yellow the previous movements and in red the most recent ones. The user may change the standard colors if desired or may desaturate the background color. In order to make visible the portion of the image occluded by the heat map, the user can make the map semi-transparent, as shown in Figure 3, or hide it temporarily.

The *Movement* area is interactive and the user can click in any part of it. By clicking on areas that do not contains any movement, nothing happens. If the user clicks on pixels with some movement, in the right area, called "Thumbs", all the video excerpts corresponding to selected pixels appear. The Sequence matrix returns an ordered list of all frames related to the selection; MotionFinder splits this list into different frame sequences in order to build a number of excerpts representing different events.

In Figure 2, the information that seven parts of the video have been detected is provided in several ways: seven thumbnails appear in the "Thumb" area, at the right of the window; it is explicitly written in the status bar, at the bottom of the window; seven markers are displayed in the timeline, at the middle of the window.

Over each thumbnail is indicated the starting time of that specific video segment. It is possible to configure MotionFinder to read metadata about the video and show real date and time in which events occurred, but often this information is not explicitly provided; it is often coded in the video file name, and the creation date and time is not always reliable. In the case of Figure 2 metadata were not provided.

Below the "Movement" and "Video" areas, there is a timeline, which displays all detected excerpts as markers. If the user clicks on the "arrow" button positioned on the left of the timeline, all videos are played sequentially in the "Video" area. The user can otherwise start a single excerpt by clicking on the desired thumbnail in the "Thumbs" area; in that case only the clicked video is played.

The user may click on several points of the "Movement" area; points can be added or removed according to the user's needs. The sequence matrix, that holds information about where and when something happened allows the tool to respond in real-time. There are some hints that can be followed in the interaction with the heat map, in order to further speed up the analysis. One of the cases in which MotionFinder can be profitably used is when, before and after the event, the scene does not change, for example in the case of a door, or a window, that is opened and closed again; this action usually last for a few seconds, so other techniques, like jumping back and forth or fast-forwarding, are not effective. By clicking on the traces of the door that opens, like the click shown in Figure 2, the frames retrieved are those that belongs to the action, plus a number of frames before and after the movement. By default MotionFinder adds one second before and one second after, but this parameter can be changed.

5 An interaction example

In the context of video surveillance, someone may have stolen a valuable object or vandalized a store, etc. The human investigator who has to analyze the video has no clue about the specific time and how the event happened. One or more stationary cameras recorded the event. In order to speed up the search, motion detection sensors can be installed into the cameras, and may provide some help, this is very effective in mostly desert areas, but the case of a streets, corridors, with several side access or in frequently accessed areas, motion detection sensors are not really useful. The human investigator should review hours of video recordings in the hope of identifying the exact moment the event occurred.

We now describe how the tool shown in Figure 2 is used in order to support the human investigator to analyze the video and identify the frames related to that specific event. The video reported in the example of Figure 2 is recorded from a surveillance camera observing a street in which a photographer has his photography studio. The video have been recorded while the photographer was working in the studio.

Figure 3. The camera theft scene with the heat map overlay set to semi-transparent

During the day, he leaves the studio and close the door without locking it. When the photographer returns, he readily notices that an expensive video camera has been stolen. He does not know who, how and when exactly the theft took place but he assumes that something must have happened while he was away. He retrieves the video recording from the surveillance system and gives it as input to MotionFinder. The system produces the summarization shown in Figure 2.

Since the street is fairly trafficked, there is a lot of activity shown in the output. The photographer supposes that the thief must have entered the studio by the door. Thus, he clicks on traces of the upper corner of the door. The system identifies several moments in which the door was opened or closed. The timeline in the bottom part of Figure 2 shows seven markers placed at the corresponding times (also indicated in the status bar at the bottom of the window). In the right part of the screen a thumbnail appears for each detected excerpt. The photographer proceeds looking to the thumbnails and watching the video associated to those who seems suspicious.

He quickly identifies the moment in which he left the store, according to the thumbnails, at the time 2.28 the theft enters in the photography studio and exits at the time 2.45. The first thumbnail shows a person entering the studio and the next one after shows the previous person leaving while holding a bag with the stolen camera. Having watched the video excerpts and confirmed his suspicions, he exports the video excerpts in which the thief is shown entering and then leaving so that he may deliver them to the authorities. Hopefully the police will be able to identify the thief.

6 Conclusions and future work

In this paper, MotionFinder, a tool for analysing video recorded by stationary surveillance cameras is presented. The tool supports the human investigator in the search for a particular event of which it is not known exactly when and how it occurred. MotionFinder summarizes the video into an interactive single image, composed by a background and a heat map showing traces produced by the movement of objects and people in the surveilled area. By interacting with this image the human investigator can see a set of excerpts related to the selection s/he made and, in seconds, get to the desired results.

One limitation of the adopted technique is about crowded scenes. The approach adopted by Pritch et al. [9] could be used to reduce the problem, but the heat map cluttering problem remains. One of the future directions could be to consider activities during time as different layers, like in the proposal of Romero et al. [10]. By providing the user with an interface for choosing the layers of interest, instead of querying the whole video the analysis may be speeded up.

During some informal evaluation sessions, some users suggested that it might be useful to incorporate the predominant direction of the movements.

7 Acknowledgments

Partial support for this research is provided by the VisMaster Coordination Action in the Future and Emerging Technologies (FET) programme under FET-Open grant number 225429.

References

[1] A. Albiol, I. Mora, and V. Naranjo. Real-time high density people counter using morphological tools. *Intelligent Transportation Systems, IEEE Transactions on*, 2(4):204 –218, Dec. 2001.

[2] P. Buono and A. L. Simeone. Video abstraction and detection of anomalies by tracking movements. In *International Conference on Advanced Visual Interfaces*, AVI '10, pages 249–252, New York, NY, USA, 2010. ACM.

[3] A. Girgensohn, D. Kimber, J. Vaughan, T. Yang, F. Shipman, T. Turner, E. Rieffel, L. Wilcox, F. Chen, and T. Dunnigan. Dots: support for effective video surveillance. In *15th international conference on Multimedia*, MULTIMEDIA '07, pages 423–432, New York, NY, USA, 2007. ACM.

[4] M. W. Green. The appropriate and effective use of security technologies in u.s. schools. a guide for schools and law enforcement agencies. Technical report, Sandia National Labs., Albuquerque, NM., 1999.

[5] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, and S. Pankanti. Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking. *Signal Processing Magazine, IEEE*, 22(2):38–51, March 2005.

[6] L. Huston, R. Sukthankar, J. Campbell, and P. Pillai. Forensic video reconstruction. In *ACM 2nd international workshop on Video surveillance & sensor networks*, VSSN '04, pages 20–28, New York, NY, USA, 2004. ACM.

[7] Y. Kameda and M. Minoh. A human motion estimation method using 3-successive video frames. In *VSMM'96, International Conference on Virtual Systems and Multimedia*, pages 135–140, 1996.

[8] D. A. Keim. Visual exploration of large data sets. *Communication of ACM*, 44:38–44, August 2001.

[9] Y. Pritch, A. Rav-Acha, A. Gutman, and S. Peleg. Webcam synopsis: Peeking around the world. In *11th IEEE International Conference on Computer Vision, 2007*, ICCV '07, pages 1–8, October 2007.

[10] M. Romero, J. Summet, J. Stasko, and G. Abowd. Viz-a-vis: Toward visualizing video through computer vision. *IEEE Transaction on Visualization and Computer Graphics*, 14(6):1261–1268, November/December 2008.

[11] O. Sidla, Y. Lypetskyy, N. Brandle, and S. Seer. Pedestrian detection and tracking for counting applications in crowded situations. In *Video and Signal Based Surveillance, 2006. AVSS '06. IEEE International Conference on*, pages 70 –70, 11 2006.

[12] A. Tang, S. Greenberg, and S. Fels. Exploring video streams using slit-tear visualizations. In *International Conference on Advanced Visual Interfaces*, AVI '08, pages 191–198, New York, NY, USA, 2008. ACM.

[13] M. Valera and S. Velastin. Intelligent distributed surveillance systems: a review. *Vision, Image and Signal Processing, IEEE Proceedings*, 152(2):192–204, April 2005.

[14] U. Vural and Y. Akgul. Eye-gaze based real-time surveillance video synopsis. *Pattern Recognition Letters*, 30(12):1151 – 1159, 2009. Image/video-based Pattern Analysis and HCI Applications.

Unified Communications Deployment Tool

Shadan Saniepour Esfahani
Cisco Systems
Advanced Services
Unified Communications Practice
ssaniepo@cisco.com

Talal Siddiqui
Cisco Systems
Advanced Services
Unified Communications Practice
tsiddiqu@cisco.com

***Abstract*- This article elaborates on the architecture of a web based tool for large scale Unified Communications Deployment. Unified Communications Deployment Tool (UCDT) is geared towards reducing implementation cost and time by automating and standardizing the process. The tool allows enforcing design and deployment best practices by standardizing and automating the process that prevents the human errors. UCDT provides one platform to provision all the components of Unified Communications (UC) such as call agent/soft switch, phones, voice mail, presence, and other collaboration platforms in the network, and hence eliminates the need for adopting multiple tools for the UC deployment. UCDT is accessible over the network that allows performing the deployment activities from remote locations and save travel costs.**

I. INTRODUCTION

Unified Communications (UC) is gaining more popularity by small and large businesses, and more enterprises are adopting UC as their main mean of collaboration. While adopting UC has enormous advantages [4][5][6], deploying UC involves some challenges, namely, implementation cost and time. Unified Communication networks are becoming complex as they include soft switches, gateways, intelligent switches to facilitate media transport with QoS settings, network voicemail, integration with multiple vendor products, and other advanced features such as presence, web sharing. It is not simply enabling a dial tone to a desk phone. In a large enterprise there will be different line of business, each with its own set of requirements. This increases the complexity of planning and designing. This also leads to ensuring that a complex architecture with multiple devices and complex feature set gets deployed accurately. Unified Communications Deployment Tool (UCDT) is geared towards addressing these issues and reducing cost and time by automating and standardizing the deployment process.

UCDT is a web based application that allows users to access it remotely over IP network with proper credentials. Web applications are widely being adopted throughout the workplace due to their flexible nature [1][2][3].The accessibility of UCDT to the target devices eliminates the need for users to be at the location where the deployment is taking place.

UCDT provides a single platform for provisioning multiple UC products. Network administrators are often concerned with the total cost of ownership of communication network and a unified tool to manage a wide range of devices and applications is highly desirable.

UCDT allows writing re-usable templates for provisioning requests, which can enforce implementation best practices. Once the templates are created and verified for one site, they can be used to provision different sites and no further verification is required. This eliminates verification time for every site, and minimizes the deployment time by utilizing re-useable templates. Furthermore, since the templates embed the intelligence, once they are finalized and put under version control, the deployment tasks can be performed by junior staff with limited technical knowledge.

Some of main advantages of the UCDT can be listed as:

1) Increases productivity by utilizing intelligent, re-useable templates.
2) Reduces implementation costs and time by means of automating the process
3) Avoids configuration mistakes by means of automating the process.
4) Reduces configuration verification time, by utilizing re-useable templates
5) Brings consistency to the survey and provisioning process, by embedding intelligence in the re-useable templates
6) Reduces admin knowledge requirement by embedding intelligence in the templates.
7) Provides a centralized database for surveyed data
8) Enforces best practices rules via standardized intelligent templates
9) Provides a single platform to provision different UC products
10) Provides a standardized deployment. A standardized deployment may address all the entitlement concerns in a large enterprise or for a service provider managing multiple clients.
11) Is a web based application and therefore eliminates the requirement of being at the location where deployment is taking place

In this article we introduce the architecture of the Unified Communications Audit Tool (UCDT), designed to facilitate the deployment of enterprise's Unified Communication applications including Cisco Unified Communications Manager, Cisco Unity Connection, Unity, Cisco Presence Server, Voice gateways, and the

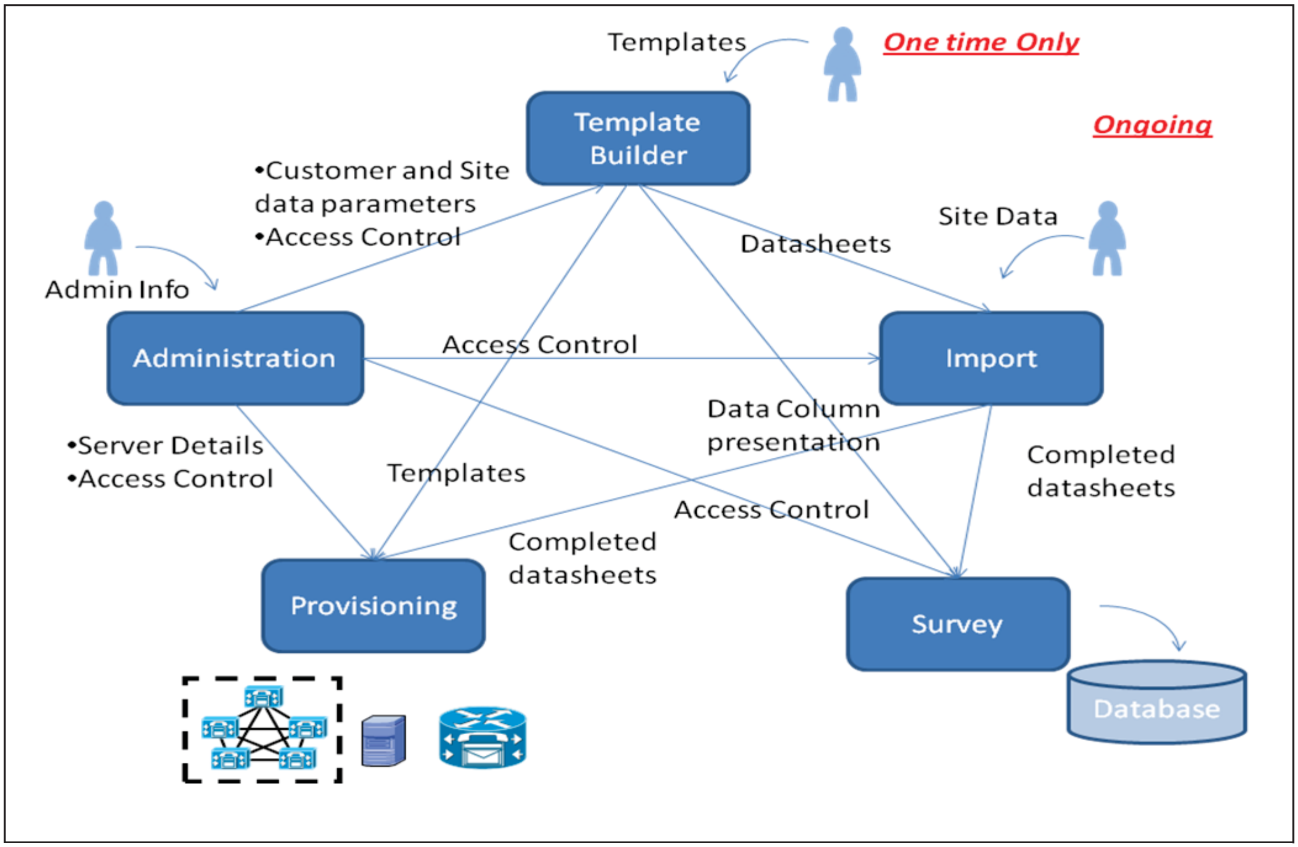


Figure 1 Unified Communications Deployment Tool (UCDT) Architecture

underlying infrastructure. We elaborate on the internals and operations, and study this architecture in practice.

II. PROBLEM STATEMENT

There are two main challenges for deploying UC in large enterprises:

1) Configuration should be done for every site according to the network design
2) Multiple device types (such as gateways, network voicemails, presence etc), possibly from from different vendors, should be configured

Automation is very desirable to facilitate the first point. Consider a financial enterprise with thousands of branches around the world. All the devices (gateways, switches, servers) should be configured based on the same design, and only few variables vary from site to site (such as usernames, IP Phone MAC IDs, directory numbers, etc). Therefore automating this task would save productivity. Customized tools can always be provided to automate the provisioning of a single device, but that means to satisfy the second point mentioned above users have to deploy multiple tools, just to perform the task of deploying UC. Therefore, the problem is how to have a single platform that can provision any device. UCDT provides a unique, one of the kind architecture that allows automation of the provisioning task for multiple device types, even from different vendors. It is a single platform that can configure ***any*** device. This reduces the cost of ownership significantly. The flexible architecture of this tool is introduced in this paper to demonstrate how it overcomes the challenge of automating deployment of multiple device types by a single tool.

III. SYSTEM ARCHITECTURE

The main components of the system are defined as: Administrator agent, Template Builder agent, Site Survey agent, Provisioning agent, Data Import agent, Remediation agent. Fig. 1 illustrates this architecture.

The Administration agent controls the access of the users to the tool based on their roles and permissions. It also allows defining parameters that can be used enterprise wide (only for a specific customer) or system wide (for all customers). System maintenance such as back up, applying software upgrades and finally uploading license is also done by this agent.

The “Template Builder” agent, allows users create templates, which are the framework of the provisioning requests. Templates are in fact provisioning request that have variables instead of the site specific data. Once user feeds the site specific data to the tool, these variables would be replaced by the data and a provisioning request would be formed. In other words, templates plus site data would provide provisioning requests. UCDT uses “Velocity” engine for the template writing language. Users can utilize all the features available by “Velocity” such as string manipulation or any kind or mathematical functions to write their templates. This would allow writing complex templates and embedding all the intelligence inside the template. This

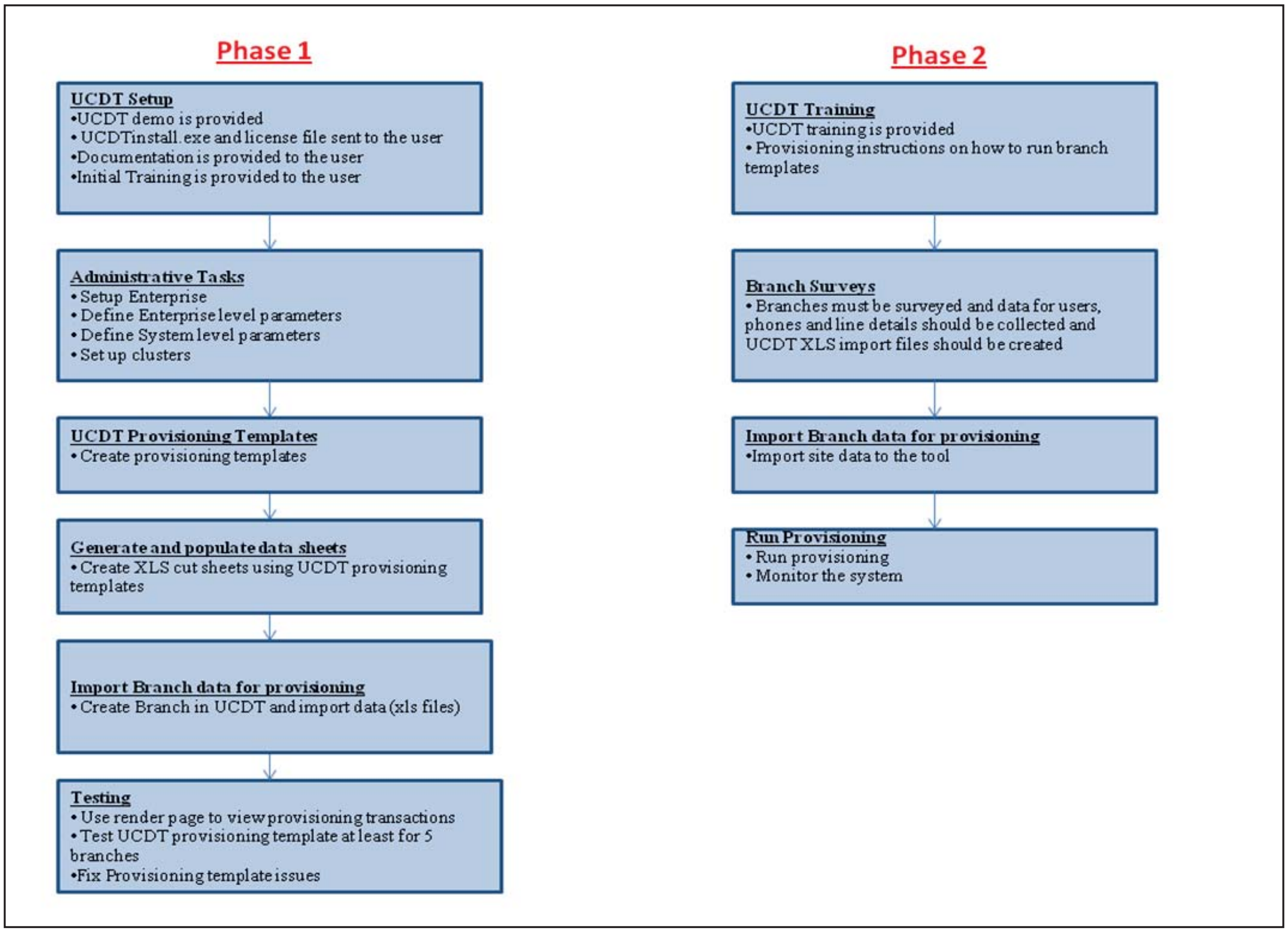


Figure 2 Process for Deploying UCDT

would save users from the need to massage the data prior to uploading it to the tool. The templates are usually written by senior engineers and contain the fundamentals of the UC network design.

Once templates are created, tool would look into the templates and would find the variables used inside the templates and would generate datasheets which contain columns as variable names that need to be provided to the system. These variables are typically the name and details of the enterprise users, or MAC ID of the phones, or any other information that is required to be fed to the UC components for provisioning. The datasheets are filled out by the user and fed to the system through the "Import" agent. Since all the intelligent has gone to the templates, the task of populating datasheets (which is ongoing tasks for all sites) can be left to junior engineers or people with limited knowledge about the design of the network.

The "Survey" agent allows reviewing the uploaded site data. It provides a centralized database that can store all the site data. This is a good mean for storing the site data, for all UC products in a centralized database prior to doing provisioning.

Once the site data is provided to the tool it is used to replace the variables in the templates and form provisioning requests. Provisioning agent allows user review completed requests before sending them off to the UC server for provisioning. Users can then launch a provisioning task immediately or schedule a time in the future to send the provisioning requests to the UC components. Once the provisioning task is completed, the results (success or failure) and details (reason for failure) are reported back to the user.

UCDT currently supports provisioning of Cisco Unified Communications Manager, Cisco Presence Server, Unity, Cisco Unity Connection, Voice Gateways and Switches. However the tool provides ability to write plug-ins to provision any third party device as well.

The flexible template builder agent design, which I based on Velocity and allows generating datasheets according to the templates, plus the ability to provide plugins to provision new devices are two essential elements of UCDT that have made possible to do deployment of any device from a single platform.

IV. SYSTEM OPERATIONS

As Fig. 2 illustrates, there are two main phases for using UCDT, namely: the initial phase, when the templates are created, and the ongoing provisioning phase.

During the initial phase (Phase 1), users go through the UC network Low Level Design (LLD) and create the templates. The templates contain the skeleton of the provisioning requests. Any value that is consistent among sites/branches should be hardcoded and any value that might differ from site to site should be defined as a variable. The values for the variables would be provided

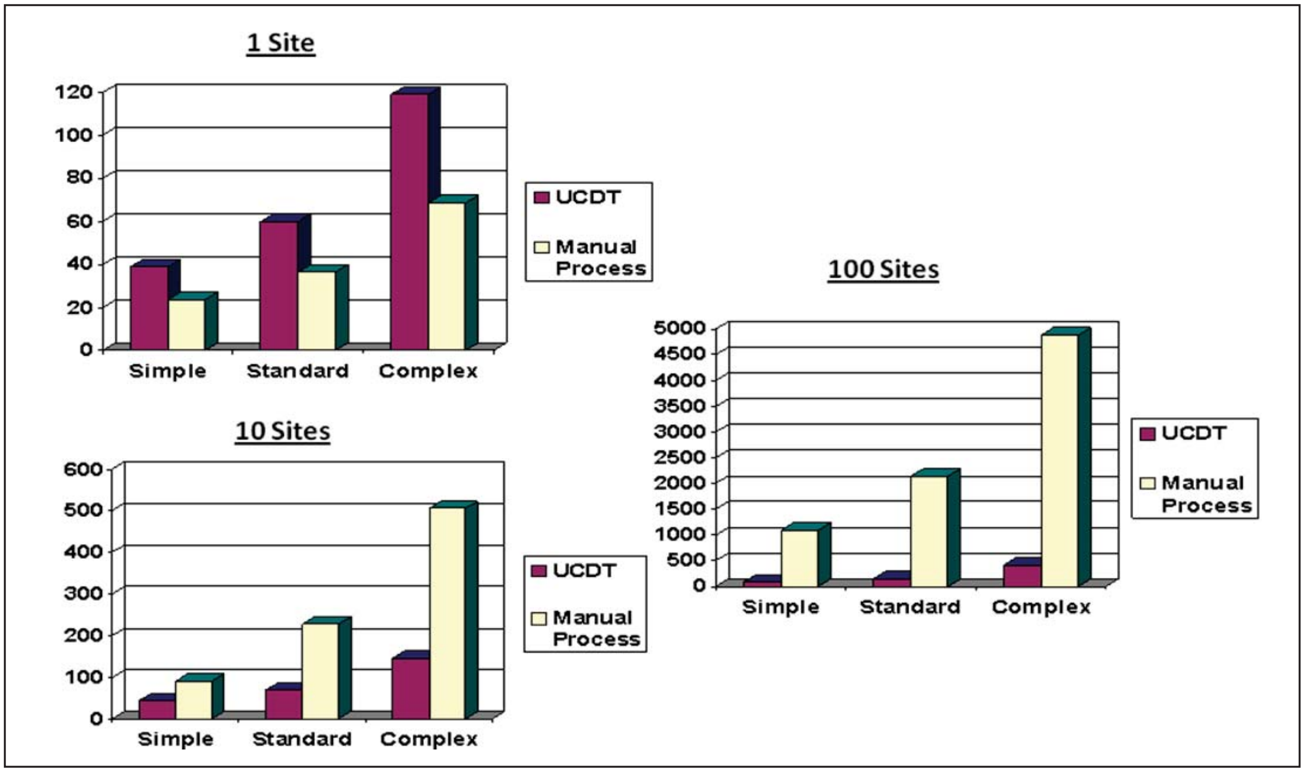


Figure 3 UCDT vs. Manual Process

to the system during site provisioning, when the datasheets are populated. Populating datasheets is a task prone to error and time consuming, so as the best practice, users are encouraged to hardcode as much information as possible inside the template. Also users are recommended to enforce best practices and avoid providing too much flexibility in the design. For example, it is recommended to assign product type based on job roles, to for example provide Cisco IP Phone 7961 to all regular employees and Cisco IP Phone 7975 to all site managers. This way we can define a rule in the template to specify the phone model based on job title, as opposed to have any kind of phone for any employee. This not only would simplify the provisioning and deployment process but also would greatly simplify the maintenance of the network.

Since templates are based on the LLD, if anything in the LLD should change, it would require template update, to reflect the same change. Therefore, it is recommended creating templates once the LLD is locked done. Experience has shown that changes in LLD are inevitable, but with this policy at least changes to the templates can be minimized.

Once the templates are created, they should be tested on a lab cluster. So datasheets are generated and then populated for the lab cluster. The provisioning task is then scheduled and launched, and the results are verified. Templates are revisited to fix any potential error or to optimize. Once templates are perfected, we can go to the next phase which is the ongoing provisioning phase.

In the beginning of the second phase (Phase 2), training is conducted to educate the people who would be responsible for populating datasheets and scheduling provisioning tasks on how to use the tool and the templates.

Upon the completion of the training, staff would be able to independently utilize the tool. They should gather the site survey data and populate the datasheets, then import to the tool and schedule and launch provisioning task.

Populating the datasheets is one of the most time consuming tasks in the whole provisioning process, as it is manual and needs human interference to make sure correct data is provided to the tool for provisioning. This is one of the factors that add delay to the provisioning process. In addition, there is a throttle speed that would be adjusted on the UCDT side, to control the speed of sending provisioning requests. This is to prevent the UC components from becoming slow under high volume of requests per second.

In order to increase the speed of the process with regard to these limitations it is suggested to populate datasheets for next sites/branches, while running provisioning tasks on previous sites/branches. Experience has shown that provisioning up to 50 sites/branches per week is practical.

System should be monitored during phase 2, but since the templates have already been verified, thorough verification is not required in this phase.

V. UNIFIED COMMUNICATION DEPLOYMENT TOOL INTERNALS

In this section we would visit three topics: security of the tool, currently supported platforms and the provisioning mechanism for each of them, and finally the template builder engine.

A. UCDT Security

UCDT is a web-based application and hence eliminates the need to be locally present at the site to conduct the provisioning. To ensure the security of the tool, it utilizes https protocol and therefore all the communication between browser and the tool is encrypted. UCDT also allows defining granular access rights to each agent of the tool for different users. The administrator of the system can grant access or deny each user from accessing specific modules. Typically all users have access to “Import” and “Provisioning” modules, whereas only a few users have access to “Template Builder” module. The reason is, templates are usually written by the senior engineers who have designed the network and other users are usually not supposed to modify or delete templates once they have been verified.

Furthermore, users can be denied from accessing/launching specific templates. The reason is some templates such as delete templates would have catastrophic impact if ran unintentionally. So the access to this type of templates would be typically only granted to administrators or senior users.

All passwords and critical information are also encrypted inside the tool. Therefore users do not have to worry about providing the credentials of their UC servers to the tool.

A. Provisioning Mechanism

UCDT currently can provision Cisco Communications Manager (CUCM), Unity, Cisco Unity Connection, Cisco Unified Presence Server (CUPS), Gateway and switches. Each of these devices provides a different interface for the provisioning. In some cases, even for one device, different provisioning methods should be used since one method is not capable of doing the job or is simply too complicated. For instance although writing SQL queries into the database will allows provisioning many items, due to the complexity of the relations in the database tables and the fact that a few tables should be modified to add an item for some components this will only be considered as an alternative method. Table 1 lists all the provisioning mechanisms we use for different components.

Table 1 Provisioning Methods

UC Platform	Provisioning Method
Cisco Communications Manager (CUCM)	Thick AXL/SOAP and SQL query through thin AXL/SOAP
Unity	SQL through ODBC access
Cisco Unity Connection	CUPI, SQL through ODBC access
Cisco Unified Presence Server (CUPS)	SQL query through thin AXL/SOAP
Cisco Unified Provisioning Manager (CUPM)	NBI
Network Infrastructure (Gateways and Switches)	Telnet, TFTP

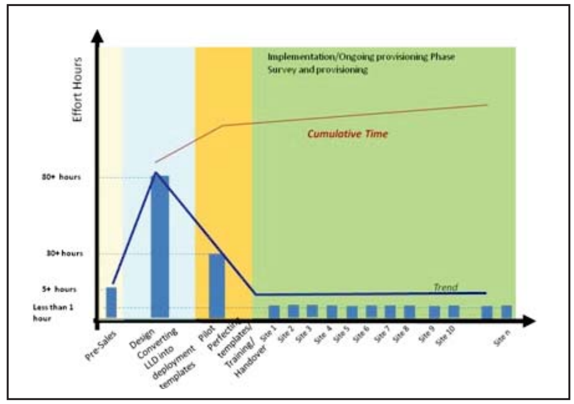


Figure 4 Effort Estimate for using UCDT

UCDT also provides the ability to write plug-ins so that support for any new or even third party devices. Once the plug-in is provided and provisioning mechanism is defined, same infrastructure for template creation, import, and survey can be used for the new device. This has a great advantage for customers since they can deploy only one tool in their network and provision multiple devices.

B. UCDT Template Builder

Template Builder is one of the major modules of UCDT. It provides a template editor and allows users write flexible templates and utilize programming language to conduct complex operations such as string manipulation or complex mathematical calculation. This would greatly simplify the provisioning process as the raw data can be fed to the tool as-is and no data manipulation has to be done over the data prior to feeding to the tool.

UCDT utilizes “Apache Velocity” engine in the template builder module. Therefore users can utilize all the commands available by Velocity to write templates. Once the templates are parsed and variables are replaced with data provided by datasheets, the provisioning requests are formed. Provisioning agent has a feature that allows users view completed provisioning requests, before sending them off to the UC server. This allows detecting parsing errors and other misses in the templates and greatly helps with the troubleshooting.

C. UCDT Datasheets

As it was mentioned, one of the most time consuming tasks for the whole provisioning process is populating datasheets, since it is a manual process and prone to errors. To facilitate the datasheet population UCDT provides concept of tags pick-lists. Tool allows users to define enterprise-wide, system-wide or even branch-wide pick-lists. Once the values are defined for a pick-list it appears as a drop down menu in the datasheet. This allows users to select a value from the menu instead of typing the value. UCDT tags are in fact variables that can be populated only once and be used for all the

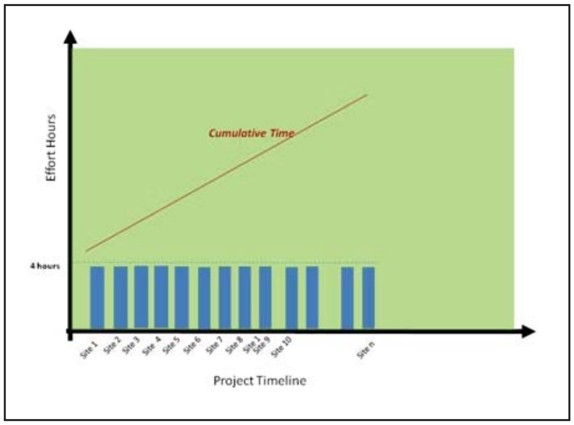


Figure 5 Effort Estimate for Manual Provisioning

templates. This would also simplify populating the datasheets.

The concepts of tags and picklists have greatly helped with facilitating the datasheet population task and have saved users from making mistakes and typos as well as saving time for typing the values.

VI. CASE STUDY

Figure 5 shows the effort hours needed to provision sites without using UCDT. Every site takes about 4 hours to be configured. So the total duration of deployment is 4 hours multiplied by number of sites.

Figure 4 shows the effort hours when utilizing UCDT. Initially around 3 weeks time is needed to create and fine tune the templates. Once templates are created, then provisioning sites would take approximately 15 minutes. So the total duration is number of sites multiplied by 15 minutes plus 3 weeks.

It can be concluded the initial time for writing templates is the most time consuming part. Once templates are created, the provisioning takes very little time. This fact makes UCDT a good candidate for deploying in large enterprises with so many sites/branches. If there are only a few branches for provisioning, UCDT might not have a great impact with respect to the time/cost saving.

Figure 3 shows the impact of utilizing UCDT on a deployment with large number of sites/branches. This is an actual case study from one of the customers.

VI. SUMMARY

This article explains the architecture, internals and process for using Unified Communications Deployment tool. The advantages of the tool are explained and the cost saving figures are presented.

References

1) Garcia, I.; Pacheco, C.; Calvo-Manzano, J.; Using a web-based tool to define and implement software process improvement initiatives in a small industrial setting, IET Software Journal, Volume 4, Issue 4, July, 2010
2) Tokiwa, Y. Nonobe, K. Iwatsuki, M. , Web-based tools to sustain the motivation of students in distance education, Frontiers in Education Conference, 2009. FIE '09. 39th IEEE, pages 1-5, Oct, 2009
3) Erasmia Kitou, Eric Masanet, Arpad Horvath, WEB-BASED TOOL FOR ESTIMATING THE ENVIRONMENTAL IMPACTS OF TELEWOR**K,** Proceedings of the 2001 IEEE International Symposium on Electronics and the Environment, Denver, CO, May 2001
4) Forrester Consulting, Unified Communications Delivers Global Benefits, September 2007, http://www.cisco.com/en/US/prod/collateral/voicesw/ps6882/ps6884/prod_white_paper0900aecd8073c9c3.pdf
5) Elizabeth Herrell, Lindsey Coit, Robert Whiteley, Forrester Research, Time To Revisit Your Unified Communications Priorities, Phase In Key UC Applications To Accelerate Business, August 2010
6) Gartner Research, MarketScope for Unified Communications for the SMB Market, North America, April 2010 http://www.bostonwireless.com/uploads/pdfs/Gartner+MarketScope+for+UC+for+the+SMB+Market+NA.pdf

MYSTORYPLAYER: SEMANTIC AUDIO VISUAL ANNOTATION AND NAVIGATION TOOL

Pierfrancesco Bellini, Paolo Nesi, Marco Serena
Distributed Systems and Internet Technology Lab, Department of Systems and Informatics
University of Florence, Via S.Marta 3, Florence, 50139 Italy, tel +39-055-4796523
nesi@dsi.unifi.it , http://www.disit.dsi.unifi.it

ABSTRACT

Most of the solutions for accessing semantically annotated media provide clients for interpreting simplified information and not to directly executing the semantic annotations. Thus, the interactivity in the navigation among the semantic concepts is closed on the server side, leaving at the client/player tool a passive role. In this paper, a uniform solution (called MyStoryPlayer) that allows modeling and playing non-linear stories by following their semantic including temporal and logical relationships, is proposed. In MyStoryPlayer, any media segment can be an annotation for another media element, plus any semantic description. The user may navigate in the audiovisual annotations and media creating its own non-linear experience/path. The resulting solution includes a uniform semantic model, a corresponding semantic database for the knowledge, a distribution server for semantic knowledge and media, and the MyStoryPlayer to be used in web applications

Index Terms— audio visual annotation, semantic model, semantic player.

1. INTRODUCTION

The content models for entertainment and edutainment are suffering a range of transformations in the recent years. Multistream TV programs such as granprix, bigbrother, etc., are proposing synchronized multiples views/streams of the same event, on which the user may decide to switch. Despite of this effort, the TV is still less attractive than Internet since the possibilities and interactivity are limited (even adopting innovative streaming solution MPEG-21 DIS, digital item streaming [Burnet et al., 2005], or by the many projects on multimedia and multicontext streaming such as LIVE [http://www.ist-live.org]). The most famous TV serials such as Lost, Flashforward, and also the more dated Odissey 5, Doctor Who, etc., may find a more attractive modality of fruition on more interactive channels for fruition. Internet fruition is mainly non-linear. To this end, new models have been proposed in the context of IPTV and/or WebTV. They may take advantage of the possibility of providing content and stream on demand and by the possible direct connection with portals of content distributors, via the so-called hypervideo and/or multimedia links from videos.

The internet capability of rapid context change and the virtually unlimited number of combinations/paths are probably the most relevant motivations for the fast penetration of Web solutions for entertainment, despite the present limitations in network bandwidth. In Internet, on web browsers or via dedicated applications, a large range of entertainment and edutainment solutions are growing in which the user may see at the same time: video and slides, video and chat, etc. Moreover, the adoption of semantic web technologies on internet are transforming the information search and access. Some of the most accessed web portals are starting using semantic web technologies; by indexing content on the basis of their semantic description, and allowing users to provide annotations and classifying content and contributions. See for example users' annotations on images of Flick, and the annotations on videos of YouTube. In the latter case, an annotation may allow to select the successive scene, passing to different videos, modeling the paradigm of Hypervideo (see MIT Hypersoap [Dakss et al., 1998], etc.). Simple paths may be created starting a video from an overlapped link placed on another. Thus, these solutions give at the user the possibility of creating a unique personal experience in the content fruition. In most cases, queries are still traditionally based on keywords, and in some cases full text indexing via some fuzzy support. Thus, the insertion of hyperlinks on videos as annotations does not correspond to enabling semantic queries. That means for example, that it is not possible to perform semantic queries such as requesting *"all videos where two men are talking each other in a car"*. To this end, semantic descriptors of the scene have to be indexed to enable semantics queries, for example in SPARQL. Moreover, for huge collections the computational complexities for semantic extraction and about the inferential processing on semantic indexing related to queries may be huge. Complexity may also depend on the semantic model and functionalities offered to the final users. Therefore, a large effort for content service improvement has been focused to the extraction and association of semantic descriptors with content.

Semantic descriptors are typically modeled as media or multimedia annotations formalized in MPEG-7 and/or RDF [MPEG-7], annotations can be also in MPEG-21 [Bellini 2011]. **Vannotea** solution has been proposed for collaborative annotation of videos [Kosovic et al. 2004].

Vannotea allows the collaborative discussion on video content in real-time. While, the execution model is linear along the video time line. The annotations database has been developed by exploiting the Annotea model and solution W3C [Koivunen et al., 2003].

In this context, critical points are: (i) the detailed modeling of the relationships among media especially along the temporal line, and thus of the ontological model adopted, (ii) the interpretation of the annotations in terms of ontological model with time, and (iii) the effort performed by the server in searching and providing this information to the rendering and exploiting client tools. In most cases, there exists a semantic gap; the server is capable to store and keep knowledge model in terms of ontologies and relationships, while the players receive simple knowledge about the semantic context, which is limited to information strictly needed to the rendering of connected media. In all cases, the capabilities of navigating into the knowledge, and thus the complexity are located on the server side. Thus, the interactivity in the navigation among the semantic concepts is confined on the server side, leaving at the client/player tool a passive role.

In this paper, a uniform solution (called MyStoryPlayer) that allows to model and play annotations and thus non-linear stories by following their internal temporal and logical relationships is presented. The idea is to model audiovisual annotations in terms of RDF including classical descriptors and temporal relationships, for media. The MyStoryPlayer may continuously obtain the successive descriptors and relationships when the user changes the context, for example, selecting media annotations. The final effect consists in allowing the user to create their own experience in navigation by executing the semantic relationships among media and annotations. The solution has been studied and developed as a generalization of the models that may be used to describe and annotate non-linear stories such as Lost, Flashforward, Odissey 5, and classical learning content in histories and humanities in which several different time relationships may be defined. With MyStoryPlayer and tools, the final user may start by making a semantic query to the MyStory Server to obtain as a result a set of possible entry points. During this experience, the user may decide to navigate interacting with them, thus, changing the context in terms of time and related annotations. This allows the user to build a personal experience on the related story.

The paper is organized as follows. In Section 2, the conceptual semantic model of MyStory is presented. Section 3 reports the general architecture, giving details about the server and clients. In Section 4, the internal details of the MyStoryPlayer client tools are presented. An example, which can be accessed via web, is presented on Section 5. Conclusions are drawn in section 6.

2. CONCEPTUAL AND SEMANTIC MODEL

The semantics of relationships among media essences and annotations can be regarded as simple descriptors or effective temporal associations to synchronize them. The annotations may be capable to create/formalize relationships from them and the annotated resources. Moreover, the annotations may consist in semantic descriptors of the content. For example a description of the scene, the description of the image (e.g., the scene is in the forest, or the forest is represented in the scene, Al and Jack are talking each other, Jack has a gun). Thus, each single element of the scene may have its corresponding descriptors as annotations.

The main idea of MyStoryPlayer is to put in the hands of the final users a tool for navigating in the annotations including temporal relationships and with a full access to the single media segments, which may refer to independent media segments of: images, video, and audio essences with associated descriptors. Each of them may be executed according to their semantics and nature. *Thus the execution of a certain scenarios with all its related annotations, leads to the execution of multiple synchronized media.*

In MyStoryPlayer navigation, several annotations and thus audiovisual elements may be executed at the same time. These audio/visual annotations can provide additional and more accurate information, may:

- explode a single time instant with a more complex scenario,
- bring back the story to reminded past events, or to possible futures,
- show a different point of view, or what happen at the same time in another place,
- provide the comment of the director,
- show a video with the scene without visual effects,
- show the historical scenarios,
- show the advertising details of the car which the present in the scene, etc.

In MyStoryPlayer solution, the main aim is on defining a general model to cope with a set of annotations which can be used to navigate among audiovisual scenes according to different possible paths – e.g., the aim of director, the linear time, the activities of a given actor, the movements of an object. The MyStoryPlayer may be used to create new content experiences and models for accessing media in which one may access to a story in a given point and from that may navigate to an indefinite network of annotations.

In MyStoryPlayer, media are synchronized via annotations and segments, and may be executed by the user at the same time, giving the opportunity to the user to jump on a different context by following those annotations, that in this manner become the main stream. The user may pass from one video/essence to another and to return back in the stack of events. The single resources may not be simply connected as temporally aligned videos such as the views of the different cameras in granprix, multiview football match, big brother multiview, etc. This means that the switch from one essence to another may change the context and thus some of the other essences in execution jumping to a different execution time and annotation domain.

The proposed MyStoryPlayer model and tool exploits a semantic model which allows a full RDF navigation among temporal models and audiovisual descriptors and reduces the gap between annotated media essences and the annotations, since in MyStoryPlayer: both of them belong to the same pool of audiovisual.

The main elements of the ontological model of MyStoryPlayer knowledge are reported in Figure 1. The figure does not include the relationships with basic elements such as: URLs, strings, date, time, integer, floats, etc. According to Figure 1, the MyStoryPlayer model is mainly focused on the concept of Annotation. Each Annotation includes a set of descriptions and a media references.

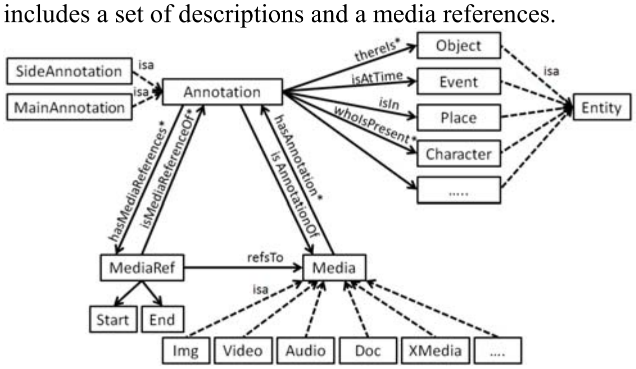


Figure 1. Simplified MyStoryPlayer semantic model.

The scene Annotation descriptor may include multiple details about Characters, Objects in the scene. These kinds of descriptions are useful to perform semantic queries of the scenes. Moreover, Places and Event are singular features. The scene representation model can be augmented with additional ontologies; thanks to the RDF flexibility other features can be added to the model in order be more accurate, complete, suitable for any specific domain and contexts. Any media essence may have multiple annotations; each of them with its own description and reference to other media elements. Among the description also a textual annotation, a starting and ending date/time instant.

The MyStoryPlayer annotations have a starting time, and an ending time. Annotations can be SideAnnotation or MainAnnotation. The two kinds of semantic annotations may be composed to create more complex cases. And the same media segment may be used as SideAnnotation and/or MainAnnotation in different contexts.

The **SideAnnotations** are associated with the main audiovisual by means of a starting time instant, and from that time instant they are played synchronously with the annotated audiovisual (but aside). They start at a given time instant and have a time duration, for that duration the two audiovisual are synchronously played: the main audiovisual in the main area and the annotation aside (the annotation may have to show some associated descriptions as text or additional information). The SideAnnotations are typically adopted for showing audiovisual and/or document/text synchronously such as slides aside the teacher video, different views of the same theatrical scene or sport event.

The **MainAnnotations** are associated with the main audiovisual at a given starting time. From that time instant the annotation is "explosive". This means that, the audiovisual annotation go forward putting in pause the main audiovisual for the duration of the annotation itself. After that, the context returns on the annotated main audiovisual. The return back is automatic if the MainAnnotation is executed up to its conclusion, while the user may decide to change the context during the execution of the MainAnnotation as well. MainAnnotations may be adopted to provide explanations, to attract the interest of the user on different scene, etc.

From the temporal point of view, the time instants of the annotation (starting and ending) are distinct concepts with respect to the Event concept associated with the Annotation. An important feature of scene description is the property *isAtTime* that links the Annotation class with Event, which gives (when possible) at the scene described a temporal collocation, and enables a semantic separation between the real temporal line. Indeed, there are many cases in which the moviemaker decides an order to represent the movie, which does not necessarily follow the occurrence in the real world. Therefore, with this feature, we can separate two parallel temporal lines, the one referred to video, and the other referring to the event ordering into the narrative, to allow the user make temporal queries in which he can ask: "all the scenes before/after event X". The Event concept is related to the narrative semantics of the scene, and represents a symbolic or a time event in the story (a symbolic event is represent by a nickname, for example the "the second goal"; a time event is something which may be defined as DD:MM:YYYY, HH:MM:SS). This means that a clear distinction from time into the story and the time relationships among the digital essences and annotations has been made. For example, it is possible to have set of media and annotations representing a possible experience related to a football game; some videos may include a linear view of the game while others may represent parallel point of views, comments of experts, jump back or forward to other events, detailed facts occurred during the advertising, etc. So that, a combination of Side and Main Annotation is needed to model all kinds of annotations. And, they can be mounted on a fewer number of video files, even in a single file.

Moreover, MyStoryPlayer annotations may be nested, and may be used to organize/create non-linearly and recursive complex paths. An additional flexibility to the model is offered by the fact that each single media file may be referred by any number of annotations, overlapped each other or not.

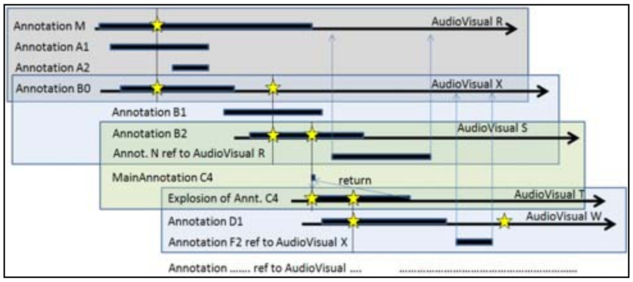


Figure 2. An example described in the next steps:

1. user started with the context of AudioVisual R (for example Video R.flv) on Annotation M (a video segment with association a set of scene descriptors and a text). During the execution of AudioVisual R also Annotation A1, A2 and B0 are executed synchronously according to the time schedule.
2. at the first yellow star (time instant marked with a star), the user decided to change the context by following the content played by Annotation B0. Thus, AudioVisual X was loaded together with its annotations: B1, B2 and Annotation N on AudioVisual R.
3. user completed Annotation B0 and continued on AudioVisual X for a while. Then, the user decided to change context passing at Annotation B2, with the corresponding change to AudioVisual S. AudioVisual S has MainAnnotation C4 (an explosive annotation) aligned at the second star. Thus, the execution passes on Exploding Annotation C4, changing the context on AudioVisual T, to returning back to AudioVisual S. The context of AudioVisual T has annotations D1 and F2. Please note that also Annotation F2 of AudioVisual T may lead again to move the context on AudioVisual X.
4. user, in the middle of executing Annotation C4, decided to pass at Annotation D1 changing again the content to go on AudioVisual W.

The example in Figure 2 represents a case in which the user starts in a certain context and changes the context (audiovisual execution) by following a number of annotations. Thus, these annotations have to be loaded from the knowledge repository by sending a Semantic Navigation Query. The advantages of MyStoryPlayer reside into the semantic model and player tools which include: the management of multiple annotations, the execution of multiple synchronous annotations with time relationships, the possibility of returning back on the stack of context changes including annotations and execution time, in the recording of the experience performed, and in the possibility of having and cumulating a portion of the knowledge on the client side to perform internal navigation without reloading in the RDF model. The possibility of receiving the RDF model of the annotations allows at the MyStoryPlayer to implement strategies on the loading and play of the successive audiovisual/media streams. In fact, it is possible to load in advance the next possible RDF descriptors without waiting for the change of context. This may accelerate the change of context and to perform a partial semantic reasoning on the client side.

3. THE GENERAL ARCHITECTURE

According to the above-presented scenarios, the user starts accessing to the MyStory knowledge by performing a semantic query on a web page. As a result, he/she receives back a list of possible annotations representing scene descriptors from which the navigation may start. Once an annotation/scene is selected, the user experience starts; for example, from video Lost-Season-3-Episode-12 at the time instant 23:34:12. The chosen annotation is put in execution with its corresponding digital media essence, thus the MyStoryPlayer automatically shows a set of synchronized annotations and thus related essences along the execution time line. Annotations are executed aside the main essence according to the time line, and the user may decide to start navigating among them. Thus creating his experience among the possible paths which are modeled by the annotations. The set of annotations are stored into the ontological knowledge database, managed by SESAME, see Figure 3. The MyStory WebPortal front-end is only a web server providing a set of web pages for collecting information, collecting queries, providing support for the client MyStoryPlayer into the internet Browser. The MyStoryPlayer is a player specifically designed and developed for executing the semantic model above presented. It has been developed in Adobe ActionScript of flash, so that it is automatically provided and loaded during the web page loading as flash player tool.

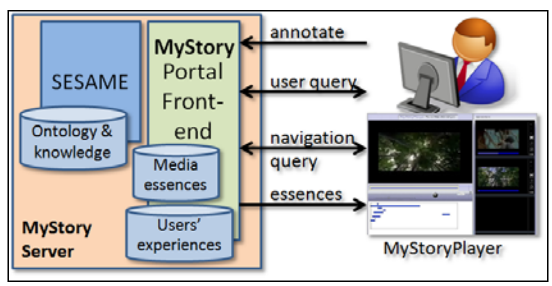


Figure. 3. General architecture of MyStoryPlayer.

The solution proposed works as a client/server application. The solution may be accessed via web browser and may provide more complex experiences including new annotations provided by other users, [Http://www.mystoryplayer.org](http://www.mystoryplayer.org) . On the other hand, the video streaming or progressive download is quite expensive for the network and thus having a number of them is presently possible only exploiting low resolutions video or very powerful connections.

The main functionalities are reported in the architecture of Figure 3. Among them the:

- **addition of annotations** to the knowledge database. The annotations are added providing a set of information such as: ID of the digital resources already in or provided file, starting point, ending point, model of play (sync or explosive), classification metadata, text description, etc.
- **SPARQL queries** on the knowledge database to get a list of possible entry point scenarios.
- **Semantic Navigation Queries** in SPARQL on the knowledge database, to acquire updated information every time the context is changed. For example, when the user click on a video/image/audio on the right side, the MyStoryPlayer poses a query to the server to get all the triples connected to the next context, to explode it in terms of related audio/visual, and thus to start the progressive download and player of the related resources.
- **Saving users' experiences on the server.** When a users takes a decision on recording the experience in navigating. This implies that for action the player has to communicate to the server the time instant at which the action has been performed on the timeline of the current media resources. This allows keeping trace of the whole user experience, and may be used to share this experience with other users. The sharing of experience is of interest for educational and entertainment purposes.

Potentially the user may navigate in any direction among the annotations contained into the knowledge database, following the links and audio/video/images which are presented along the experience and navigation. The recorded users' experiences may be re-proposed in a second moment to the same user or to other users that may be interested to make the same experience.

4. INSIDE MYSTORYPLAYER

One of the most interesting tools of the proposed solutions is the MyStory client tool, namely the MyStoryPlayer. In order to satisfy the above-described main features, the MyStoryPlayer has been designed to perform:

- interpretation of RDF information and query results coming from the MyStory Server,
- understanding and recognizing the different kinds of annotations,
- interpreting media references and thus relationships among elements,
- supporting Main and Side Annotation semantics,
- execution/interpretation of multiple annotations and thus of multiple audiovisual streams related to main media essences and Annotations.
- change of context when the user click on an Annotation (video, audio, image), implies to dynamically perform the access to the media of the next Annotation on WebPortal of MyStory.
- visualization of information and text associated with the current main audiovisual and of the Annotations; Textual descriptors to the single annotation and resources can be used to make in evidence specific aspects and to start discussions.

The MyStoryPlayer client tool has been designed by using object-oriented technology, see Figure 4. It has been designed to be loaded into a large range of Internet Browsers and thus in a range of web based solutions and applications. According to the design, the player is capable to receive annotations in RDF, where they are internally interpreted according to their semantic: audio, video, image, etc.

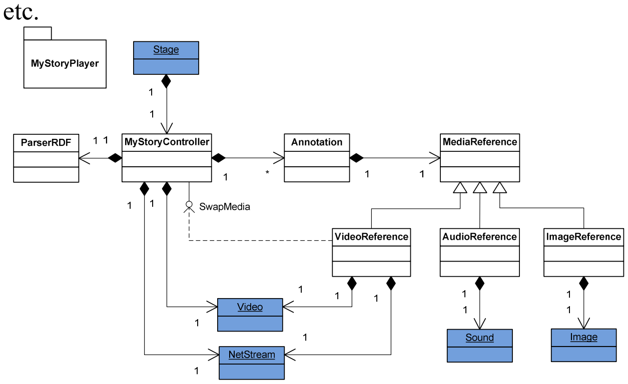


Figure 4. The main classes of MyStoryPlayer client tool and their relationships

5. USING MYSTORYPLAYER

According to the semantic model adopted, the user may perform semantic queries in SPARQL. For example, we have coded and annotated a subset of Lost TV serial to show the mechanism, and the user may perform the following SPARQL queries to get all scenes (the authors are requesting the rights to publishing the example proposed by using Lost serial, while other examples can be used as well): in which John and Jack are in a Forest; in which is present a Knife and are on the Beach. More complex semantic models could be adopted supporting also queries as: in which Kate has a Gun after the Second Airplane Disaster (as Event); on the Beach or in the Forrest, etc. For example, an example of semantic query in SPARQL to get "All the scenes where is present Jack and there is a gun", is reported in the sequel:

SELECT DISTINCT ?Video ?Annotation WHERE {
?Video a msp:Video. ?Video msp:hasAnnotation ?Annotation.
?Annotation msp:whoIsPresent ?character. ?character rdfs:label "Jack".
?Annotation msp:thereIs ?object. ?object rdfs:label "Gun".
}

After some seconds of the main video execution, other 4 video annotations start and are executed at the same time on the right side of the MyStoryPlayer in web page. The videos on the right side may contain explanations, simple flashback, contemporaneous events, as well as related issue. Each of them presented a different starting point and duration. In this context, the user may decide to interact with the presented media objects and controls.

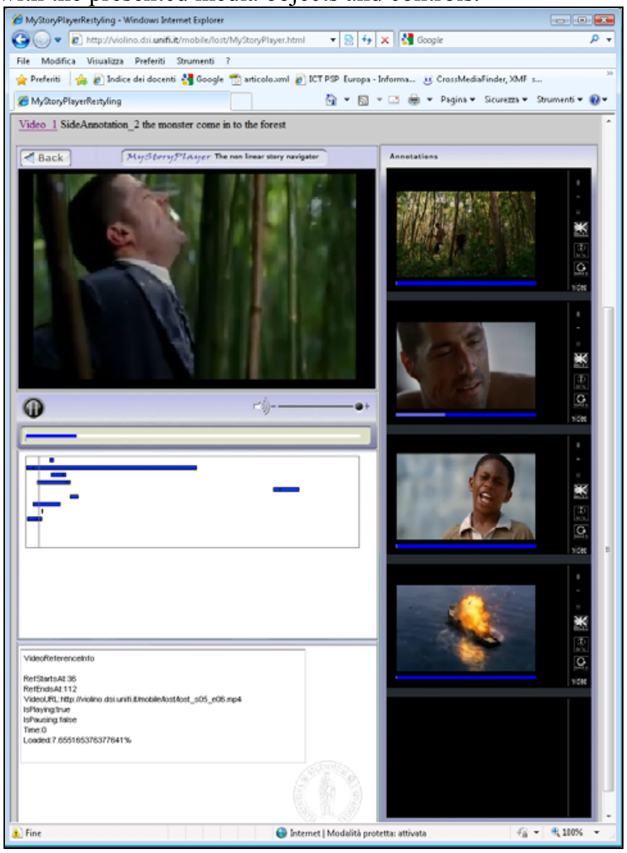


Figure 5. MyStoryPlayer at work

Figure 5 shows a case in which an annotation started with a video (the experience can be performed on http://www.mystoryplayer.org by looking for all scenes with Jack and a gun).

For example: to follow the play of the multiple streams by keeping active the audio on one or more of them, or may decide to change the context; to see the textual annotations associated with one to the video by selecting corresponding "info" button. By clicking one of the videos (audio or images as well) on the right provokes the change of context, bringing the clicked media in the first place (on left). From the point of view of the ontological model, the change of context means to contact the MyStory Server with a query to get the annotations of the selected media and to recreate the knowledge into the MyStoryPlayer. The knowledge may grow around the present network of annotations and the past annotations may be discharged after some jumps. It is evident that new annotations have been loaded while the backlog of the past annotation is still accessible. In order to guarantee a number of back jumps to rewind the experience, the stack of the activities performed have to be maintained, storing also time instants in which the changes of context have been performed.

6. CONCLUSIONS

The paper presented MyStoryPlayer model and tool to execute non-linear stories by following their internal temporal and logical relationships formalized via semantic annotations. In MyStoryPlayer, any media segment can be an annotation for another media element. The user may navigate in the audiovisual annotations creating its own experience. The resulting solution includes a uniform semantic model, a semantic database, a distribution server for semantic knowledge and media, and the MyStoryPlayer to be used in web applications. The solution has been tested on a number of scenarios related to the modeling of non-linear story telling. An example is also accessible via web for the reviewers of the paper. In the short future, the solution would be adopted for modeling educational material such as those that can be created for medical and theatrical environments and in particular for ECLAP PsP project of the European Commission. In both cases, the scenes are typically recorded from more than one point of view, additional explanation video have also to be added showing the historical aspects (on different time and spaces) or about the basic technologies/interpretations or comparison.

Acknowledgement

The authors would like to thanks to ECLAP project partners.

7. REFERENCES

Bellini, P., I. Bruno, P. Nesi, "Exploiting Intelligent content via AXMEDIS MPEG-21 for modeling and distributing news", on Int. Journal of Software Engineering and Knowledge Engineering, vol.21, n.1, 2011

Burnett, I.S., Davis, S.J., Drury, G. M., "MPEG-21 digital item declaration and Identification-principles and compression", IEEE Transactions on Multimedia, Vol.7, N.3, pp.400-407, 2005.

Dakss J., Stefan Agamanolis, Edmond Chalom, V. Michael Bove, Jr., ``Hyperlinked Video," Proc. SPIE Multimedia Systems and Applications, v. 3528, 1998.

Koivunen, M., Swick, R., and Prud'hommeaux, E., "Annotea Shared Bookmarks", 2003, In Proc. of KCAP 2003, http://www.w3.org/2001/Annotea/Papers/KCAP03/annoteabm.html

Kosovic D., Ronald Schroeter, Jane Hunter, "FilmEd: Collaborative Video Annotation, Indexing and discussion tool over Broadband networks", International Conference on Multimedia Modeling archive Proceedings of the 10th International Multimedia Modelling, Page: 346 , 2004.

MPEG-7: http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm

Feature Model Debugging based on Description Logic Reasoning

Mahdi Noorian[1], Alireza Ensan[1], Ebrahim Bagheri[1,2], Harold Boley[3], and Yevgen Biletskiy[1]

University of New Brunswick, Fredericton, Canada[1]
Athabasca University, Edmonton, Canada[2]
National Research Council Canada[3]
m.noorian@unb.ca, alireza.ensan@unb.ca, ebagheri@athabascau.ca, harold.boley@nrc-cnrc.gc.ca, biletski@unb.ca

***Abstract*—Software product line engineering refers to the concept of sharing commonalities and variabilities of a set of software products in a target domain of interest. Feature models are one of the prominent representation formalisms for software product lines. Given the fact that feature models cover all possible applications and products of a target domain, it is possible that the artifacts are not necessarily and always consistent. Therefore, identifying and resolving inconsistencies in feature models is a significant task; especially, due to the fact that a large number of possible products and complex interactions between the software product line features need to be checked. To address these challenges, in this paper, we propose a framework with an automated tool to find and fix the inconsistencies of feature models based on Description Logic (DL) reasoning. The basic idea of our approach is to first transform and represent a feature model using Description Logics. The second step is to identify the possible inconsistencies of the feature model using DL reasoning and then recommend appropriate solutions to a domain analyst for resolving existing inconsistencies.**

I. Introduction

Software Product Line (SPL) engineering is a paradigm that models families of software products with similar characteristics [12]. These characteristics, known as features, can characterize the specifications of a given domain. In fact, SPLs employ features to model all of the possible configurations and applications of a domain. The use of these features that allow for the sharing of commonalities among features of a domain strengthens the capability of SPLs to be used in re-use-based software development [15]. SPLs are usually represented using Feature Models (FM) that can indicate variability and commonality of product lines in a graphical representation. FMs have a tree structure where each feature is a node of the tree and its edges are the possible variabilities of the features.

Within the software development lifecycle of SPL, the process of developing a feature model for a given domain is referred to as the Domain Engineering phase, whereas the process of selecting the desirable features from the designed feature model for developing a new applications is called the Application Engineering phase. In both steps, inconsistent definitions and designs can be encountered that need to be automatically detected and semi-automatically resolved which requires the development and customization of AI techniques specifically for this purpose.

The task of inconsistency identification and resolution within a feature model is the main concern of this paper. Several approaches have been proposed to verify the validity of a feature model and its products; but only very few techniques exist that address the issue of inconsistency resolution within feature models. Furthermore, there is no automated tool available that is able to perform all of the tasks of inconsistency checking, debugging and resolution within feature models; so this fact shows the necessity for the development of a comprehensive automated AI-centric tool for this purpose. To this end, we propose a framework that describes the whole process of FM design and product generation based on Description Logics. It can check the validity of a given model, and all of its products; also it gives resolutions for invalid products or inconsistent models.

In real world applications such as MS Office, millions of features exist in the feature model. To generate new applications from this huge number of features, a comprehensive and reliable method is needed because feature model validation in this case would need too much time to be properly performed by a human analyst. So if a domain analyst is responsible for validating a large product line, he would need an automated tool that would help him/her by suggesting the best configuration and resolution strategies in an efficient amount of time. For this purpose, the strength of DL reasoning can be useful to guide domain analyst within the process of domain verification.

In this work, we propose a formal framework for performing inconsistency checking, validation and resolution based on DL reasoning for SPL. Using DL reasoning techniques which are implemented within Pellet [16] and also based on Reiter's algorithm [13], we can find the solutions for resolving inconsistencies within a feature model. The main contributions of this work can be summarized as follows:

- We show the importance of using Description Logic variants especially OWL-DL in the area of software product lines;
- We propose a comprehensive framework for inconsistency checking by employing the Pellet reasoning engine. Also, we utilize Reiter's algorithm to propose solutions for resolving the identified inconsistencies.

- Based on our framework, we develop a tool called AUFM which supports the design of FMs in both domain and product levels. Hence, it identifies and resolves inconsistencies in the FM and its related products. Indeed, developing such a tool in the area of software product line can be helpful for domain analysts who are dealing with feature model validation and verification.
- We ground the problem of inconsistency detection and resolution on sound Description Logic representation and reasoning mechanisms that would allow us to ensure that the proposed revision solutions are at least syntactically acceptable.

The remainder of this paper is as follows: The next section reviews some preliminaries. Then, our approach for inconsistency resolution in feature models is described in Section III. In Section IV, we support our approach by analyzing a sample feature model. Subsequently, we evaluate the proposed approach with a case study in Section V. After that, in Section VI we discuss related work in this area. Finally, Section VII is devoted to conclusions and some future work.

II. Preliminaries

A. Feature Models

Features are important distinguishing aspects, qualities, or characteristics of a family of systems [11]. They are widely used for depicting the shared structure and behavior of a set of similar systems. To form a product family, all the features of a set of similar/related systems are composed into a feature model. A feature model represents the possible configuration space of all the products of a system product family in terms of its features. Feature models can be represented both formally and graphically; however, the graphical notation depicted through a tree-like structure is more favored due to its visual appeal and easier understanding.

In a FM, features are hierarchically organized by *Structural Constraints* which can be typically classified as: 1) *Mandatory*: a feature must be included in the description of its parent feature; 2) *Optional*: a feature may or may not be included in its parent description given the situation; 3) *Alternative feature group*: one and only one of features from the feature group can be included in the parent description; 4) *Or feature group*: one or more features from a feature group can be included in the description of the parent feature. In some case, the tree structure of feature models falls short at fully representing the complete set of mutual interdependencies of features; thus, additional constraints are often added to feature models and are referred to as *Integrity Constraints*. The two most widely used integrity constraints are: *Includes* - the presence of a given feature (set of features) requires the inclusion of another feature (set of features); and *Excludes* - the presence of a given (set of) feature(s) requires the elimination of another (set of) feature. In the following sections, the term 'constraint' refers to both integrity and structural constraints unless it is specifically mentioned.

Inconsistent Feature Model (IFM): An IFM is a FM that is vulnerable to generate invalid products with regard to structural and integrity constraints [18], [19]. In other words, an IFM violates some constraint (both integrity and structural) simultaneously. As discussed in [18], inconsistency in the FM can happen in two levels including domain and product configuration level when some defined constraints are violated.

B. Description Logics (DL)

Description Logic as a subset of First Order Logic is a knowledge representation formalism that can help to effectively perform reasoning over a knowledge base. A knowledge base modeled using Description Logics could be defined as $\Psi = (\mathcal{T}, \mathcal{A})$, where $\mathcal{T}$ denotes TBox and comprises of a set of general inclusion axioms and $\mathcal{A}$ stands for ABox and comprises of a set of instance assertions. This kind of knowledge representation contains a set of all concept names (C_N), role name (R_N) and individuals (I_N). The semantic of DL-knowledge base is defined by an interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, .^{\mathcal{I}})$ where $\Delta^{\mathcal{I}}$ is a non-empty set of individuals and $.^{\mathcal{I}}$ is a function which maps each $C \in C_N$ to $C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, each $R \in R_N$ to $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ and each $a \in I_N$ to an $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$. An interpretation $\mathcal{I}$ satisfies a TBox axiom $C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. An interpretation $\mathcal{I}$ is a model of a TBox $\mathcal{T}$ iff it satisfies all of its axioms. Furthermore, an interpretation $\mathcal{I}$ is a model of knowledge base Ψ if it satisfies every TBox axiom and ABox assertion of Ψ. A concept C is *unsatisfiable* with regards to TBox $\mathcal{T}$ iff $C^{\mathcal{I}} = \varnothing$ for all models $\mathcal{I}$ of $\mathcal{T}$. In addition, a TBox $\mathcal{T}$ is called *incoherent* iff there is an unsatisfiable concept in $\mathcal{T}$. For detailed introduction to description logic, interested readers can refer to [1].

C. DL-knowledge base Diagnosis and Debugging

In the area of knowledge representation and specially the semantic Web, knowledge base quality assurance is a prominent task. In the literature, various model-based diagnosis methods have been proposed for the purpose of developing a consistent model of a DL knowledge base. Schlobach et al [14] have utilized Reiter's algorithm for identifying the underlying reasons of inconsistencies in a DL knowledge base, which is based on generating the *conflict sets* and the related *minimal hitting sets*. A set is called conflict set when it includes such elements that causes incoherencies. A hitting set is further defined based on the concept of a conflict set such that it contains at least one element from the collection of the conflict sets. A hitting set is *minimal* iff it does not have any subset which is a hitting set. In order to implement Reiter's algorithm, the concept of Minimal Unsatisfiability-Preserving Sub-TBoxes (MUPS) is used, which refers to minimal conflict sets in the TBox. MUPS of a TBox for unsatisfiable concept A are the subsets of the TBox where A is satisfiable. MUPS can be used to compute the Minimal Incoherence- Preserving Sub-TBox (MIPS), which explains the incoherence of a TBox [3]. Consider the TBox $\mathcal{T}$ in Table I [14].

The unsatisfiable concepts are A_1, A_3, A_6, A_7, e.g., MUPS for A_3 is $\{ax_3, ax_4, ax_5\}$. Also MIPS for TBox $\mathcal{T}$ is

TABLE I
AN INCOHERENT TBOX $\mathcal{T}$

$ax_1 : A_1 \sqsubseteq \neg A \sqcap A_2 \sqcap A_3$	$ax_2 : A_2 \sqsubseteq A \sqcap A_4$
$ax_3 : A_3 \sqsubseteq A_4 \sqcap A5$	$ax_4 : A_4 \sqsubseteq \forall s.B \sqsubseteq C$
$ax_5 : A_5 \sqsubseteq \exists s.\neg B$	$ax_6 : A_6 \sqsubseteq A_1 \sqcup \exists r.(A_3 \sqsubseteq \neg C \sqcap A_4)$
$ax_7 : A_7 \sqsubseteq \forall s.B \sqcap C$	

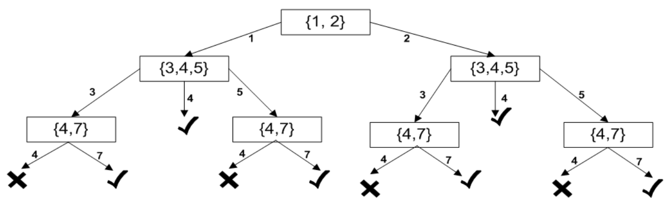


Fig. 1. Reiter's hitting set tree for MIPS in TBox $\mathcal{T}$.

$\{\{ax_1, ax_2\}, \{ax_3, ax_4, ax_5\}, \{ax_4, ax_7\}\}$. Fig.1 represents the Reiter's hitting set tree for MIPS in the above sample incoherent TBox $\mathcal{T}$. Since MIPS is a minimal conflict set, the minimal hitting set tree can be developed based on the selection of at least one element from each of its subsets.

The check mark symbols in the leaves indicate the successful diagnosis path from root of the tree. Thus, we can have the following minimal hitting sets:
$\{\{ax_1, ax_4\}, \{ax_2, ax_4\}, \{ax_1, ax_3, ax_7\}, \{ax_2, ax_3, ax_7\}, \{ax_1, ax_5, ax_7\}, \{ax_2, ax_5, ax_7\}\}$
In order to make the TBox $\mathcal{T}$ coherent, we can simply choose one element form the produced hitting set and omit the axioms of the chosen set. Note that, the members of the hitting set can be selected based on syntactic measures such as different scoring functions on axioms of MIPS.

III. The Proposed Approach

In our approach, we focus on the representation of software product line feature models as Description Logic knowledge bases, and consequently use formal reasoning approaches on such a knowledge base to identify and possibly resolve any inconsistencies. In order to support our theoretical framework, we have developed an automated tool from which domain analysts can benefit for the purpose of domain and product consistency validation. First, the domain analysts can begin by designing a feature model using the *AUFM FM Editor*. Next, for the purpose of consistency validation, the feature model which is in SXFM format (SXFM is XML standard for representing a FM) can be sent to the *AUFM Feature Model Debugger*. Then, by applying three sequential processes: 1) *Transformation*; 2)*Inconsistency Identification*; and 3) *Inconsistency Resolution*, the domain analysts receive verification feedback. The feedback message contains, what the inconsistent features/constraints are, why those features are/cause inconsistency, and how the existing inconsistencies can be resolved. This information can be used by the domain analysts to correct the inconsistent FM and make it consistent. Note that, feature model verification is an iterative process and it will continue until the domain analysts are satisfied with their design and product configuration. An overview of this process is shown in Fig.2.

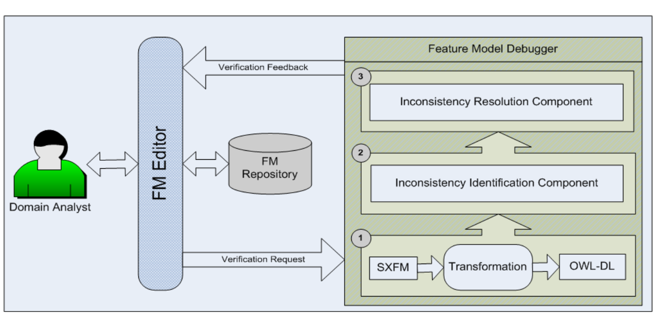


Fig. 2. A semi-automated inconsistency management framework.

TABLE II
DL-KNOWLEDGE BASE STRUCTURE DERIVED FROM A FEATURE MODEL.

Feature relation or integrity constraints	Description logic modeling
$G \sqsubseteq T$,and $GRule \sqsubseteq T$ $F_1 \sqsubseteq T$,and $F_1Rule \sqsubseteq T$ ...$F_n \sqsubseteq T$, and $F_nRule \sqsubseteq T$	hasG $\sqsubseteq$ ObjectProperty , T $\sqsubseteq$ $\forall$hasG.G, and GRule$\equiv$ $\exists$hasG.G. Also,
Also, $G \sqsubseteq \neg F_i$, for $1 \leq i \leq n$, $F_i \sqsubseteq \neg F_j$, for $1 \leq i, j \leq n$ where $i \neq j$.	has F_i $\sqsubseteq$ObjectProperty , T $\sqsubseteq$ $\forall$has $F_i.F_i$, and $F_iRule \equiv \exists hasF_i.F_i$. for $1 \leq i \leq n$.

According to the framework, *AUFM feature model debugger* consists of three main components. In the following subsections, we will explain each of them in detail.

A. The FM to OWL- DL Transformation Component

In order to resolve inconsistencies in a feature model through Description Logic reasoning, the first step is to convert the feature model representation into some variant of Description Logics. We have chosen OWL-DL for this purpose. We utilize some theoretical concepts discussed in [20] to represent a FM in OWL-DL to benefit from the reasoning and expressiveness provided by it. The conversion is implemented in both the domain and the configuration levels.

The domain conversion phase includes several tasks that need to be performed in order to transform a FM into OWL-DL. The first step of conversion is to assign an OWL class named *Feature Class* for each feature in the domain; and all *Feature Classes* are considered mutually disjointed. The second step is defining a *Rule Class* for each *Feature Class*. Such *Rule Class* is related to its corresponding *Feature Class* with *necessary* and *sufficient* condition which is restricted by existential restriction. For example, assume that F is a *Feature Class*. So, corresponding $FRule$ class is defined as $FRule \equiv \exists hasF.F$. All relations that an individual *Feature Class* might have with other *Feature Classes* are embedded in its corresponding *Rule Class*. The aforementioned possible relations can be defined as structural or integrity constraints. In other words, each *Rule Class* represents all relations that the related *Feature Class* has with its children in the necessary condition definition; this is the main reason that a *Rule Class* is defined. Consequently, for a root feature G and its children $F_1, F_2,..., F_n$, the associated DL-knowledge base has the general structure represented in Table II.

TABLE III

LOGIC REPRESENTATION FOR STRUCTURAL AND INTEGRITY CONSTRAINTS IN FMS

Structural or Integrity Constraints	**Description Logic Modeling**
Mandatory	*GRule*$\equiv \exists\, hasF_1.F_1$... *GRule*$\equiv \exists hasF_n.F_n$
Optional	Does not impose any constraints
Or	*GRule*$\sqsubseteq \sqcup(\exists hasF_i.F_i)$ *for*$1 \le i \le n$
Alternative	*GRule*$\sqsubseteq \sqcup(\exists hasF_i.F_i)$ *for*$1 \le i \le n$
	GRule$\sqsubseteq \neg \sqcup (\exists hasF_i.F_i \sqcap \exists hasF_j.F_j)$ *for* $1 \le i \le j \le n$
Include(Require)	*GRule*$\equiv \exists hasF_1.F_1$...*GRule*$\equiv \exists hasF_n.F_n$
Exclude	*GRule*$\equiv \neg(\exists hasF_1.F_1)$...*GRule* $\equiv \neg(\exists hasF_n.F_n)$

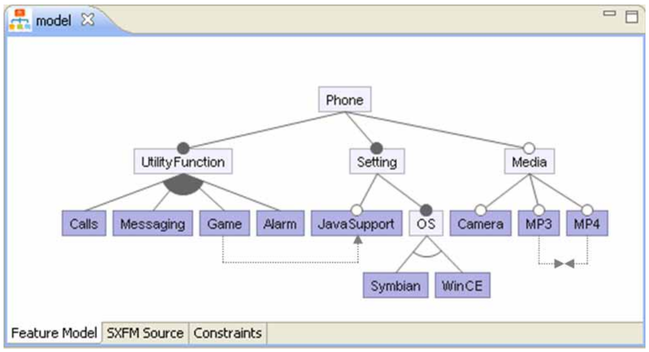


Fig. 3. The AUFM feature model editor.

As discussed before, there are two types of constraints including structural and integrity constraints in a FM. For a parent feature G and its children $F_1, F_2,..., F_n$, DL representation for all kinds of constraints are summarized in Table III. It is worth noting that, *Optional* relations do not impose any extra constraints to the feature model and DL-knowledge base, so we are not concerned with *Optional* relationships in the DL-knowledge base model.

The meaning of parent feature and its children is similar in an *Include/Excludes* constraint which means that if feature G is selected for one product, it is necessary that features $F_1, F_2,..., F_n$ are included/excluded in that product.

The products conversion phase is a straightforward task in light of the domain conversion step. In order to model a product of a FM using DL, we can add a class called $ProductClass$ and consider which features exist in that product and which do not. Mathematically speaking, for a FM with root feature G and set of features $F_1, F_2,..., F_n$, a product that includes features $F_1, F_2,..., F_i$ and does not contain features $F_{(i+1)}$,$F_{(i+2)}$,...,F_n is defined as follows:

$Product \sqsubseteq GRule$

$Product \equiv (\sqcap(\exists hasF_j.F_j\ for 1 \le j \le i \le n) \sqcap (\neg\exists hasF_k.F_k fori < k \le n))$

B. The Inconsistency Identification Component

Basically, this component is responsible for finding the inconsistencies in the FM and its related products. For a given feature model, the source of inconsistencies can be two-fold: *structural constraint violation* and *integrity constraint violation*. After transforming feature model to OWL-DL, we can perform DL reasoning with any standard DL reasoner to pinpoint the inconsistent parts of the feature model. Several DL reasoners are already available such as FaCT++ [17], RACER [9], and Pellet [16] that implement Tableau based reasoning [1]. Pellet has been seamlessly plugged into our designed tool and we can take advantage of its facilities to implement Reiter's algorithm for inconsistency resolution, which will be discussed in the next subsection. Thus, we have employed Pellet as an integrated reasoning engine in AUFM and are using it for the purpose of inconsistency checking.

C. Inconsistency Resolution Component

The inconsistency resolution component is an important part of the *AUFM feature model debugger*. The aim of this component is to find the minimum corrections needed to make an inconsistent FM into a consistent one. In other words, we are looking for a minimal subset of axioms in TBox (here TBox refers to the converted FM into DL) that need to be repaired or removed to render a correct FM and thus make it consistent again. Note that here the TBox contains all *Feature Classes*, *Rule Classes*, and their correlated properties which are discussed in Tables II and III. Also, at the product configuration time, the TBox will be extended by adding *Product Classes*. For any incoherent TBox (inconsistent FM), the reasoning engine returns the list of unsatisfiable axioms (features and constraints). At this point, AUFM automatically performs debugging and exposes the main reasons of inconsistency and provides reasonable solutions for resolving them. All these tasks are carried out via a user friendly interface.

We adopt the theory of debugging and diagnosis from [14] within our work. Based on this, for a given inconsistent FM, AUFM finds the minimal conflict set for corresponding TBox. Note that, here conflict set can be defined as a MIPS set. Afterwards, Reiter's hitting set algorithm is used to find the minimal hitting set. Indeed, MIPS set expresses why the feature model is inconsistent according to errors that are found by the reasoning engine. In addition, the hitting set represents how these errors can be resolved. AUFM supports both services inherently and domain analyst can benefit from them automatically.

IV. A SAMPLE PROCESS DEPICTION

To demonstrate the process of our framework, in this section we illustrate the resolution of a sample inconsistent feature model. According to our proposed approach, domain analysts should go through the following steps to validate any target feature model and its related products.

1. Design a feature model using the AUFM editor: We have provided the AUFM feature model editor which can

TABLE IV

RESULTS FROM INTERACTIVE DEBUGGING PROCESS

Steps	Inconsistencies	Recommended solutions	More Inconsistency?
1	Exclude violation	$MP3Rule \equiv \exists hasMP3.MP3$	Yes
2	Alternative violation	$OSRule \equiv \exists hasOS.OS$	Yes
3	Mandatory violation	$PhoneRule \subseteq \exists hasUtilityFunction.UtilityFunction$	No

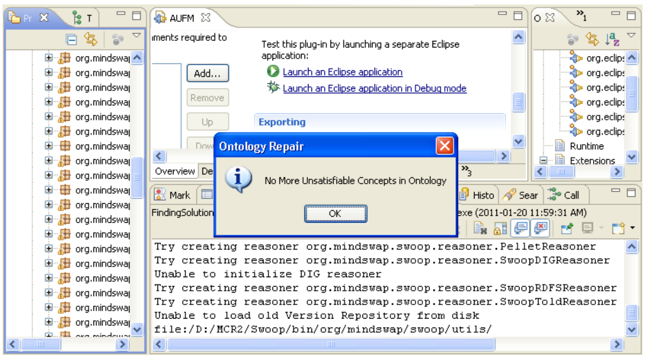


Fig. 4. Consistency checking result for Phone feature domain.

be used as a plug-in in the Eclipse platform. As depicted in Fig.3, the *Phone* domain is graphically designed with 16 features which is a simple part of a real phone domain. The corresponding SXFM file will be generated automatically by AUFM. We will use this feature model as an example for presenting our framework. The SXFM file for this can be accessed from *http://falcon.unb.ca/*$\sim$ *j41z9/files.html*.

2. Transformation from SXFM to OWL-DL: In this stage, we need to transform the *phone* feature model into its DL representation. For this matter, the produced SXFM file will be used and fed as an input to the *transformation* phase. The result of transformation component is an OWL-DL knowledge base that represents the input feature model. The generated OWL file for *phone* feature model is available at *http://falcon.unb.ca/*$\sim$ *j41z9/files.html*.

3. Consistency checking and inconsistency resolution: Consistency checking can be performed in two levels, feature domain and configuration levels. Here first, we perform consistency evaluation for *Phone* feature model in the feature domain level. Furthermore, by configuring a sample product we will show how AUFM can identify probable inconsistencies. In both levels, Pellet has been used as plug-in in AUFM to perform tableau based reasoning to check consistency in the TBox which is generated from the *phone* feature model.

1) Consistency checking in feature domain level: Fig.4 represents the consistency checking result for the *Phone* feature model at the feature domain level. As illustrated, the AUFM reasoning engine does not find any unsatisfiable concepts. Thus, *Phone* feature model is consistent in terms of feature model structural and integrity constraints.

2) Consistency checking and inconsistency resolution in configuration level: Before starting consistency checking in this level, we need to configure a product. We configured an inconsistent product, *P*, which violates one *Mandatory*, *Alternative*, and *Exclude* constraints. Consider*P* as follows:

P= $\{$*Phone, Setting, Calls, Messaging, OS, Symbian, WinCE, Media, MP3, MP4, Camera*$\}$.

The OWL file for product *P* can be accessed from *http://falcon.unb.ca/*$\sim$ *j41z9/files.html*. According to the structural and integrity constraints of the *phone* feature model, product *P* cannot have both of the *Symbian* and *WinCE* features at the same time. Also, *MP3* and *MP4* are related with *Exclude* constraint. Thus, these two features cannot be included in product *P* simultaneously. In addition, *UtilityFunction* is defined as a *Mandatory* feature and it must be included in product *P*, while it is not.

AUFM performs consistency validation which is an iterative process. In each iteration, the system detects any possible inconsistencies and provides its corresponding solution and then sends a report back to the domain analysts. Then domain analysts are able to look at and resolve inconsistency (based on the recommendations from AUFM) and send the edited FM to *feature model debugger* for further inconsistency identification. This process will continue until there no more inconsistency could be found in the designed FM.

Product *P* has three inconsistencies. Thus, the consistent product can be gained after three steps of product debugging. In the first step, AUFM finds that the domain analysts have violated the *Exclude* constraint, which connects *MP3* and *MP4* features. Afterwards, the solution (which is *MP3Rule*$\equiv$ $\exists$*has MP3.MP3*) for resolving this error would be recommended by AUFM. This problematic axiom is found based on Reiter's algorithm. As discussed in Section III, to implement the Reiter's algorithm, hitting set trees should be drawn and traversing in the edge weighted tree can determine the valid solutions for resolving inconsistencies. The drawn hitting set tree for resolving the first inconsistency which is *Exclude* violation is available at *http://falcon.unb.ca/*$\sim$ *j41z9/files.html*. In order to resolve this inconsistency, domain analysts need to remove this axiom from the knowledge base. With regards to the meaning of the feature model, *MP3* should be removed from the product *P*. In other words, domain analysts interpret the meaning of the proposed solutions based on the semantics of the feature model. In product *P*, *MP3* and *MP4* are related with an *Exclude* constraint; in order to resolve this inconsistency, domain analyst needs to omit one of them from the configured product *P*. After domain analysts resolve the first error, they may ask whether there is more inconsistency or not. This process will continue until the second and third inconsistencies which are *Alternative* and *Exclude* violation are resolved. At the end of this interactive debugging process, product *P* is turned into a

TABLE V
THE CASE STUDY RESULTS (TIME IS BASED ON MINUTES)

Student	Number of feature	Number of structural constraint	Number of integrity constraint	Number of inconsistency	Elapsed time for debugging
st1	32	12	4	3	2.21
st2	27	10	6	7	5.10
st3	35	15	6	5	3.66
st4	30	11	3	4	2.45

consistent model. Table IV represents the inconsistencies and corresponding recommended solutions which are provided by AUFM.

V. EXPERIMENTAL EVALUATION

In this section, we report on a case study to investigate the suitability of the proposed approach in terms of its usability. Here the proposed framework is evaluated to find out whether our approach can be practically helpful for domain analysts in the case of feature model consistency checking and debugging. In the following, we first describe the initialization process for usability evaluation and then we present the results and provide discussion in this regards.

Initialization process: In this case study four participants were invited to comprehensively model a *Smart Phone* domain and design corresponding feature models using our tool, AUFM. Two of these participants were undergraduate students which had little background about domain modeling. The other two participants were graduate students which their research areas were software product line engineering.

We recorded each participant's activities based on some metric criteria such as number of features, integrity constraints, structural constraints, and inconsistencies in the *Smart Phone* domain. Also, we recorded the elapsed time of the iterative debugging process which mainly includes inconsistency identification and inconsistency resolution process.

Result and discussion: The case study observations are summarized in Table V.

As seen in Table V, students subjectively design *Smart Phone* feature models with average 31 features and 17 constraints. Also, the number of inconsistencies that has been discovered by AUFM varies between 3 to 7. In fact, the source of this variation for inconsistency is twofold. First, each individual participant has various levels of knowledge about domain modeling and second, the number of features and constraints that have been used by the students could affect the number of inconsistencies. That is, the more features and constraints, the more the designed model is susceptible to error.

The last column of Table V represents the elapsed time for the debugging process. The debugging process begins once the students send their initial inconsistent feature models to *AUFM feature model debugger*. The process continues until all the inconsistencies are resolved. In this case study, we have observed that the debugging process is varying almost between 2 to 5 minutes. In fact, we noticed that for the experienced users, the elapsed times are fairly acceptable for the feature model with average 31 features. Although, by considering the fact that st2 is the least experienced in comparison with the others; 5 minute of debugging could be considered acceptable.

After evaluating the results and discussing with students, we conclude that the AUFM tool is a user-friendly tool such that students with different levels of knowledge could easily design whatever they have in their minds. AUFM provides the means for students to interactively identify the inconsistencies of their designs and resolve them based on the recommended solutions. However, some of the students pointed that the recommended solution is at times ambiguous and it takes time to understand them in some cases. The main reason for this problem is that the recommended solutions are based on the DL language. They suggest that, it would be more efficient, in terms of time, if the users have this solution according to the feature modeling language rather than DL. Thus, they would be able to just concentrate on their design and directly correct their designed domain.

VI. RELATED WORK

Since the advent of feature models in the early nineties [10], a lot of work has been done in this area. Researchers have devoted their efforts to various stages of software product line engineering including works such as SAT-based feature configuration [21], AHP-based model ranking and preference elicitation [2] and machine learning-based external attributes quality prediction [4].

Some attempts have also recently been made to to find inconsistencies in feature models. Some of them have concentrated on the conversion of feature models to logical representations to benefit from the reasoning capabilities that exist in these domains [19], [5], [6], [8]. In [5], [6], feature models have been transformed into propositional logic and non-functional domain attributes have been modeled with fuzzy variables; and an semi-automated method uses propositional logic feature selection approach to check the satisfaction of domain's constraints. The characteristics of First Order Logics (FOL) and constraint programming have been applied to the analysis of feature models in [7]. In [8], Description Logics has been used to build an ontology for features and some external services. Later on, using ontology matching algorithms, the proper services can be bound to corresponding features. Then using a DL-based reasoning engine the consistency validation is applied on feature models. Most work that deal with inconsistent feature models have focused on validity checking or finding inconsistencies in feature models. There is the lack of approaches to propose solutions to resolve inconsistencies in feature models. The most important approach that suggests solutions for inconsistency resolution in feature models is [19]. Wang et al in [19] have extended an incremental algorithm, called SkyBlue. They propose a dynamic approach to prioritize constraints in a feedback control system that adjusts itself based the opinions of the domain analysts. In contrast to [19], we have transformed feature models into Description Logic (OWL-DL) to utilize the strength of DL reasoning approaches for solving inconsistency in feature models. Furthermore, by employing Reiter's algorithm the proper solution is provided to the domain analysts.

VII. Conclusion and Future Work

In this work, we strive to design and implement a framework to identify and resolve inconsistencies in feature models based on DL reasoning. In our approach, feature models can be automatically converted into OWL-DL in order to perform reasoning and check the consistency of the models. We have seamlessly integrated Pellet as a plug-in into our AUFM tool to discover inconsistencies in the feature model. Based on the proposed solutions, domain analysts can simply remove or keep the inconsistent features. Our preliminary evaluations demonstrate that our framework can help a domain analyst model and correct the desired domain dynamically.

One of the main contributions of our work is that we provide a correspondence between feature model representations and Description Logic knowledge bases and benefit from formal DL reasoning for identifying and resolving inconsistencies in feature models. Given the clear semantics of DL revisions, we have been able to guarantee meaningful revisions in feature models through DL based semi-automatic inconsistency resolution algorithms. For future work, we believe that our tool support, AUFM, needs more evaluations in terms of scalability testing. It is required to examine what would be the performance of AUFM and basically DL based reasoning methods for handling a feature model with large number of features. In addition, based on the feedback we obtained from the case study participants, we find that the resolution process would be more efficient if AUFM provides recommended solutions in FM language rather than in pure DL. So, we also need to think about this matter and improve the usability of AUFM at this point.

Acknowledgments

The authors would like to acknowledge the work of Mostafa Karami Bekr and his contributions to AUFM.

References

[1] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.

[2] Ebrahim Bagheri, Mohsen Asadi, Dragan Gasevic, and Samaneh Soltani. Stratified analytic hierarchy process: Prioritization and selection of software features. In *The 14th International Software Product Line Conference*. Springer, 2010.

[3] Ebrahim Bagheri and Faezeh Ensan. Evidential reasoning for the treatment of incoherent terminologies. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, SAC '10, pages 1381–1387, New York, NY, USA, 2010. ACM.

[4] Ebrahim Bagheri and Dragan Gasevic. Assessing the maintainability of software product line feature models using structural metrics. In *Software Quality Journal*. Springer, 2010.

[5] Ebrahim Bagheri, Tommaso Di Noia, Dragan Gasevic, and Azzurra Ragone. Formalizing interactive staged feature model configuration. In *Journal of Software Maintenance and Evolution: Research and Practice*. Wiley, 2010.

[6] Ebrahim Bagheri, Tommaso Di Noia, Azzurra Ragone, and Dragan Gasevic. Configuring software product line feature models based on stakeholders' soft and hard requirements. In *The 14th International Software Product Line Conference*. Springer, 2010.

[7] D. Benavides, Ruiz A. Cortés, and P. Trinidad. Using Constraint Programming to Reason on Feature Models. In *The Seventeenth International Conference on Software Engineering and Knowledge Engineering (SEKE'05)*, July 2005.

[8] Marko Boskovic, Ebrahim Bagheri, Dragan Gasevic, Bardia Mohabbati, Nima Kaviani, and Marek Hatala. Automated staged configuration with semantic web technologies. In *International Journal of Software Engineering and Knowledge Engineering*, volume 20, pages 459–484. World Scientific, 2010.

[9] Volker Haarslev and Ralf Möller. Racer system description. In *Proceedings of the First International Joint Conference on Automated Reasoning*, IJCAR '01, pages 701–706, London, UK, 2001. Springer-Verlag.

[10] K. Kang, S. Cohen, J. Hess, W. Nowak, and S. Peterson. *Feature-Oriented Domain Analysis (FODA) Feasibility Study*. 1990.

[11] Kyo C. Kang, Jaejoon Lee, and Patrick Donohoe. Feature-oriented product line engineering. *IEEE Software*, 19:58–65, 2002.

[12] Klaus Pohl, Günter Böckle, and Frank J. van der Linden. *Software Product Line Engineering: Foundations, Principles and Techniques*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[13] R Reiter. A theory of diagnosis from first principles. *Artif. Intell.*, 32:57–95, April 1987.

[14] Stefan Schlobach, Zhisheng Huang, Ronald Cornet, and Frank van Harmelen. Debugging incoherent terminologies. *Journal of Automated Reasoning*, 39:317–349, 2007. 10.1007/s10817-007-9076-z.

[15] Richard W. Selby. Enabling reuse-based software development of large-scale systems. *IEEE Transactions on Software Engineering*, 31:495–510, 2005.

[16] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. Pellet: A practical owl-dl reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):51 – 53, 2007. Software Engineering and the Semantic Web.

[17] D. Tsarkov and I. Horrocks. FaCT++ description logic reasoner: System description. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4130 LNAI:292–297, 2006.

[18] T. von der Maen and H. Lichter. Deficiencies in feature models. In *Proceedings of SPLC'04 Workshop on Software Variability Management for Product Derivation*, 2004.

[19] Bo Wang, Yingfei Xiong, Zhenjiang Hu, Haiyan Zhao, Wei Zhang, and Hong Mei. A dynamic-priority based approach to fixing inconsistent feature models. In *Model Driven Engineering Languages and Systems*, volume 6394, pages 181–195. Springer, 2010.

[20] Hai H. Wang, Yuan Fang Li, Jing Sun, Hongyu Zhang, and Jeff Pan. Verifying feature models using owl. *Web Semant.*, 5:117–129, June 2007.

[21] Jules White, Brian Dougherty, Douglas C. Schmidt, and David Benavides. Automated reasoning for multi-step feature model configuration problems. In *SPLC*, pages 11–20, 2009.

A Management Framework for Context-Aware Multimedia Services

Ichiro Satoh
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
ichiro@nii.ac.jp

Abstract

This paper presents a framework for building user interfaces for context-aware service management. It enables us to easily deploy context-aware services at computers through user-friendly manipulations to edit documents and monitor changes and services in the real world. It is constructed as a compound document framework whose the structural composition of visual components can be adapted to changes in the real world. Since components in the framework are programmable entities, they can directly monitor and control their target devices. This paper presents the design for the framework and describes its implementation and several practical applications with it in the real world.

1 Introduction

Ambient intelligence provides highly proactive or interactive environments that use embedded computation to observe and participate in activities that have never previously involved computation. Computers in such environments remain invisible to users and system services are obtained by means of context-awareness. Therefore, rooms, offices, classroom and homes should be provided with their own entities and improve the quality of life of their inhabitants, helping them in their daily tasks. Moreover, services must be adapted to tasks, the environment, its occupants, and available resources.

Many researchers have explored context-aware services as one of the most typical applications of ubiquitous/pervasive computing. However, They have paid scant attention to the management tasks of context-aware services. For example, a context-aware system consists of many heterogeneous computers and sensing devices connected through wired or wireless networks in a house or office. The requirements of applications in such environments tend to depend on their targets, e.g., users, houses, or offices. To support context-aware services, their management systems need to know the context and process this in the real world, e.g., in terms of people, locations, and time. Furthermore, some existing context-aware services contain multimedia content, such as text, images, and video. However, existing context-aware services systems lack professional administrators unlike other network systems. Therefore, end-users themselves are required to customize their own context-aware services environments to their individual requirements and applications.

We propose a component framework to rapidly and easily build graphical user interfaces (GUIs) for management systems in context-aware services. Our framework not only provides GUIs that enable (non-professional) administrators to manage context-aware services and sensing or computing devices in ubiquitous/pervasive computing systems but also provides GUIs for enabling end-users to use the services. The framework itself is constructed as a distributed multimedia system. For example, some context-aware services support video cameras for computer vision or security purposes.

2 Example Scenario

Suppose a context-aware visitor-guide system is installed in a museum. Most visitors to museums lack sufficient knowledge about the exhibits there and they need annotations on these. However, as their knowledge and experiences are varied, they may become puzzled (or bored) if the annotations provided to them are beyond (or beneath) their knowledge or interest. User-aware multimedia annotation services, including text, images, and video, about exhibits are required. For example, when a user stands in front of an exhibit, a multimedia annotation service about the exhibit is provided in his/her personalized form on a stationary terminal close to the exhibit. Museum curators, who have no professional knowledge about context-aware services, deploy and customize context-aware multimedia annotation services and test the services. As exhibitions are often changed in museums, even curators should be able to easily and naturally configure the systems, e.g., the topology of sensor network and filter the data measured by these sensors.

3 Design Principles

There have been a few attempts to enable end-users to easily manage their networks. Nevertheless, networks for context-aware services environments in houses and offices are usually administered by end-users, who may have no knowledge or experience with networks. User-friendly interfaces are therefore needed so that end-users can easily manage networks in their context-aware services environments. Several commercial or academic systems for network management have provided visual interfaces for professional administrators rather than end-users. Furthermore, they have explicitly or implicitly assumed that they were being used without any other network management systems. That is, their visual interfaces have not been able to coexist with those of other systems. We need to be able to seamlessly unify visual interfaces for different network management systems.

Although commercial or academic Web-based tools for network management have recently been used, they cannot always monitor and control their target network systems in a real-time manner. This is because they periodically query the target systems and update their visual interfaces displayed within Web browsers executed on client-side computing devices through http-based protocols. Ajax technology may be able to reduce latency between the target systems and the visual interfaces, but most network devices or sub-systems do not support the technology, because they only support simple or particular protocols.

Administrators or users must support heterogeneous network devices, sub-systems, or services that are different. Our framework needs to be independent of any network management systems and open to various network management protocols. Networks in context-aware service environments are evolving in the sense that network devices and services are dynamically being added to and removed from them. Therefore, a user-friendly management system supports the dynamic evolution of network systems or context-aware systems.

4 Compound Document Framework

This section presents a component framework for building and operating the visual interfaces for network management systems by using compound document technology.

4.1 Basic approach

The framework is constructed based on a compound-document framework, called *MobiDoc*, developed by the author [7]. It enables one document to be composed of various visible parts, such as text, images, and video created by different applications, like other compound-document frameworks, e.g., COM/OLE [1] and Bonobo [4]. Compound-document technology is useful for constructing visual interfaces for network management, because it enables end-users to easily and dynamically assemble visual components into a seamless interface.

Like other compound document frameworks, the framework presented in this paper enables components to maintain their own content within them and to be dynamically assembled into one document or component. It also supports GUI-based manipulations enabling it to edit individual components and to layout components on GUI windows or control panels so that end-users can create GUIs for their networks. Unlike other existing frameworks, e.g., COM/OLE, OpenDoc, CommonPoint, and Bonobo, it provides each component with its own program code enabling it to view and edit its content within the component. Therefore, such a component itself can implement network management protocols to communicate with its target network device, sub-system, and service. For example, when a user wants to manage a new network device in his/her networks, he/she drags and drops the visual component that can define the network protocol to monitor and control the device as well as the visual interfaces on his/her control panel.

Although the framework inherits many features of our previous compound document framework, `MobiDoc`, it has been extended to support network management. It consists of two parts: a visual component and a component runtime system. The former is defined as a collection of Java objects and the latter is executed on the Java virtual machine (VM). Since the Java VM and libraries abstract away differences between underlying systems, e.g., operating systems and hardware, components and runtime systems can be executed on different computers, whose underlying systems may be different.

4.2 Component runtime system

Each runtime system governs all the components within it and provides them with APIs for the components in addition to Java's classes. It assigns one or more threads to each component and interrupts them before the component terminates, or is saved. Each component can request its current runtime system to terminate and save itself and its inner components in secondary storage. This framework provides each component with a wrapper, called a *component tree node*. Each node contains its target component, its attributes, and its containment relationship and provides interfaces between its component and the runtime system (Fig. 1). When a component is created in a runtime system, it creates a component tree node for the newly created component. When a component migrates to another location or duplicates itself, the runtime system migrates its node with the component and makes a replica of the whole node.

A hierarchy is maintained in the form of a tree structure of component tree nodes of the components (Fig. 2). Each node is defined as a subclass of `MDContainer` or `MDComponent`, where the first supports components, which can contain more than one component inside them and the second supports components, which cannot contain any components. For example, when a component has two other components inside it, the nodes that contains these two inner components are attached to the node that wraps the container component. Component migration in a tree only occurs as a transformation of the subtree structure of the hierarchy. When a component is moved over a network, on the other hand, the runtime system marshals the node of the component, including the nodes of its children, into a bit-stream and transmits the component and its children and the marshalled component to the destination.

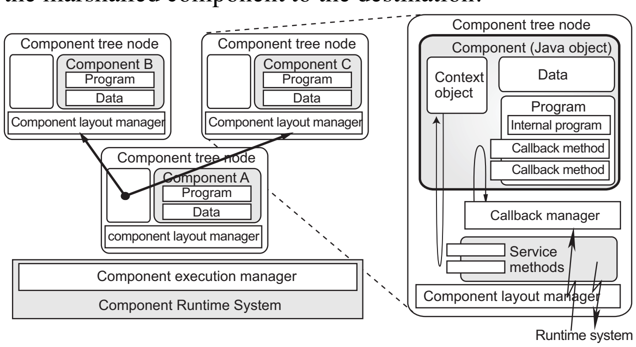


Figure 1. Component hierarchy and structure of components.

4.3 Visual component

As we can see from Fig. 1, each component is a collection of Java objects wrapped in a component and has its own unique identifier and image data displayed as its icon. All the objects that each component consists of need to implement the `java.io.Serializable` interface, because they must be marshaled using Java's serialization mechanism. Each visual component needs to be defined as a subclass of either the `java.awt.Component` or `java.awt.Container` from which most of Java's visual or GUI objects are derived. To enable existing software to be reused, we implemented an adapter to use typical Java components, e.g., Java Applets and JavaBeans, which are defined as subclasses of the `java.awt.Component` or `java.awt.Container` class within our components[1] Most Java Swing and AWT GUI Widgets can be used as our components in the framework without having to make any modifications, because they have been derived from the two classes.

[1]This is not compatible with all kinds of Applets and JavaBeans, because some of these existing applications manage their threads and input and output devices depreciatively.

The runtime system can invoke specified callback methods defined in components when they are created, relocated, and terminated and it can assign more than one active thread to the components. We can define program codes to communicate with network systems or devices within these callback methods. In fact, we constructed several components for network management through basic protocols. For example, an HTTP server (or client) component plays the role of an HTTP server (or client) to monitor and control network devices as HTTP clients (or servers) and a Telnet component can connect to its target device through a telnet protocol. Since these components are defined as abstract classes, we can define visual interfaces for network management by using Java Swing or AWT Widgets. We also developed various components, e.g., a text viewer/editor component and a JPEG, GIF, and MPEG viewer component and an audio-player component. Note that visual components allow their content to be in arbitrary as well as standard formats, because they have codes for viewing and modifying content. Components can support further application-specific protocols. For example, the Video Stream Player component in Fig. 2 supported a Real-Time Protocol (RTP) to receive a video stream and displayed the stream on its visual rectangle with a GUI control panel to stop, play, forward, back and pause the stream.

4.4 Component manipulation

Each component can display its content within the rectangular estate maintained by its container component. The node of the component, which is defined as a subclass of the `MDContainer` or `MDComponent` class, specifies attributes, e.g., its minimum size and preferable size, and the maximum size of the visible estate of its component in the estate is controlled by the node of its container component. These classes can define their new layout manager as subclasses of the `java.awt.LayoutManager` class.

This framework provides an editing environment for manipulating the components for network processing, as well as for visual components. It also provides in-place editing services similar to those provided by OpenDoc and OLE. It offers several value-added mechanisms for effectively sharing the visual estate of a container among embedded components and for coordinating their use of shared resources, such as keyboards, mice, and windows. Each component tree node can dispatch certain events to its components to notify them when certain actions happen within their surroundings. `MDContainer` and `MDComponent` classes support built-in GUIs for manipulating components. For ex-

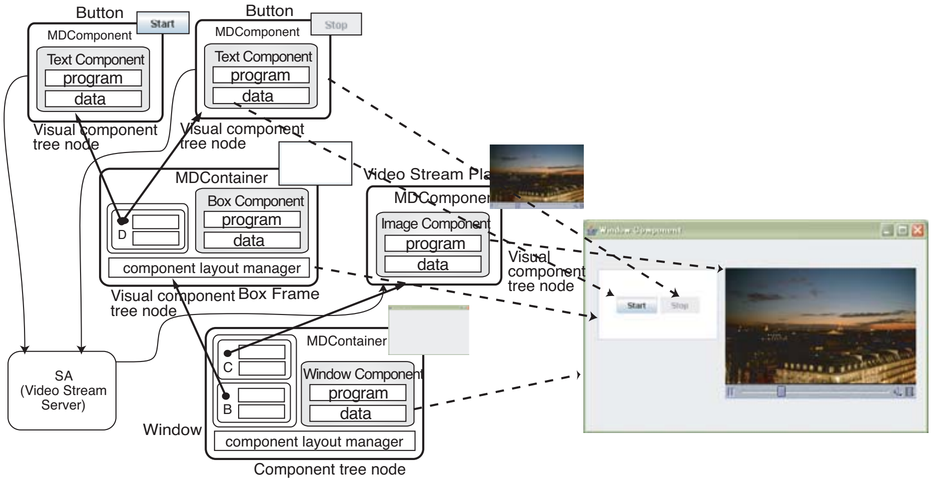


Figure 2. Component Hierarchy

ample, when we want to place a component on another component, including a document, we move the former component to the latter through GUI manipulations, e.g., drag-and-drop or cut-and-paste.

4.5 Component persistence

By using mobile agent technology, the runtime system can save or duplicate a component, its children, and information about their containment relationships and visual layouts into a bit-stream and can then later unmarshal the components and information from the bit-stream. When a component is saved or duplicated, its code and state, e.g., instance variables, can be marshalled by using the Java object serialization package. The package does not support the capturing of stack frames of threads. Consequently, our system cannot marshal the execution states of any thread objects. Instead, the runtime system (and the Java VM) propagates certain events to components before and after marshalling and unmarshalling them. The current implementation of our system uses the standard JAR file format for passing components that can support digital signatures, allowing for authentication.

5 Early Experience

We developed various components for monitoring and controlling computing devices, appliances, and sensors, in addition to basic visual components, e.g., text viewer/editor components, JPEG or GIF viewer components, and stream-video player components. Most Java Swing and AWT GUI Widgets can be used as our components in the framework without modifications.

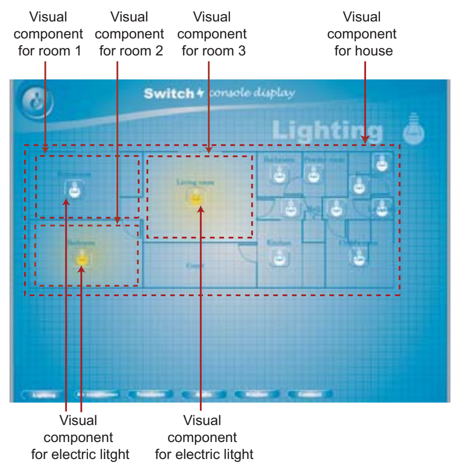


Figure 3. Screenshot of remote control interface.

5.1 Remote controller for home appliances

The first example is a remote controller for the power outlets of lights through a commercial protocol called X10. The lights are controlled by switching their power sources on or off according to the X10-protocol. We provided all lights with their visual components to switch them on or off. Each component communicates with an X10-base server, which controls an X10-module connected to the power outlet to switch the outlet on or off, and each displays its own visual interface to turn the outlet on or off as shown in Fig. 4. As we can see from Fig. 3, the component corresponding to the house contains the components corresponding to the rooms in the containment relationship between these physical spaces and entities. This system is connected to a smart meter, called TED 5000C, and we provide a component for display the power consumption measured by the meter. Since it treats the amount of power consumption as context, it can control appliances according to the amount of power they consume.

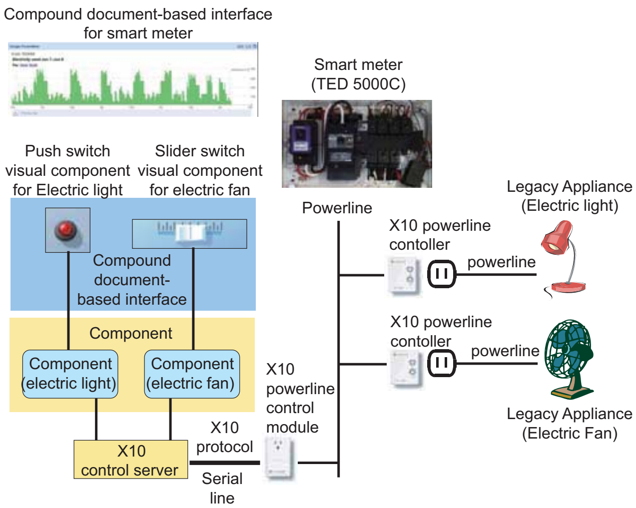


Figure 4. X10-based power-outlet controlling system

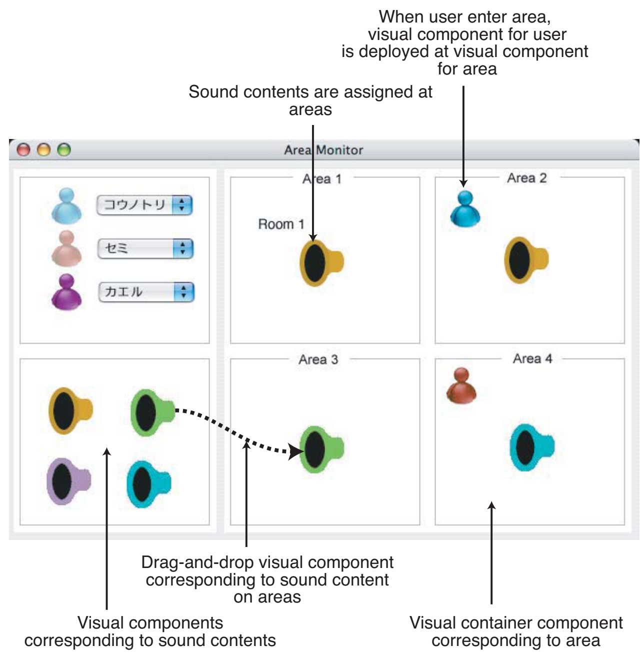


Figure 5. Screenshot of monitor system window for user/location-aware visitor assistance system with sensor network

5.2 Management system for context-aware services

The framework has already been used in a management system for context-aware user-assistant services in public museums, e.g., at the National Science Museum in Tokyo and the Museum of Nature and Human Activities in Hyogo, Japan. The system was constructed as a sensor network where RFID tag readers were connected through a wireless network. These readers were located at specified spots in several exhibition spaces at these museums. Visitors were provided with active RFID-tags to track their locations. When they came sufficiently close to various exhibits, e.g., zoological specimens and fossils, located at the spots, they could listen to multimedia content that provided annotation about the exhibits.

The RFID-tag readers identified all the visitors within their range of coverage, i.e., a 2-meter diameter and sent the identifiers of their detecting RFID tags to a service-provider computer through TCP sessions. All service-provider computers had databases for storing multimedia content and they selected and played content according to their visitor's knowledge and interests when they received the identifiers of the tags from the readers. Fig. 5 shows a screenshot of the visual interface for the management system. The interface enables users to deploy services at areas by using drag-and-drop manipulations. For example, the exhibition had more than 200 visitors daily and the system continued to monitor and manage RFID-tag readers and location-aware services for a week without experiencing any problems.

The interface consisted of four visual components that monitored four RFID-tag readers located at spots throughout the exhibition consisting of four spots. As we can see from Fig. 5, the window was implemented as a window component that contains six components corresponding to the frame boxes in it. Four of the box-frame components represented the spots and had programs that communicated with their readers through TCP sessions to monitor the presence of tags within their coverage areas, where RFID readers could work as TCP servers to send the identifiers of such tags. Fig. 6 shows the visual interface for user/location-aware systems used at the National Science Museum. An image view component drew a map of the exhibition room and it contained six management components for RFID readers connected through a network.

When a visitor with an RFID-tag entered a spot, the component corresponding to him or her was deployed at the component corresponding to the spot. We could dynamically add/remove location-aware services to/from spots. We deployed software to define the service at the component corresponding to a spot by dragging-and-dropping the visual component of the software on the visual components corresponding to the places in which the services should be provided.[2] Curators, who may have no knowledge about context-aware systems, can therefore easily and intuitively

[2]When more than one user enters a spot, the system selects and plays content according to the combination of people in the spot. The mechanism is presented in another paper [9].

change context-aware multimedia services at exhibitions.

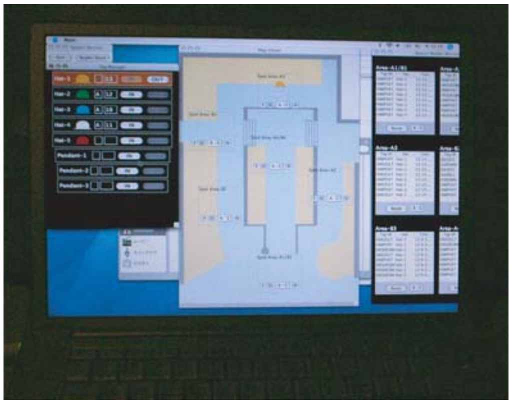


Figure 6. Monitoring computer for six spots at National Science Museum.

6 Related Work

Several projects have proposed scripting or markup languages for building and configuring GUIs [10]. However, they were intended to define static GUIs, whereas our framework supports dynamic GUIs in the sense that the structural composition of visual components is modified according to changes in the real world.

Several researchers have explored toolkits for context-aware services [3, 2], but most of them have aimed at building user interfaces on mobile devices, including smart phones. Other projects have focused on building context-aware systems but not on building user interfaces. There have been several mechanisms for automatically generating GUIs for controlling devices [6, 5]. Most existing approaches can provide GUIs individual devices and can support them, being dynamically generated for devices that may be added. They assume the target devices, which they should provide visual interfaces for, can support their own control protocols for visual interfaces. However, most network devices do not support such protocols. Instead, they can be controlled and monitored through their own favorite protocols, e.g., SMNP, HTTP, and Telnet, in addition to management-specific protocols. Therefore, the framework presented in this paper is required to support various protocols. Network devices and ubiquitous computing devices only have limited resources, such as restricted amounts of CPU power and memory. However, client-side devices, i.e., PCs and PDA, have sufficient resources with their input/output devices, e.g., displays, keyboards, and mice. Therefore, visual interfaces should be managed and executed as much as possible in external systems, including client-side devices, rather than in network devices.

7 Conclusion

We presented a component framework for rapidly building and operating visual interfaces for network management. The framework can dynamically assemble visual components into a visual interface. It enables components to communicate with their target network sub-systems through the sub-systems' favorite protocols, since these components contain their own program code in addition to the content inside them. Since it provides GUI-based manipulations for editing visual components, end-users can easily add and remove the visual interfaces for network management on their control panels. We designed and implemented a prototype-system based on the framework and demonstrated its effectiveness in a network management system for sensor networks in public museums.

Acknowledgments

This research was supported in part by a grant from the Promotion program for Reducing global Environmental loaD through ICT innovation (PREDICT) made by the Ministry of Internal Affairs and Communications in Japan.

References

[1] K. Brockschmidt, Inside OLE 2, Microsoft Press, 1995.

[2] T. Butter, M. Aleksy, P. Bostan, and M. Schader: Context-aware User Interface Framework for Mobile Applications In Proceedings of International Conference on Distributed Computing Systems Workshops 2007.

[3] A. Dey, D. Salber, and G. Abowd, A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications, Human-Computer Interaction (HCI) Journal, vol. 16, no. 2-4, pp. 97-166, October 2001.

[4] The GNOME Project, Bonobo, http://developer.gnome.org/ arch/component/ bonobo.html, 2002.

[5] T. D. Hodes, R. H. Katz, E. Servan-Schreiber, L. Rowe, Composable ad-hoc mobile services for universal interaction, Proceedings of International Conference on Mobile Computing and Networking (MobiCom'97), pp.1-12, 1997.

[6] J. Nichols, B. A. Myers, M. Higgins, J. Hughes, T. K. Harris, R. Rosenfeld, and M. Pignol, Generating remote control interfaces for complex appliances, Proceedings of Symposium on User Interface Software and Technology (UIST'02), pp.161-170, ACM Press, 2002.

[7] I. Satoh, A Component Framework for Document Centric Network Processing, IEEE International Conferences on Symposium on Applications and the Internet (SAINT'2007), IEEE Computer Society, January. 2007.

[8] I. Satoh, A Visual Framework for Deploying and Managing Context-aware Services, International Conference on Grid and Pervasive Computing (GPC'2007), Lecture Notes in Computer Science (LNCS), May, 2007.

[9] I. Satoh, Context-Aware Services for Groups of People, Proceedings of Pervasive and Embedded Computing and Communication Systems (PECCS'2011), pp.54-62, March, 2011.

[10] User Interface Modelling Language, http://www.uiml.org.

A Visual Approach supporting the Development of MicroApps on Mobile Phones

Stefania Cuccurullo, Rita Francese, Michele Risi, Genoveffa Tortora
University of Salerno
Fisciano (SA), Italy
{scuccurullo, francese, mrisi, tortora}@unisa.it

Abstract — **The definition of an approach supporting an End-User in the development of mobile applications is a hard task because of the characteristics and the limitations of mobile device interfaces. In this paper, we present a visual approach to enable End-Users to compose visually their own applications directly on their mobile phone. To this aim, we propose a touchable interface and an ad-hoc visual language, enabling the user to compose simple focused applications, named MicroApps. The user has not in charge the creation of the user interface that is automatically generated. Moreover, we present the results of a preliminary usability study that revealed a good satisfaction degree of all the involved subjects, whereas an empirical analysis highlighted that the MicroApp visual approach is effective and efficacy.**

Keywords- Visual languages; Mobile End-User Development; Mobile Applications

I. INTRODUCTION

Mobile phone applications are becoming more and more popular, and Gartner research expects a market volume that can reach $30 billion in 2013 [6]. Indeed, new services and innovative interaction modalities, including gesture detection, device movement and context-based control are continuously proposed [9]. These innovations are mainly due to the novel and cheap equipments offered by the last mobile phone generation, such as on-board camera, accelerometers, compass, GPS, etc., their increased processing power and fast internet connectivity.

However, the more smartphones become smarter, the more the design of complex applications covering the various user needs becomes a not easy task. Indeed, few of today's available applications take into account that mobile users have different preferences and employ the applications in various situations. In many cases, it is impossible to foresee all the context of use and the actual requirements are often difficult to describe.

One possible solution is to provide to the End-User the means to easily compose and customize mobile applications, starting from a set of available simpler features. Some existing generation tools create the application on the Personal Computer (PC) and successively download it on the mobile device [4]. In addition, as in [1], the proposed composition approaches often requires the user to compose the application user interface.

The main idea is to avoid writing the source code. In particular, a mobile application requires the knowledge of different technologies, such as a particular programming language (i.e., Java or Objective-C), different software development kits related to different mobile devices (i.e., Android, iPhone and so on). This involves a high learning curve with a subsequent high effort for End-Users or novice programmers. Moreover, several mobile applications can be modeled as a composition of pre-existing applications/ services available on the different mobile devices. In fact, an appropriate handling of these services allows the user to define by himself more complex applications meeting different needs. To this aim, the composition of these applications can be visually modeled through graphical symbols, associated to a particular application behavior and to a specific user interface. By opportunely connecting these graphical symbols, the user can describe complex behaviors.

This paper aims at supporting an End-User in the creation of focused mobile applications, called MicroApps [3]. A naive End-User generates and uses a MicroApp directly on the mobile phone. A MicroApp is designed by graphically composing the functionalities offered by the various phone applications, such as taking an image from the *Camera* object and saving it (i.e., the *Camera.Take* and *Image.Save* actions), retrieving the contacts list from the *Contacts* object (i.e., *Contacts.List* action), and sending an email using the *Mail* object (i.e., *Mail.Send* action), etc. Each action exposes a description of its user interface that enables to generate automatically the MicroApp user interface. Moreover, the proposed tool assists the user in the composition of the functionalities provided by the mobile device. The result is a specific application based on the user needs that customizes the usage of the mobile device. Finally, a workflow engine enacts the composed process when the application is invoked. In particular, this engine, named MicroApp Engine, is a suitable mobile application that manages and executes modeled MicroApp specifications.

This paper is structured as follows: Section 2 discusses the related work, Section 3 presents the proposed approach for the generation of mobile applications, while Section 4 provides the results of a preliminary usability study to assess the proposed approach. Finally, Section 5 concludes the paper with future works.

II. RELATED WORKS

Mobile End-User development is at the beginning phase and presents new issues mainly due to the characteristics and the limitations of mobile device interfaces. The research interest towards this topic is mainly due to the growing preference revealed by the user towards the services offered by these devices and the need of customizing their applications [4].

In [7] and [9] an approach has been proposed supporting the user in the definition of context-action rules aiming at activating mobile phone functions when the rule conditions are satisfied. Differently from them, we adopt a visual interface to compose a MicroApp that can be more complex than the pattern event-action.

Jigsaw programming has been largely investigated in literature. In this kind of approach, program constructs are represented using icons that look like jigsaw pieces, and only icons that fit together can be composed to form legal programs, see as an example [2], [5] and [12]. More recently, Google has proposed AppInventor [1], that adopts a block editor to create simple programs on the PC that should be downloaded on the mobile device. This editor enables the user to program by using the OpenBlocks programming language [13]. Blocks enable the user to program repeating actions, conditions, information storing, etc. The approach we propose is very similar. It does not require the user to compose the interface and the PC usage, because MicroApps are directly created on the mobile device, and provides a form of assistance in the composition considering the compatibility of the input/output of the various actions.

Authoring tools have also been proposed to compose user-generated mobile services. In [4] Microservices composed by the user can be shared and downloaded. Microservices are created considering two user expertise levels: beginners, enabling a template-based development of a Microservice, and advanced, based on a XML-based language. We propose an approach that assists the user in the application composition, as better detailed in Section III.

An emerging End-User service composition technique is based on the Spreadsheet metaphor [8]: the process is modeled putting the service invocation in the cells. The HUSKY tool [14] enables the users to compose logic spatially arranging of component services within spreadsheet cells. Time progresses from left to right and from top to bottom in cell blocks. Thus, a set of adjacent cells makes a sequence of events. In this case, the information concerning the compatibility among input-output services is not graphically depicted.

III. THE PROPOSED APPROACH

In this paper, we focus on the creation of customized mobile applications, named MicroApp, that the user is able to compose directly on the mobile device [3]. The framework works into two main configurations: model and enactment. In the first configuration, the user composes the application using a Visual Editor. During the enactment configuration phase, a micro application is executed on the mobile device touching its icon or when a specific event occurs.

The proposed approach helps the users to manage the complexity of their activities performed with the mobile device by composing simple applications. The users do not concentrate in managing the dataflow and in the designing of the user interface, but only on the sequence of the actions needed to model the required MicroApp.

Composition approaches have been classified in three groups [15]: Control Flow-based, Data Flow-based and Assisted. The approach we propose is hybrid: on one side, the tool assists the composition by enabling the user to select an individual action from a wide range of actions available on the phone and, on the other side, it allows the user to compose the actions using an ad-hoc developed visual language. A Visual Editor and a Visual Language, adopting a Jigsaw-based programming approach, support the composition of a MicroApp.

In the rest of this section, we describe in detail the development process, the visual composition and the user interface generation of a MicroApp.

A. The Development Process

The overall process to develop and execute a new MicroApp mobile application is composed of three subsequent phases: MicroApp Definition, MicroApp Modeling and MicroApp Deployment, as shown in Figure 1, where the rounded rectangles represent process phases, whilst the rectangles represent the intermediate artifacts produced at the end of each phase.

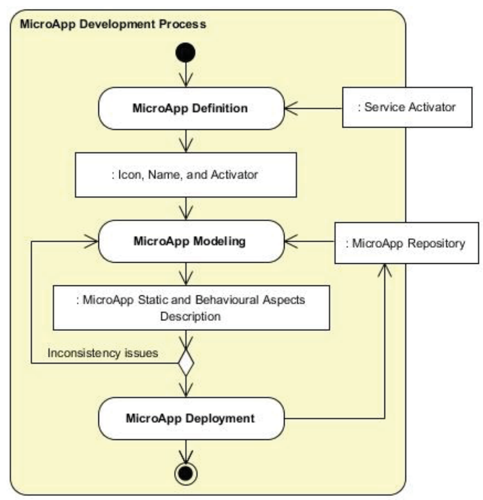


Figure 1. The MicroApp development process.

The first phase, *MicroApp Definition*, allows the user to describe the scaffolding of the MicroApp to be developed. First of all, this phase enables the user to choose how the MicroApp is represented on the device. In particular, the icon representing the application and the visualized name are provided. To complete this phase, a user has to choose an activation modality from the *Service Activator* list provided as input to the process. In particular, examples of activators are: application menu, user gestures, location, etc. The output of the first phase is an empty mobile MicroApp expressed using an XML notation. An example of MicroApp definition user interface is shown in Figure 2(a). In particular, the camera icon of the developing MicroApp has been selected to represent the application which has been entitled "Take and Send". The selected activation event is a circular finger gesture. When the device detects this gesture, the application is loaded and then executed.

The *MicroApp Modeling* phase defines the behavior of the mobile application. In this phase, the MicroApp components are taken from a repository and are composed connecting them, respecting the input/output parameter constraints. The output of this phase is a description of the static and dynamic aspects of the mobile application to be generated. The *MicroApp Deployment* phase stores the MicroApp description in the *MicroApp Repository* of the mobile device and registers it in the action list that are launched when the selected activation event happens.

B. *MicroApp Visual Editor and Language*

An ad-hoc developed mobile Visual Editor, designed considering the limited size of the device screen, supports the user in the modeling of the behavior of a MicroApp by composing its application logic.

The main idea is to eliminate all the textual components of a programming language, providing a suitable visual language easy to use and not restricted to simple functionalities. Users can select from a wide range of actions and do not have to define dataflow among them as these aspects are automatically managed. Indeed, to enable the user to appropriately compose a MicroApp, the selection of the actions is supported by an underlying computational algorithms that manage the compatibility of already selected actions.

(a) **(b)**

Figure 2. The MicroApp definition and the Visual Editor layout.

In Figure 2(b) a screenshot of the Visual Editor is shown. The main area of the screen is the *Composition Area*, where the MicroApp is composed by dragging and dropping the action icon of the selected application. This area is divided in columns to allow the user the sequential and parallel composition of actions. This editor supports only these types of composition rules avoiding more complicated structures as cycles, because one of the usability objective of the proposed approach is that the language has to be easy to use and to learn and that it has to enable the composition of simple applications.

The user adds a new action by pressing the *Add* button, represented by the "+" icon in the middle lower part of the Composition Area. Then, the editor opens a new window showing a list of the available applications. The list of all available applications is provided by the MicroApp Repository. When the user selects an application (e.g., *Camera*, *Contacts*, *Net* and so on), the editor shows the list of the actions available for that application. As an example for the Camera application the available actions are *Camera.Take*, *Camera.Preview*, etc.. The Visual Editor assists the selection of an action by highlighting the action icons whose input is compatible with the inputs/outputs of the blocks already positioned. Once selected the action, the user clicks on a specific column of the Composition Area to add the action. The editor still assists the user disabling the columns that are not compatible with the action to be inserted. However, an action can always be inserted in the first empty column on the right, highlighted differently from the others, as shown in Figure 2(b). In this case a new empty column is automatically added in the rightmost part of the editor.

The proposed approach allows the users to manipulate and connect the actions to build their mobile applications. Thus, the compatibility concept is related to the type and number of inputs provided (outputs received) by the application actions.

Graphically, an application action available on the mobile device is represented by a rounded square containing the application icon, such as *Camera* and *Mail*, and the name of the action, such as *Preview* and *Send*, respectively. The input/output parameters are represented by colored bullets. In particular, as shown in Figure 3, the input parameters are depicted in the higher part of the square, whereas the output parameters are shown on the bottom.

(a) **(b)** **(c)** **(d)**

Figure 3. Action block examples.

The parameters are differently colored, depending on the type of the corresponding object. As an example, the pink colored parameter represents an Image object, while the cyan colored parameter represents a Contact object, containing the contact data (e.g., name, surname, address, email, cellular phone and so on). Similarly, the red bullet represents a text string and, finally, the yellow bullet represents an email object.

Figure 3 shows some examples of application actions. In particular, in Figure 3(a) the *Preview* action takes as input an Image object, displays it and returns the same object as output. The generated user interface is described in the next section. In Figure 3(b) the red bullet parameter, corresponding to a text string, is used to fill the subject field of an email. Moreover, the circled bullet denotes a variable number of parameters of any type. The objects associated to these parameters will be used to compose the email body, and could be indifferently Image and/or Text objects. Once the mail is sent, it will be provided as output.

Let us note that in Figure 3(c) the *Contacts* action exposes a triangle parameter in the left hand side. This kind of parameter has to be assigned during the application composition, and the associated value will be fixed for each execution of this MicroApp. This means that at design time the user has to select a contact present in the contact list.

Figure 3(d) shows the *Take* action, that has no input parameter. The Image object provided as output is obtained by pressing the *Take* button at execution time. In addition, each input parameter type has associated a default action that is able to generate its value. The default action is automatically invoked in case the user at design time leaves it unassigned. It is worth noting that it can be hard for a novice user to remember the meaning of each color parameter. Thus, the editor assists the user providing a description of all the parameters when the user performs a long press on the action icon.

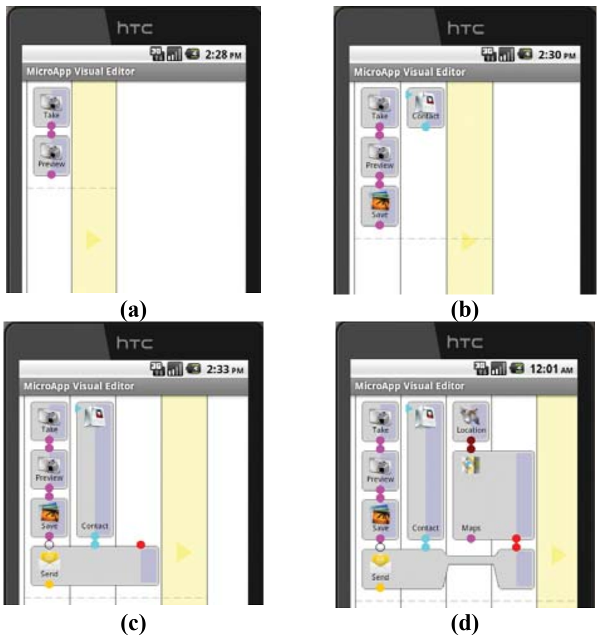


Figure 4. Example of composition rules.

Figure 4 shows some composition rules available to create a MicroApp. In particular, Figure 4(a) depicts an example of a successful sequential action flow. In fact, the output parameter of the topmost action is compatible with input parameter of the lower action. Figure 4(b) shows an example of a parallel action flow, whereas Figure 4(c) depicts an example of joint action flow. In particular, a new action is added to collect the outputs of the actions in the first and second columns. The user initially drags and drops the action on a particular column (i.e., the first column), and successively clicks on the other columns for associating the input parameters to the action, respecting the number and the type of the inputs. Note that if, as in the case of the second column, there is an empty space, the action *Contacts* is automatically lengthened. If the user provides no action in the third column, the text input will be provided by the default action associated to the text parameter (i.e., the *Text.Input* action), otherwise only an action with a textual output parameter should to be successively added.

Note that, the association between the action parameters is automatically resolved taking into account their compatibility.

In case an action does not need to use a specific parameter, the editor allows to connect the input parameters of an action with the output parameters of actions that are not in consecutive columns. This case is shown in Figure 4(d), where the Image object provided as output by the *Map.Maps* action is not provided to the *Mail* action. This Image object could be used by other subsequent actions. Let us note that the Visual Editor implements the *undo* operation to remove the last action or input/output parameter association.

It is important to point out that the composition can be performed selecting the blocks in different orders. Indeed, the user could first add the block *Mail.Send* and then position the other blocks. In particular, if there exists a block in the first row of the Composition Area that has at least an input parameter the Visual Editor automatically adds a new empty row by shifting vertically all the blocks by one position, as shown in Figure 5.

Figure 5. Example of assisted editing.

Once the user terminates the MicroApp modeling phase, he selects the *Deploy* command. The Visual Editor verifies the inconsistency issues, such as triangle parameters not specified and if all the circled bullet inputs have associated an output parameter. In case of an inconsistency problem, the editor provides the appropriate warning messages; otherwise, if the verification phase terminates successfully, an XML description of the composition process is stored in the MicroApp Repository, ready to be enacted by the MicroApp Engine.

The translation of the MicroApp composition process into the XML description is performed by analyzing the actions in the Composition Area from top to bottom and from left to right directions. This allows to linearize the execution of the modeled MicroApp by sequencing the execution of all the actions. During this step, the forward and backward communication messages between the actions are checked and implemented. When the MicroApp Engine loads the XML description, it translates the description into an execution sequence by instantiating the action objects, and then running the process over them. The control flow is driven by a mechanism that synchronizes the execution of the actions.

Finally, after the end of the deployment phase, the application icon appears in the main application menu of the mobile device. In case a different Service Activator has been selected in the definition phase, the application is also executed when the selected activation event occurs.

C. The MicroApp User Interface

The composition of the user interface is not an easy task for a non programming user. Indeed, there is the need of model the user interaction in terms of GUI elements, such as windows, pull-down menus, buttons, scroll bars, iconic images, wizards, etc.

In this paper we partially follow the approach proposed in [10] for service composition. In particular, each application action is annotated by a user interface. The MicroApp user interface is automatically generated by combining the annotated actions presentation frontends. When generating a MicroApp, the interface is created following the Microsoft PowerPoint presentation approach: each action is presented as a slide.

(a) **(b)**

Figure 6. The user interface of the *Camera.Take* and *Mail.Send* actions.

As an example, Figure 6(a) shows the interface generated for the *Camera.Take* action. In particular, the user is able to get an image, touching the *Take* button, or he can go *Back* in the control flow or eventually *Exit* the MicroApp by pressing the appropriate buttons. In practice, the *Back* button corresponds to an *undo* operation. Similarly, Figure 6(b) shows the user interface for the *Mail.Send* action. In particular, a preview of the mail is provided to the user that sends the mail by touching the *Send* button. The *Back* and *Exit* buttons have the same behavior as those of *Camera.Take*. In particular, in case the user touches the *Back* button the process is moved a step back to the execution of *Text.Input* action.

In case an error occurs during the execution of an action (e.g., no Internet connection is available), an error message alerts the user providing the cause that generated the problem, and successively the MicroApp Engine asks the user to perform an *undo* operation in order to repeat the action again or to exit the application.

The user interface associated to each action is specified using an XML description, that is automatically managed by the MicroApp Engine. As an example, Figure 7 shows the XML description used by the *Camera.Take* action to manage the layout of its user interface. In particular, each user interface component is specified by an XML tag that represents the graphical widget identified by a unique id. In fact, the *TextView* tag is used to inject the title of the performed action (i.e., "Camera Take"), the *FrameLayout* tag is used to visualize interactively the camera preview, and finally the *Button* tags are properly customized to implement the *Take*, *Back* and *Exit* behaviors.

However, the XML description in Figure 7 is also used by other MicroApp actions, such as *Camera.Preview*. Moreover, it is worth noting, that some actions do not require a user interface. As an example, the *Image.Save* action that is executed in background, directly stores the Image object taken as input in the image gallery of the mobile device, and provides the same Image object as output parameter.

```xml
<?xml version="1.0" encoding="utf-8"?>
<LinearLayout xmlns:android="http://schemas.android.com/apk/res/android"
   android:id="@+id/layout"
   android:orientation="vertical">

   <TextView android:id="@+id/title" />

   <FrameLayout android:id="@+id/surface" />

   <LinearLayout android:orientation="horizontal">
       <Button android:id="@+id/button_1" />
       <Button android:id="@+id/button_2" />
       <Button android:id="@+id/button_3" />
   </LinearLayout>
</LinearLayout>
```

Figure 7. An XML description used by the Camera.Take action.

IV. A Preliminary Evaluation

The proposed approach has been implemented in a prototype supporting users in the composition of MicroApps. In particular, the application is running on an Android based HTC device, by using the SDK version 2.2.

We conducted a preliminary usability study to assess the effectiveness and efficacy of the MicroApp development tool for End-Users during the composition of mobile applications. The satisfaction of the users has been also investigated here. To this aim, the presented study adopts the combination of two techniques: a questionnaire-based survey and an empirical analysis. The context of the user study was constituted of Bachelor students in Communication Science at the University of Salerno. In particular, data for the study have been gathered considering a group of ten volunteers. All the involved subjects had not procedural or object-oriented programming experience, but they are practical with touch mobile phones.

The study has been divided in three steps and performed in one-to-one session (i.e., a supervisor for each subject) using the think aloud technique. In the first step, a lesson of 15 minutes introduced to all the subjects the principles of editing MicroApps and the main features of the prototype. In the second step, we asked to the subjects to perform a task. The task concerned the creation of a MicroApp that exploited the Image, GPS, Calendar and Camera objects posting to a preselected Facebook profile information that states where the user is currently located, together with a photo and the current date. The application uses the camera actions that allow the user to take and save a photo, and successively enables the user to post the information to own Facebook personal area. In particular, the application posts the photo on the user wall, together with the GPS and Calendar data. In the third step, the subjects have to fill in a post experiment questionnaire to collect information on their satisfaction.

During the experiment the supervisor did not provided any help to the subjects to avoid biasing the experiment. He only wrote the comments and problems of the subjects, when they spoke aloud. For each subject the needed time to accomplish the experiment was annotated as well.

The questionnaire we used in this study contained 14 questions, shown in Table I, arranged in three categories: subject experience, MicroApp editing satisfaction, performed task. All the questions expected closed answers according to a five point Likert scale [11]. In particular, the values range from 1 (positive) to 5 (negative).

TABLE I. THE QUESTIONNAIRE-BASED SURVEY

Id	Question
q1	How do you judge your programming experience?
q2	How do you judge your touch mobile experience?
q3	The time to learn a functionality of the prototype is appropriate.
q4	The prototype functionalities were easy to use?
q5	The user interface is pleasant and easy to understand.
q6	The prototype commands are easy to understand and enact.
q7	Did you need any external help during the experimentation?
q8	I would use the MicroApp prototype in the future.
q9	Was the task simple?
q10	The number of operations to accomplish the task were appropriated
q11	Did you have any problems to find the needed objects?
q12	Was the connection between the objects and the parameters easy to perform?
q13	The messages and the warnings were easy to understand.
q14	How hard is to perform the task without using the MicroApp?

For the empirical analysis we have considered two variables: *(i)* the number of minutes required to perform the task, *(ii)* the number of *undo* operations performed by the user during the editing of the MicroApp.

In order to represent graphically the distribution of the size measures of the considered usability study, we have adopted the boxplots. This kind of diagrams is widely employed in exploratory data analysis since they provide a quick visual representation to summarize the data using: the median, upper and lower quartiles, minimum and maximum values, and outliers. In particular the answers are visually summarized in Figure 8, whereas Figure 9(a) shows the time performance results and Figure 9(b) the number of performed *undo* operations.

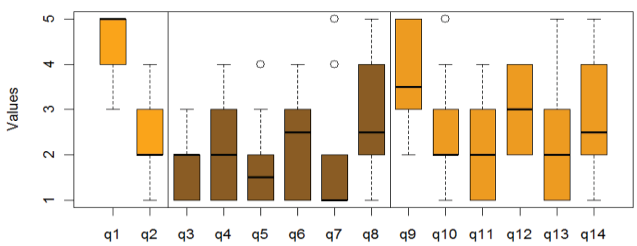


Figure 8. Boxplots of the questionnaire answers.

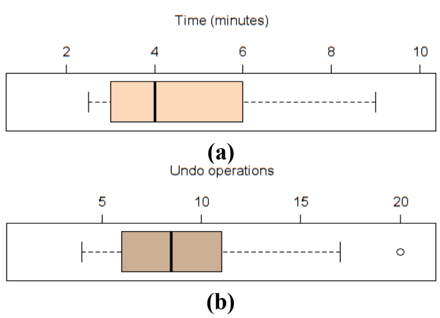


Figure 9. Boxplots of the empyrical analysis.

All the involved subjects finished the task, but two subjects generated a wrong MicroApp. The survey shows a good satisfaction degree of all the involved subjects (i.e., question from q3 to q8), even if they consider the task not simple to perform (i.e., the question q9), but hard to complete without using the prototype (i.e., question q14). Moreover, the empirical analysis reveals that the MicroApp visual approach effectiveness and efficacy is appropriate. In fact, 4 minutes is the mean time to accomplish the task performing in average less than 10 *undo* operations.

V. CONCLUSION

In this paper we presented a mobile application and a development process that support the user in the visual composition of customized applications for mobile devices. The proposed Visual Editor was designed for naive users and does not require the user involvement in the specification of the user interface. Moreover, the results of a preliminary usability study revealed a good satisfaction degree and the effectiveness and efficacy of the MicroApp visual approach.

In the future, we plan to improve the usability evaluation of the proposed approach, also comparing it with other existing similar ones, like AppInventor [1]. We are also investigating how the proposed methodology can be extended to other mobile platforms and support Service Oriented Architecture.

REFERENCES

[1] AppInventor, Google, http://appinventor.googlelabs.com/about/ retrieved on November 2010.

[2] J. Bonar and R. Cunningham, "Bridge: Tutoring the programming process", In Intelligent Tutoring Systems: Lessons Learned, J. Psotka, L. D. Massey, and S. A. Mutter, Eds. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1988, pp. 409-434.

[3] S. Cuccurullo, R. Francese, M. Risi, G. Tortora, "MicroApps Development on Mobile Phones", in Proceedings of 3rd International Symposium on End-User Development, Brindisi, Italy, M.F. Costabile et al. (Eds.), LNCS 6654, 2011, pp. 289-294.

[4] J. Danado, M. Davies, P. Ricca, A. Fensel, "An Authoring Tool for User Generated Mobile Services". Lecture Notes in Computer Science, Volume 6369/2010, 2010 , pp. 118-127.

[5] E. P. Glinert, "Out of Flatland: Towards 3-D Visual Programming". In Proceedings of FJCC '87-Fall Joint Computer Conference. IEEE Computer Society. Dallas, Texas, 1987, pp. 292-299.

[6] J. Grant, $6bn bonanza for mobile apps in 2010, says Gartner, 2010, http://www.computerweekly.com/Articles/2010/01/18/239992/6bn-bonanza-for-mobile-apps-in-2010-says-Gartner.htm

[7] J. Häkkilä, P. Korpipää, S. Ronkainen, U. Tuomela, "Interaction and End-User Programming with a Context-Aware Mobile Application", in Proceedings of Interact 2005, pp. 927-937.

[8] D. D. Hoang, H. Paik, B. Benatallah, B., "An Analysis of Spreadsheet-Based Services Mashup", in Proceedings of 21st Australasian Database Conference (ADC), Brisbane, Australia, 2010.

[9] P. Korpipää, J. Häkkilä, E.-J. Malm, T. Rantakokko, V. Kyllönen, J. Kela, I. Känsälä, J. Mäntyjärvi, "Customizing User Interaction in Smart Phones", IEEE Pervasive Computing: Mobile and Ubiquitous Systems, 5 (3), 2006, pp. 82-90.

[10] T. Nestler, M. Feldmann, A. Preuÿner, A. Schill, "Service Composition at the Presentation Layer using Web Service Annotations", First International Workshop on Lightweight Integration on the Web, 2009, pp. 63-68.

[11] A. N. Oppenheim, "Questionnaire Design, Interviewing and Attitude Measurement", Pinter Publishers, 1992.

[12] N. Osawa, "Jigsaw-Puzzle-Like 3D Diagrams to Represent Syntactical Constraints", Sixth International Conference on Information Visualisation , 2002, pp. 421-426.

[13] R. Roque, "OpenBlocks : an extendable framework for graphical block programming systems", Master Thesis Massachusetts Institute of Technology, 2007. http://dspace.mit.edu/handle/1721.1/41550.

[14] D. Skrobo, "HUSKY: A Spreadsheet for End-User Service Composition", 2011. http://www.fer.hr/_download/repository/DanielSkrobo_KvalifDrIspit_HUSKY.pdf.

[15] U. Wajid, A. Namoune, N. Mehandjiev, "A Comparison of Three Service Composition Approaches for End-Users", in Proceedings of the International Conference on Advanced Visual Interfaces, 2010.

3D Virtual-Reality Object of Noh-Mask for Mobile Devices

Billy Pham*, Hisato Kobayashi**, Kazuo Yana**, Momori Ban**, and Yoshiyuki Hino**

*Hosei Univesity Research Institute, California
800 Airport Blvd. Suite 504
Burlingame, CA 94010, U.S.A.

**Hosei Univesity IT Research Center, Tokyo 102-8160, Japan
Correspondence: bpham@huric.org

Abstract:

Recently 3D technology has been applied to various fields, including higher education. Many universities have created 3D virtual environments to provide effective interactive teaching materials and allow students to interact virtually in real time. At the same time, mobile devices have rapidly improved, giving users convenient yet powerful media for communication. This paper describes a method for creating Noh-masks, which are used for traditional Japanese performances, as 3-D virtual objects viewable either on a standard web browser or mobile devices such as iPad or Android terminals. These virtual Noh-masks have significant cultural values. The techniques described below are useful for creating 3D virtual replicas of similar valuable assets that can be placed in virtual museums.

Keywords: *3D virtual reality object, Objective-C, OpenGL ES for mobile devices.*

I. Introduction:

Hosei University has long been actively researching and collecting materials related to Noh Theater, and established the Noh Theater Research Institute in 1952. The collection holds over 40,000 items found mostly in Japan, including rare and extremely valuable manuscripts dating from the Muromachi and the Edo eras. The Institute's continuing research carries a meaningful mission to introduce Japanese culture to people around the world.

By 2015, 80% of people accessing the Internet will be doing so from mobile devices [1]. As the use of mobile technology increases, it gives educators around the world a new way to deliver lectures remotely. The potential of mobile computing is already being demonstrated in many projects at institutes of higher education [1]. An every-time, anywhere, and anyone system called mobile learning is being created. It is an updated version of e-learning in which there is no requirement for physical study locations. With touch screen user interface, ten- hour batteries, and desktop computer features, tablet PCs or smart phones are indispensible devices for the future of education.

At the same time, research in 3D multimedia is advancing at a rapid pace [4]. 3D media content technology simulates every aspect of reality [3]; this is known as virtual reality. In the virtual environment, users easily recognize and interact with each other to study, or collaborate. In order to create the virtual world, many virtual objects must be made. Virtual objects fall into two categories: one, pure virtual objects, which are created by people's imaginations, or two, digital replications of real objects. The second is what we will discuss in this paper. The 3D Noh-mask for mobile devices is a project that captures real Noh-masks in many different angles, and uses 3D software to create a 3D object model. The user can then control the app with the fingertips, or mobile accelerometer to see the object at any angle.

II. Technique Procedure:

There are many ways to capture a real object and generate its 3D model; however, this paper introduces the use of Strata Foto 3D CX and Blender. The Strata Foto 3D CX is commercial software, which exports its 3D data so that Blender, open-source software, can recognize it. Then Blender exports the 3D data object to iPhone, and iPad. It's important to have accurate data in every photograph so that the software will work properly.

1. Capture a real object

1.1 Studio set-up

1.1.1 Mat, turntable, and stand

1.1.1.1 Mat:

The Strata Foto 3D CX uses a flat sheet of paper printed with dots to determine the sequence in which each photo was taken. The mat can be printed from the software's help file. A fifteen dot-pattern is on the mat, and the first dot-pattern is marked by a line.

1.1.1.2 Turntable:

The turntable is a plate that is the same size as the mat, and can easily spin 360 degrees.

1.1.1.3 Stand

The stand is a cube of the same color as the mat, used to raise the object away from the marking on the mat and reduce shadows.

The mat, turntable, stand, and object should stick together as a single unit when they are rotated.

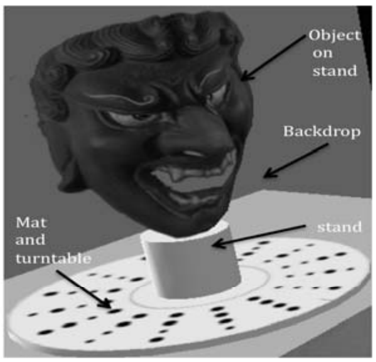


Figure 1: Object, stand, mat, turntable, backdrop setup

1.1.2 Solid color backdrop

Depending on the object's color, the color of the backdrop must be an opposite color. For example, in this project, we use a plain green background because the mask appears in a dark brown or black. The backdrop is always placed behind the object in every photo that is taken, making it easy for the software to detect the object and the background.

1.1.3 Lighting

There must be no shadow on the backdrop when the object is placed on the stand. In order to have a good quality photo shoot, light in the studio should be distributed evenly on the object and its surrounding area. This is best achieved when all lights used are of the same type.

1.1.4 Camera

Photographs may be taken using a standard digital camera. In order to have the same angle for the photo shoot, the camera must be mounted on a tripod. Here are some recommended pre-settings for the camera:

- Disable the camera's flash
- Use diffused light to make sure that there are no hot spots on the object
- Disable the digital zoom
- Lastly, clean the camera's lens

1.2 Photograph the object

1.2.1 Test before taking the picture

From the camera's angle, turn the turntable in many different directions to make sure that the backdrop covers the object's background. In any shot, the view of the entire object must be captured; the mat dot-pattern must be shown in the photo so that the software can recognize the object. There are three views: side-view, high-angle-view, and top-view.

1.2.2 15 photos from the side-view

15 photographs from the side-view provide the most important data for the software to build the object's 3D model. In the picture below, the angle between the camera's center line and the object's horizontal line is zero degrees. This is the angle of the side-view. After all configurations are checked, take the first shot. Then turn the object clockwise to the next dot-pattern, and take a second shot. Repeat the steps until all fifteen photos are taken.

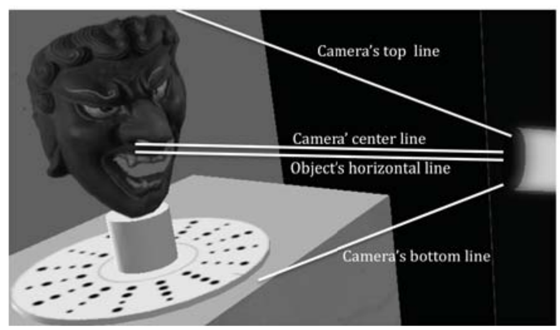


Figure 2: Camera position for side-view photographs

1.2.3 3 photos from 60-degrees

In order to maintain the distance between the object and the camera after taking the side-view photos, keep the tripod with the camera in the same (x,y) position. Next raise the camera up and tilt it until the camera's center line and the object's horizontal line are about sixty degrees. At the same time, make sure the first dot-pattern is facing the front, and then take the first shot. Next move the turntable 120 degrees, or to the fifth dot-pattern, and take the second picture. Finally move the turntable 120 degrees, or to the tenth dot-pattern, and take the third shot.

1.2.4 1 top-down photo

There is only one photo to show the top view of the object; therefore, it is important to do it carefully. In this case, the camera's line and the object's horizontal line make a ninety-degree angle. The camera's top and bottom line should be little outside the edge of the turntable so that the dot-pattern is shown in the photograph. After everything is aligned, shoot the photo.

1.2.5 1 bottom photo

As with the top-down photo, the bottom photo has to be taken carefully in order to show the detail on the base of the object. Turn the object upside down, and repeat the process for taking the top-down photo.

1.3 Creation of the 3D object model

Import all images from the digital camera into one folder and then open the Strata Foto CX and a new project. Next, import all images into the project. Use the software's wizard to mask all images so the software automatically detects the photos that have been imported. The software makes those photos' background transparent. If the software could not completely change the background, especially around the edge of the object, go to those photos and manually clean the remaining pixels. This process will improve the quality of the photos so that the software can build a better 3D object from them.

From the software menu, generate a 3D wire frame model. This is required for the completion of a 3D object. Finally, apply texture over the entire wire frame model.

1.4 Export the object

Export the object's data as .3ds from Strata Foto CX. Next, open a new project in Blender, open-source software, and then import the .3ds object into the Blender. Then, from Blender, export the data object as .obj. Finally, convert the .obj file into an Objective-C header by using a Perl script.

2 Coding

2.1 Development environment

2.1.1 Xcode, and Simulator

Usually in iPhone or iPad development, a lot of time is devoted to building a user interface by using UI Builder, a user interface software which is included in Xcode. A developer may code the app by using Object-C, a language of Xcode, to integrate. However, in this project, we used Xcode as the main workplace, and all interfaces are built in Xcode.

Simulator is used for iPhone or iPad testing. However, Simulator cannot handle all iPhone, or iPad features such as the accelerometer. Use of a real iPhone, or iPad is recommended for a live test. In order to use real devices for testing, a developer must register for an Apple developer account. Then, he or she can download code samples, create project members, or generate provisioning to integrate with real devices.

2.2 Coding procedure

2.2.1 Create a new OpenGL ES project

OpenGL ES is a cross-platform standard 3D API for handheld and embedded devices. OpenGL ES is a C-based API, and, in Mac OS, it integrates seamlessly with Objective-C based Cocoa Touch applications. When a developer develops an OpenGL ES application, his or her OpenGL ES content is rendered to a special Core Animation layer, known as a CAEAGLLayer object. In the project, we have used OpenGL ES 1.1, which provides a standard fixed-function 3D application.

There are a couple of ways to create a new OpenGL ES 1 from Xcode wizard or from Apple OpenGL ES 2.0 example code. In the project, we used the example code. Then, we imported all masks' Objective-C header files and JPEG textures into the project's Resource folder. When we had all necessary data, we overrode some of the project's classes.

- **ESRenderer.h**

This is the class to define functions of ES1Renderer. In this project, there is no implementation for this class.

- **ES1Renderer.h, and ES1Renderer.m**

The ES1Renderer class creates an OpenGL ES 1.1 context and draw using OpenGL ES 1.1 functions.

They are the main classes to handle the external data such the object's JPG map texture, and Objective-C header file of the vertex array. There are some declarations in the header file, then implement them in the method file:

- "***loadImageFileL(NSString *)name ofType:(NSString *)extension texture☹uint32_t)texture***" function is to import the external JPG texture map into OpenGL ES app when the app is running.
- "***touchedAtX: (float)x andY:(float)y***" is used to detect user's touched-coordination
- "***drawTexture***" is re-drawing the screen whenever any graphic is updated.
- "***renderMain:model:(uint32_t)model;***" is the active place in OpenGL ES app after the "***init()***" function completed its task.

Especially in ES1Renderer.h, all vertex coordination for the object header files need to be imported. For example, mask1.h.

- **EAGLView.h, and EAGLView.m**

The EAGLView class is a UIView sub-class which renders an OpenGL view. The class first tries to allocate OpenGL ES 2.0. If it fails, then it allocates an OpenGL ES 1.1.

In these two classes, "***ES1Renderer.h***" file must be imported so that EAGLView can use the OPENGL ES 1.1 context.

In EALView.h, "***id*** **<*ESRenderer*>** ***renderer;***" should be replaced by "***ES1Renderer*renderer;***"

- **"ProjectName"AppDelegate.h and "ProjectName"AppDelegate.m**

When creating a new project, "ProjectName" is the name a developer gives to his or her project. Then, Xcode will concatenate the "ProjectName" and AppDelegate to make the "ProjectName"AppDelegate header file, and its implement file. They are the main classes to tie everything together for running. There is no modification in these classes.

2.3 UI design

In this project, we tested two different ways to interact with users. The first way is to use an accelerometer. When a user moves the iPhone or iPad device, the mask will display according to the angle of movement. When he or she wants to view the back side of the object, a shake of the device is needed to turn the object 180 degrees. However, some users may find it difficult to control the object in a wanted angle. Therefore, we also implement a swipe-out and automatic 360 degree rotation. In this way, the mask object will display its front view as the default setting, then it will be rotated slowly. At any time, the user can stop the rotation for observation by touching the screen.

Figure 2: A screen shoot for a Noh-mask running on an iPad

III. Evaluation:

1. Disadvantage

There are some objects which may not work well with Strata Foto CX such as colorless glass, extremely complicated items like hairbrushes, mirrors or other highly reflective items, and objects with deep recesses like coffee cups. It takes a long time to build a 3D object from taking photographs to creating the data for OpenGL ES. There are many tasks that have to be done manually.

2. Advantage

Although some objects such as Japanese Noh-masks are difficult to replicate virtually, once they are replicated in 3D they can be used in further development as components in mobile games or any 3D documentary content.

IV. Conclusion:

On January 8th, 2011, we had a demonstration of the iPad application for audiences in the Traditional Arts at a State of the Art Technology seminar held in Burlingame, California. After a real Noh performance, we let users try our iPad app in which nine different Japanese Noh masks had been built. The audience was impressed to see those masks from many different angles. It was a great experience.

For future development, we are working to cut down the time it takes to create 3D reality-objects. We want to build a studio that can take all photos, and import them into a computer automatically. In addition, we plan to write scripts to create Objective-C header files from .3ds data files. We are also researching to create educational games, and other 3D content from virtual-reality objects.

Reference:

[1] "2011 Horizon Report," EDUCAUSE Learning Initiative, p11-15, 2/2011.
[2] Aaftab Munshi, Dan Ginsburg, Dave Shreiner, "OpenGL ES 2.0 Programming," Addison-Wesley, p101-125, August 2009.
[3] Barney Dalgarno, "Spatial Learning in Virtual Environments – An Exploration of the Distinguishing Characteristics of 3D Virtual Environments and the Contribution of these Characteristics to Spatial Learning," VDM Verlag Dr. Muller Aktiengesellschaft & Co. KG., p10, 2008.
[4] Dinei Florencio, "Recent Progress in 3D Multimedia Research," Hot3D'2010, July 23, 2010.
[5] Phillip Rideout, "iPhone 3D Programming," O'REILLY, pages 1-42, 2010.
[6] Tony Mullen, "3D for iPhone Apps with Blender and SIO2-Your Guide to Creating 3D Games and More with Open-Source Software," Wiley Publishing, Inc. , p29-56, 2010.

Cloud design for learning and mobility services

Giorgio Valle
Dipartimento Scienze Informazione
Università degli Studi di Milano
Milan, Italy
Giorgio.Valle@unimi.it

Bruno Apolloni
Dipartimento Scienze Informazione
Università degli Studi di Milano
Milan, Italy
Bruno.Apollini@unimi.it

Francesco Epifania
Dipart. Elettronica e Informazione
Politecnico di Milano
Milan, Italy
Epifania@elet.polimi.it

***Abstract*—**Two Business case studies in University Laboratories on learning and mobility present the opportunity to experiment new multimedia data management, data analysis, distributed collaboration. Software engineering practices based on cloud technology techniques acquire from the Web what is needed for rapid prototyping, capability of satisfying the needs of customers, market validation of pricing policies, marketing strategies.

As an alternative to standard software engineering programming, this approach allows to specify the projects' concepts by reaching stakeholders of needed ICT components to implement a prototype relying on SW components available elsewhere. Two sample cases clarify this cloud-engineering approach to the development innovation: 1) Program for Recovery Insufficient Grades in High-School (PRISC); 2) Social Mobility (SM) to reduce dependence on owned cars to satisfy mobility needs.

The PRISC strategy is to focus on the learner and on how proficiency is assessed through grades. The action steps: 1) understanding is focused on the students' study on schoolbooks; 2) assessment is delegated to exercises and tests managed in interactive Web environment on screens of mobile telephones or PCs; 3) students with excellent grades act as tutors by exploiting the social networks.

The SM strategy focuses on exploiting the empty seats available in most cars in metropolitan cities with air pollution problems. Key features: 1) direct negotiation via mobile devices between car drivers and ride seekers; 2) short range contacts insuring prompt satisfaction; 3) GPS ride surveillance.

Keywords: mobile applications – web services – distributed collaboration – distance learning – user profiling, reasoning and recommendations

I. INTRODUCTION: CLOUD ENGINEERING

From the standpoint of a University environment, the attention that Cloud computing receives on scientific publications and on daily press as well, indicates a major potential for business innovation and job opportunities for the young graduates. This goal is achieved once a win-win situation is established which could leverage on cloud technology techniques to acquire on the Web what is needed for rapid prototyping of business models concepts, while reducing time-to-market to satisfy the needs of customers.

Technical aspects are widely covered in the IEEE International Conference on Cloud Computing CLOUD 2010 [1] held on July 5–10, 2010 in Miami, Florida, sponsored by The IEEE Technical Committee on Services Computing in IEEE Computer Society.

Basically, Cloud Computing seems a replication of the client-server innovation in the past decades in computing resource allocation, while the focus now moves to acquiring what is needed over Internet. With the additional present capability of the users' interaction on demand over mobile devices without any time and location constraints. Companies are discovering that traditional in house applications can be handled with significant reduction in Hardware and Software costs, without turning to teams of expert employees for installation, testing, execution, protection and updating. This approach increases capacity or add capabilities without investing in new infrastructure, training new personnel, or licensing new software. Cloud computing encompasses any subscription-based or pay-per-use service that, in real time over Internet, extends IT's existing capabilities.[2]

Cloud based services require to open a browser, address a provider, open an application, and start using it [3]. Benefit is particularly significant for small, medium enterprises which benefit of IT services never previously available, enabling them to compete in innovation with much bigger companies because of lower budgets needed.

Vendors bet on Cloud technology: CA Technologies, Salesforce, Microsoft [4], 7thFLOOR, Cloud Zone, Google, Seeweb, IBM, Siemens Enterprise Communications, Capgemini Worldwide are only a few players in a wider list. The problem is that - as with Web 2.0 - everyone seems to have a different definition. As a metaphor for Internet, "the cloud" is a familiar cliché, but when combined with "computing," the meaning enlarges and gets fuzzier.

Cloud computing is at an early stage, with a variety of services, from full-blown applications to storage services: *SaaS***,** *Utility computing***,** *Platform as a service*, *Service commerce platforms, etc.* Cloud computing is also assumed to be a form of green computing. There are ongoing studies to substantiate how outsourcing the servers affects the environmental effects. Companies are served by a variety of data centres whose resources are very under-used, where the server utilization level is less than 20% by many estimates [5].

Two sample cases illustrate how this cloud approach has engineered two applications related to:

- *education* where users are high-school student with bad grades,

- *mobility* where users are those leaving their house with a car that requires fuel and parking space.

These two applications present several topics of interest in DMS 2011 conference, such as media coding, digital rights management, privacy and security issues, mobile intelligent applications, human-computer interaction, semantic computing, web services, multimedia databases and information systems, cross media authoring, distance learning, e-commerce, user's profiling.

II. PROGRAM FOR RECOVERY INSUFFICIENT GRADES IN HIGH-SCHOOL (PRISC)

High School Education is a major concern for Governments since economy globalization and technological innovation in industrial processes require staff and managers able to anticipate unpredictable market changes. (On the other end,) While lower government budgets and economy recession force to prototype new educational formats where ICT and social networking make education more effective.

To deliver a better overall education in High-Schools requires a co-ordinated action on the stakeholders: teachers, students, textbook publishers, school administration board, parents, laboratories and their equipment. [6]

PRISC strategy focuses on the learner and on how proficiency is assessed through grades, limiting its action to the last 4 years before either looking for a job or entering the University. Learners interested in PRISC action are both those with bad grades and those with excellent grades. The problem addressed is how a student with bad grades in specific subjects succeeds in recovering them in order to proceed to the next higher level class. In this initial phase, PRISC focuses on a few subjects: Mathematics, Physics, Computer Science.

At present bad grade recovery happens either with additional classroom-based teaching organized by the School free of charge to students (in Italy these teachers are paid order of 50 euro/hour) or by private action where parents look for teachers or University students) for private tutoring usually paid cash with no receipts. In Italy this process brings to one of the worldwide most expensive School systems ranking Italy as 37th Country out of the first 40 main Industrial States in the world in performances proposed by OCSE and PISA.

Target groups are teenagers still studying in High Schools and their parents who pay for private teachers when needed in case of bad grades. Final beneficiaries could include the University Systems as well since a better basic understanding of scientific subjects and humanities could be a good starting point for an excellent student's University curriculum.[7]
The overall objectives of PRISC project, while taking care of the students' recovery of bad grades when in High School, are:

1) enhancing topic understanding focused on the students' study of the topic where a bad grade was received using the text book adopted for the school year;

2) assessment of how well a topic has been understood is delegated to exercises and tests managed in an interactive WEB environment on screens of mobile telephones or PC;

3) create a student social network, where students with excellent grades act as teachers (same age) in this extracurricular additional school-like activity;

4) school textbook enhancement: all exercises and explanations are indexed and stored in a central database. Over time, similar bad grades could be recovered by automatic access to it with the potential of adopting artificial intelligence techniques to monitor over time learning improvement and to leave to an "expert system agent" the task to cope with big numbers of students in need of insufficient grades recovery;

5) assessment of better students' study performance happens in the same school environment where a bad grade was received at the beginning;

6) this "social learning community" is an obvious target for the exploitation of recent digital social networks like: LinkedIn, Facebook, Plaxo, etc [8]

7) the same happens with telecommunication technologies, such as video-telephone group calls among group of students and teachers.

All the procedures are managed on the cloud by definition of their characteristic data (see Table 1). The resulting system is inherently distributed (users do not know where data and procedures are stored), where the central management interacts with a set of web-services provided by the cloud interface. He doesn't know directly the actual students (both with insufficient or with excellent grades). He only interacts with standard tools reckoning the student activities and with a standard platform for teaching material exchanging and repository. Thus, there is not any venture capital previous investment, but only a pay-per-use fee that is tightly commeasured to the real PRISC activity. In particular: a) Study of specific topics not well understood on text-books involve text-book Publishers to edit relevant exercises b) Modern ICT digital video-interactive communications, over screens of mobile phones or PCs, involves Mobile Telecommunications Carriers c) Web-based support to a learning community, where students with good grades help other students to recover from bad-grades, involves Social Networks platforms d) Implementation involves partnership with Education stakeholders at national, local government, and private sector.

TABLE I

Service	***Provider***	***project***
Cloud platform	Heroku, Amazon EC2	PRISC- Sustainable Mobility
Database	Cloud Database, MongoHQ, Cloudant	PRISC - Sustainable Mobility
Video encoding	Panda Stream, Zencoder	PRISC
CDN service	Amazon AS3, Amazon CloudFront	PRISC
Queue system	Amazon SQS	PRISC
Push Notification	Pusher, PubNub	Sustainable Mobility
Payments	Paypal, Google Checkout, Amazon Payments	Sustainable Mobility
Mail and SMS notification	Sendgrid, Cloudmailin, Moonshado	Sustainable Mobility
Search engine	Websolr, Elasticsearch	Sustainable Mobility
Social authentication	Facebook, Twitter	Sustainable Mobility

III. THE SM (SUSTAINABLE MOBILITY)

SM strategy focuses on exploiting the empty seats available in most cars in metropolitan areas with air pollution problems [9]. Its implementation assembles system components acquired on the Web leveraging on a cloud technology approach for

- short range contacts for prompt satisfaction,
- GPS ride surveillance.

It is widely documented how car traffic creates pollution, higher expenses because of the raising cost of fuel and the daily stress to drivers. Many actions have been taken at Government, Region, and City levels to implement mobility strategies which reduce the systematic use of private cars. Sustainable Mobility is a paradigm for these strategies reducing environmental impact of vehicular traffic.

Although programs for promoting Sustainable Mobility in urban areas have been heavily financed, no solutions for traffic problems are currently available, resulting in continuous exceeding of the pollution indices defined by the European Community and lower individual productivity due to the waste of time in house-to-work commuting.

Since 2003 the lack of public transport on the desired routes has caused a progressive change in the strategies of the local public transport companies to provide mobility services. This implies managing the whole mobility within a circumscribed area satisfying many fragmented, different and hard-to-forecast mobility needs, via the so-called Advanced Transport Telematics (ATT) area identification [10]. The technologies needed range from telecommunications and informatics to sensoristics, from localization systems to correlate automation process technologies (defined also as ITS, Intelligent Transport Systems), which allow a centralized management of public transport services (knowing in real time the localization of vehicles and finding alternative solutions in case of failures or accidents), real-time information to users, payment via smartcards and the offer of personalized on-demand transport services. In spite of the potential of these applications, ATT systems are not yet widespread.

The distinguishing strategy through which SM plans to overcome this drawback is the raising up of a Social Mobility community capturing the consensus of the customers thanks to common ICT instruments, such as Internet and GPS, on the one hand, and socio/economic motivations on the other [11].

The approach is intrinsically distributed, having cloud computing as an ideal middleware. Indeed, the community autonomously emerges from the grouping of persons without any previous direct contact. They simply interact through an id which attests them simultaneously moving in the same geographic range and information tools in the clouds which handle transactions like the following:

1. the member ID_0 needs to go from place A to place B and therefore launches a call through a mobile terminal to all who want to supply him/her with a ride (paid or not) in the next 10 minutes.

2. The ID_1… , ID_n members are in the same area of ID_0 and declare their availability for a partial or the total cover of the ride.

3. The cloud devices assign the issue to one of the ID_i (or to a sequence of them), according to its/their credibility and to the options of the caller (payment, exchange of services, urgency, etc).

4. Analogous devices follow the correct behavior of the transaction via GPS and on this basis update the credibility and the credits of the involved actors.

The described transactions are usual, nowadays, for airway or railway long trips. Today the wide availability of connectivity and traceability on many personal devices allows keeping this approach for mobility services on micro-distances. The registration to the community requires little effort: i.e. filling of a profile form, acceptance of behavior norms, and acknowledgment of a minimum handling charge to the cloud team which handles the community. Moreover, in force of the strong advantages deriving from belonging to a community it is possible to require the respect of "ad hoc" norms and requirements. First of all the limited pollution rate of the carriers, then the usability of some data about traffic to define antipollution and antitraffic plans for specific critical zones, etc.

The model is intrinsically distributed, where the managers of the transactions are exactly the actors who access remote resources through a portal installed through APIs on their mobile terminals (see Table 1). This will allow the achievement of the following basic goals: A) extemporaneity of the transaction. We assume the user to be waiting for approximately 10 minutes between when he plans to move and when he identifies the way to do it. Otherwise he decides to move with his own means. B) Efficiency of the transaction. It is a matter of operational and technological efficiency to guarantee the above wait, but also the energetic and ecological efficiency as a keen quality parameter of the proposed solutions. C) Reliability of the transaction. We consider aspects of both reliability and security of the service. The latter will be favored by the tracking off the rides and the triggering of emergency procedures in case of anomalies.

The GPS availability allows to divide the community in users grouped by location (and consequently by time), so that we could combine demand and offer in an optimal way both in terms of quality of the services (QoS) and of energetic consumption.

From the methodological point of view the realization of the project requires to reach the following synergic objectives:

1) study of the micro-mobility phenomenon given by the transport request [10]. On the one hand this requires the characterization of the needs related to the personal micro-transport and of the inertial behavior that induces the use of one's own car. On the other hand we must study the statistical models for the mobility with the aim to synthesize a spatial-temporal model that is asymptotically stationary even if resulting by the composition of not stationary processes engaged by the single users [12].

2) the realization of an 'ad hoc' information system. The focus of this implementation will be the complete commitment of the transaction success to a simple API exploitation within a self-correcting procedure supported by remote, totally automatic information tools. We need to realize: a) a distributed information support for the formulation of the ride requests and offers. The peripheral hardware will be constituted by new generation mobile terminals endowed with Android operative system. An important part of this support will be the message cyphering and antintrusion systems and the transaction sampling and log compacting systems. b) a cloud database for the collection of the requests, their interface with the availabilities and finally the dispatching of the service orders to the transport suppliers. The DB might also contain the anagraphical data of the community members with their preferences profile and credibility, updatable on line. The DB will store the traces of the transactions, including GPS traces, for a time sufficient for security control needs, but not exceeding the limits of the privacy of the members.

3) the design and implementation of the security system guaranteeing the passenger and driver in a non invasive but sufficiently robust way. Main points are:

3a) the transaction enrolment from the driver-passenger assignment to the end-of-transport message by both the actors. The transaction will be followed both through the service messages and the GPS traces of the driver and of the passenger compared with the optimal routes

3b) a library of rules for the agent system. With this library we will face both unpredictable event caused by traffic, and anomalous behavior of either the passenger or the driver. We will use fuzzy rules to be calibrated both in batch mode and online with neurofuzzy learning mechanisms [13].

4) the assessment of algorithms for the optimal resource allocation. The problem is complex and multitask and depends on variables that are highly correlated to one another [14].

5) the design of a business and governance model that guarantees the advantage deriving from the system to all the community members including transport suppliers and cloud team. It will be a system with characteristics very similar to electronic commerce but with a core business represented by the assignment of acquisitions on the basis of competitions in which the offers are partially automatic (based on the passenger and driver profiles), and partially defined in an extemporary way (with an acceptance declaration and with the objective spatial location of the two actors). The technological aspects of the transaction will contribute to the definition of the model re the following aspects: a) the composition and structure with reference to the stakeholders; b) the perimeter of the offered services and the revenue model; c) the activities directly ensured and certified; d) the economic and financial sustainability.

IV. CONCLUSIONS

In a recent inquiry made by IBM over 2000 IT professional in more than 87 countries, the 91% of returns forecast that cloud computing will play the role of principal IT achievement within 2015. Analysts of this sector expect a significant growth of cloud computing services as well.

The new IBM Laboratory recently opened in Hursley, UK, to help business partners to take advantage of cloud computing is a further support to an estimated increase of this market from the current 68 billions of dollars to 150 billions in the next 4 years.

Similarly, a Cloud Business Services study featuring data from over 1000 respondents and conducted in cooperation with the Outsourcing Unit at the London School of Economics shows how Cloud Business Services will change the way business and IT will work together. It forecasts that future centers on low cost 3rd party "utility style" become safe as outsourcing governance matures and new roles, skills and capabilities inside and outside enterprises make the cloud based operating model work.

REFERENCES

[1] "IEEE International Conference on Cloud Computing (CLOUD)".Thecloudcomputing.org http://www.thecloudcomputing.org.

[2] Luis M. Vaquero, Luis Rodero-Merino, Juan Caceres, Maik Lindner "A Break in the Clouds: Towards a Cloud Definition" ACM SIGCOMM, Volume 39, Number 1, January 2009

[3] Jörn Altmann, Björn Rupp and Pravin Varaiya : Internet demand under different pricing schemes, pp 9 -14, ACM Conference of Electronic Commerce, 1999

[4] http://www.microsoft.com/windowsazure/pricing accessed on January 25, 2011.

[5] R. Smith, "Power companies consider options for energy sources," The Wall Street J., p. A.10, Oct. 29 2007.

[6] Valle, G. & Epifania, F. (2008). An Interactive VideoJournal to complement Learning & Training. In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008 (pp.1164-1169). Chesapeake, VA: AACE.

[7] Epifania, F. (2007). An experience in distance learning for university and secondary school. In G. Richards (Ed.), Proceedings of World Conference on E- Learning in Corporate, Government, Healthcare, and Higher Education 2007 (pp. 6786-6791). Chesapeake, VA: AACE.

[8] Fovet, F. "Impact of the use of Facebook amongst students of high school age with Social, Emotional and Behavioural Difficulties (SEBD)" 2009 Proceedings - Frontiers in Education Conference, FIE , art. no. 5350786

[9] http://www.erideshare.com/

[10] A. L. Bristowa; A. D. Pearmana; J. D. Shiresa. An assessment of advanced transport telematics evaluation. Transport Reviews, 17, 3 , 177 – 205, 1997.

[11] G. Elofson, P. M. Beranek, and P. Thomas. An intelligent agent community approach to knowledge sharing. Decision Support Systems, 20(1):83–98, 1997.

[12] D. D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. Nature, 439:462–465, 2006.

[13] B. Apolloni, S. Bassis, D. Malchiodi, and P. Witold. The Puzzle of Granular Computing, volume 138 of Studies in Computational Intelligence. Springer Verlag, 2008.

[14] B. Apolloni, S. Bassis, and A. Zippo. Processing of information microgranules within an individuals' society. Human-centric Information Processing Through Granular Modelling, pages 233–264, 2009.

Proceedings
International Workshop on Distant Education Technology (DET 2011)

Co-Editors

Oscar Fuhua Lin, *Athabasca University, Canada*
Teresa Roselli, *University of Bari, Italy*
Veronica Rossano, *University of Bari, Italy*

The Zone of Proximal Development between Adaptive Learning and Reputation-based Group Activities

M. De Marsico, A. Sterbini
Dept. of Computer Science
Sapienza University of Roma, Italy
{demarsico, sterbini}@di.uniroma1.it

M. Temperini
Dept. of Computer and System Sciences
Sapienza University of Roma, Italy
marte@dis.uniroma1.it

***Abstract*— Vygotskij's educational concept of the Zone of Proximal Development (ZPD) calls for an adaptation of learning activities to the present state of learner's knowledge and abilities. At the same time, Vygotskij's educational model includes a strong bent towards social and collaborative learning. In practice, a tight integration between personalized and collaborative learning is called for. Along this line, we exploit two previously implemented prototypes of systems for web-based e-learning, to investigate the integration of adaptive e-learning with social learning activities, both at individual and group level. LECOMPS is a web-based e-learning environment for the automated construction of adaptive learning paths. SOCIALX is a web-based system for shared e-learning activities, which implements a reputation system to provide feedback to its participants. We propose a two-way tunneling strategy to integrate the above prototypes: the LECOMPS student model is used to select the set of social activities (met in SOCIALX) according to the present individual learner state of knowledge; on the other hand, the solution of exercises, and the associated reputation gained in SOCIALX, are used to update the LECOMPS student model. Under the social perspective induced by the integration, we present a mapping between the student model and the definition of Vygotskij's Autonomous Problem Solving and Proximal Development regions, with the aim to provide the learner with better guidance, especially in the selection of available social learning activities.**

Keywords: zone of proximal development; adaptive e-learning; social collaborative e-learning; reputation system; group modeling

I. INTRODUCTION

The aim of the design strategy presented in this paper is to merge personalized learning paths and acquired learning reputation under the auspices of Vygotskij's theory of proximal development [1][2], while further strengthening the sense of community in a class. In order to achieve these goals, we rely on the integration and exchange of data, about students' performances and achievements, between two formerly developed web applications: LECOMPS, an environment for personalized e-learning, and SOCIALX, supporting collaborative and social learning activities.

Personalized learning paths allow to better respond to learners' needs and preferences, while reputation based techniques tend to increase the student's motivation. In our model, the learner's activity in the socio-collaborative reputation system (in SOCIALX), and the reputation itself, are used to refine the individual student model (in LECOMPS). In the other direction, the learner's overall profile (of LECOMPS) is used to enrich a course with social-learning activities. This strategy fully agrees with principles and guidelines of Learner Centered Design [3][4].

II. MOTIVATIONS AND BACKGROUND

A great deal of interest and research is presently dedicated to supporting automated construction, maintenance and delivery of adaptive e-learning courses. On the other hand, collaborative learning is considered a winning methodology to allow the development of meta-cognitive abilities, such as critical thinking in learners, and to foster the acquisition of new knowledge. Moreover, it appears to support better retention and deepening of knowledge over time [5]. In a collaborative environment the learners are prepared for team-based working activity, both by sharing their common experience and combining their skills [6]. Thus it becomes an aspect of interest the social dimension of e-learning, viewed as a community in which social activities take place and social interaction skills are developed by the *participants*. Along this line, in connectivism perspective [7] as well as in situated learning model [8], knowledge is a system which is accessed through people participating in (ecologically significant) activities. An example of e-learning environment implementing Organizational Learning is the environment E-MEMORAe2.0 [9]. Its aim is to prepare the students to work in team, and above all to share and exchange knowledge. Such skills are important both in formal and informal learning, and the same holds in the perspective of lifelong learning [10]. Moreover, while any organization, including a class of students, cannot learn without its members being continuously learning [11], it seems crucial for the organization to act as a Community of Practice (CoP), in order to better enable its members to share experiences, knowledge and skills, in a true peer-to-peer learning. So, development of CoPs is surging ahead. A CoP is "a group of people who share an interest in a domain of human endeavor and engage in a process of collective learning that creates bonds between them: a tribe, a garage band, a group of engineers working on similar problems." [12]. CoPs develop along a number of dimensions. Wenger [13] lists eight dimensions, classifying technological products which may support CoPs to a different extent: ongoing integration of work and knowledge, (collaborative) work, social structures, conversation, fleeting interactions, instruction, knowledge exchange, documents. As we can notice, these are the same core components of the collaborative side of an advanced e-learning system. The main difference is that, while CoPs rise and grow in a completely spontaneous and quite unsupervised way, e-learning systems still need the role of the teacher as main inspirer, supervisor and guiding figure. The integration of SOCIALX for collaborative aspects and LECOMPS for supervised ones is a first attempt towards the long term goal of fully developing an e-learning system including all required CoP-like features, in other words, a Learning Community of Practice (LCoP). A reputation system is a natural component for building a LCoP, suitable in particular to motivating and transforming a class (or some classes) of learners into a real community. The reputation captures the contribution of each learner, and makes it apparent to the community (the group, the class and the course). The reputation is both a motivational tool and a way to evaluate and understand learner's psychological characteristics (learning and communication style) as well as preferences, relations with others, and ability to analyze/judge others' work (meta-cognitive achievements).

III. VYGOTSKIJ'S COOPERATIVE MODEL

One of the main achievements of Vygotskij's research [1][2] is a model aiming to demonstrate that cooperation provides the basis for the individual development. Even during childhood, targeted as well as casual interactions activate and spur the cognitive processes. The presence of people in the same environment, and the cooperation with peers, induces a reflection and an auto-regulation of one's own behavior. Once such processes are interiorized, they become part of the child's autonomous evolution. Social learning therefore precedes individual competencies and determines and prepares cognitive development. As a matter of fact, according to Vygotskij, the typical direction of learning is from outside to inside: the knowledge interiorization process mainly happens through the social "co-construction" (social learning), and then proceeds through a progressive transfer of the exterior social activity to the interior control. Once knowledge and processes are interiorized, the learner will be able to proceed in an autonomous and independent way. It follows that concrete growth can occur only in the Zone of Proximal Development (ZPD), which is characterized as [14] "the distance between the actual development level, as it is determined by the autonomous problem-solving, and the level of potential development, as determined by the problem-solving under an adult's guidance or in collaboration with one's own more capable peers". In practice, when the learner has consolidated a region of Autonomous and independent Problem Solving (APS), it is useless to further suggest exercises related to the same level of difficulty. On the other hand, it is even more useless to suggest exercises which are completely out of reach for the learner (Unreachable Problem Solving – UPS). The right zone of complexity, is the one including exercises that the learner can solve, according to her own competence and/or with a moderate support and/or in collaboration with companions (Zone of Proximal Development – ZPD) (See Fig. 1).

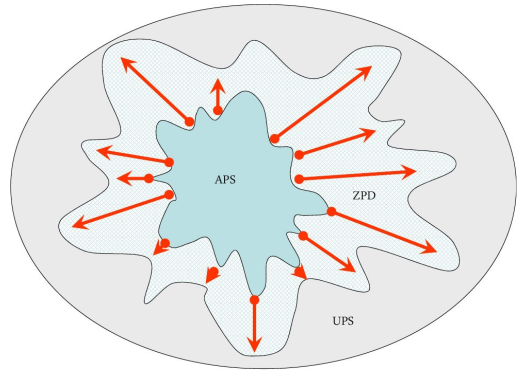


Figure 1 Cognitive zones in Vygotskij's model.

IV. TWO WEB-APPLICATIONS FOR PERSONALIZED SOCIAL LEARNING

A. *LECOMPS*

LECOMPS [15] (Fig. 2) is a web-based e-learning environment, supporting learners and teachers in a variety of activities:

- authoring of learning objects (*Learning Components* – LCs) and their organization in pools defining "subject matters";
- enrolment of learners and evaluation/maintenance of their personal *Student Model* (SM)
- automated construction and adaptive delivery of personalized courses, tailored over the learning goals (*Target Knowledge* – TK), stated for the subject matter, and the (evolving) individual SM.

The whole set of LCs dedicated to the subject matter composes the *Learning Domain* (LD) for the courses. A course *C* is a sequence of LCs, a subset of LD, built to let the learner bridge the gap between her initial state of knowledge and the TK.

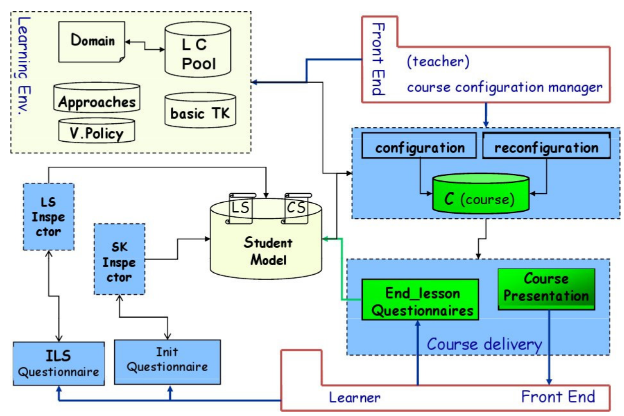


Figure 2 Architecture of the system LECOMPS.

The knowledge associated to a learning component (*required* to study its content, or *acquired* during its study) is represented through the conceptual device of the *Learning Objective* (LO) [16], that is a predicate such as

LO(level, keyword, {concepts}, context)

where *level* and *keyword* are cognitive characteristics of the *concepts* (topics about which the LO does express a skill), and *context* designates the learning context of the concept(s). (Keywords and level are grounded on the taxonomy of Bloom [17]).

In detail, a learning component, *lc*, includes:

- *learning content* (an XHTML-formatted resource, that can be given in different, LS-wise, versions;
- two sets of LOs: *lc.RK* (required knowledge), and *lc.AK* (acquired knowledge, see earlier);
- *questions*, related to the LOs in AK, used in questionnaires to assess knowledge acquisition;
- *effort*, informal measure of the LC content.

Among LOs, inference rules hold, such that the possession of certain LOs (such as a set $\{l_i\}_{i\in I}$) can imply that some other LOs are possessed too: $\{l_i\}_{i\in I} \vdash \{l_j\}_{j\in J}$. For instance, a skill possessed at a certain cognitive level implies the possession of the same skill at lower level:

LO(3, apply, cpt, cxt) $\vdash$ *LO(2, describe, cpt, cxt)*).

The set of all LOs associated to LCs in the LD is the *Knowledge Domain* (KD) for the subject matter.

The student model SM denotes the state of knowledge (CS – *cognitive state*) of the learner and her *Learning Style* (LS) preferences, as couple <*CS,LS*> where

- *CS* is a set of pairs $\{<lo_i, cert(lo_i)>\}$: LOs presently "owned" by the learner are listed and labelled by their estimated certainty (see later).
- *LS* is a 4-tuple $<d_1,v_1; d_2,v_2; d_3,v_3; d_4,v_4>$, where v_i are in the range [0,11], and d_i are values in the dimensions of Felder-Silverman's model [18] (active/reflexive, sensing/intuitive, visual/verbal, sequential/global.).

The SM possibly changes during the study, allowing for an adaptive retuning of the course.

The acquisition and continuing possession of a LO in CS, and its *certainty*, are determined basing on learner's answers to end-lesson tests (where a lesson is a segment in the sequence *C*). The system manages a set of parameters (configurable by the teacher on each subject matter) to drive CS updates during the course taking. In particular, when a *lo* is added to CS after a test, it is assigned certainty c_{IN}; then, after each further test, *cert(lo)* is decreased/increased, by c_{KO}/c_{OK},, depending on answers; should *cert(lo)* eventually be under/over thresholds $c_{Promote}/c_{Demote}$, though, it would be respectively extracted from CS or permanently included in it (with no further tests).

B. SOCIALX

SOCIALX [19] (Fig. 3) is a web-based system designed for collaborative and social aspects of learning. It supports the practical/exercise experiences of a course, through the management of socio-collaborative learning activities, in the framework of a reputation system [5] [6][20] [21].

In particular, in SOCIALX the learners can contribute by:

- *providing solutions* to available exercises;
- *using others' solutions* to develop their own;
- *evaluating* others' and one's own work;
- *discussing* exercises in contextual micro-forums, in which direct reward (*tokens*) is earned for the perceived usefulness of one's contributions (*questions & answers*);
- *participating to group-based projects*, with a "social" bent. Several projects are available, structured in common stages; at each stage a group is expected to produce a deliverable, also by exploiting deliverables from previous stages of the same project; after stage completion, the group moves to the following stage of another project.

The reputation in SOCIALX is a representation of the following learner's characteristics and qualities demonstrated during the interactions with the system:

- *involvement*, i.e. the active participation in the system (through the abovementioned contributions);
- *usefulness*: how a learner's work is beneficial for others (e.g. her/his solutions are taken for extension and reuse, her/his groups deliverables are not harmful for the further advancements of other groups in the project);
- *competence*: measured on the basis of grades and judgments coming from other learners and, especially, from the teacher;
- *judgment* and *self-judgment*, i.e. the consistence with others' (and especially teacher's) judgments;
- *critical appraisal*: the ability to select others' contributions to be extended or corrected;
- *group_reputation*, which measures aspects of the learner's activity performed during group work (number of products delivered by the group, number of marks given by the members of the group to products received from other groups, average teacher's marks on the group's products).

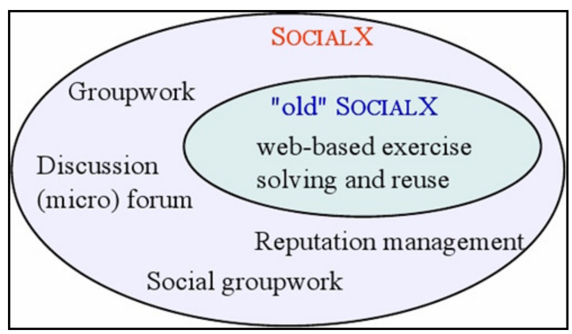


Figure 3 SOCIALX functionalities.

V. MODEL OF INTEGRATION

It is worth pointing out that, though Vygotskij's model is frequently cited in literature about e-learning, at the best of our knowledge a concrete attempt to formalize and put into practice its features in a working software framework seems to be still to come. A first step in the integration of the web-based systems, mentioned in Sec. IV, along the lines of Vygotskij's research, is the extension of the definition of SOCIALX exercises according to the structure of an LC, by RK/AK/effort. Quite straightforwardly, this allows admitting the SOCIALX learning assets (LCs that are actually mirrors of the exercises) into the LECOMPS pools and courses. They have no questions embedded, though, so we have to allow for a dedicated method of assessment of their AK: in brief, it is the reputation gained by the learner in SOCIALX, as related to the given exercise, which allows to state that its LOs are to be included, with according certainty, in the CS component of the student model.

Assuming that a course C can now contain both LCs and exercise-LCs, we thought it essential to let the learner have a more personal support to the navigation of the LCs in the course, besides the default sequencing provided by LECOMPS: that is important due to the presence of exercise-LCs, where prolonged, interactive and collaborative activities are supposed to be met during the course span.

In order to provide such a support, a classification of the course LCs (or, rather, of their Learning Objectives) under the "zones" discussed in Sec. III is helpful.

In the following we assume that $C=\{c_1, c_2, \ldots, c_n\} \subseteq$ LD is the course personalized for a given student, in the learning domain, according to stated learning goals TK. CS will be the personal Cognitive State of the student. Denoting the knowledge acquired through a LC c_i as c_i.AK, the overall knowledge provided by C is $C.AK = \cup_{c \in C} c.AK$. Accordingly, we'll call *projection of the knowledge domain over the course* the set of all LOs relevant to the course:

$$\prod^{KD,C} = C.RK \cup C.AK$$

(notice that $C.RK$ and $C.AK$ are not necessarily disjoint).

First we define the Autonomous Problem Solving region as the set of all the LOs that are in CS with maximal certainty

$$\text{APS} = \{lo \in \text{CS} \;/\; cert(lo_i) = c_{Promote}\}$$

As of the zone of proximal knowledge, we have to define which LOs of the KD (projected over C) are not "too distant" from the SM, i.e. from the knowledge in the CS. Those LOs are the ZPD, that can be shown the learner, to help her in selecting next LCs to take in C.

In this we have to rely on two further definitions:

- a distance metrics: $d(CS, lo)$ measures how far is a given LO, lo, from the learner's grasp (that is from the CS set);
- a "daring threshold" $p()$ on whose respect declaring that a given LO is "not too distant" from CS (and so be in ZPD).

In order to acquire a given $lo \in \prod^{KD,C}$, the learner is supposed to study along a path of LCs in C. Each LC, c, demands an effort, $c.effort$, and the distance between CS and lo is computed as the (minimal) sum of such efforts. So, let $G(CS,lo)$ be the minimal (wrt effort) subset of LCs in C, to be studied in order to have CS be extended so to comprise lo:

$$G(CS,lo) = \{c_{h1}, \ldots, c_{hk}\} \subseteq C$$
$$\text{with} \quad lo \in c_{hk}.\text{AK}$$
$$\text{and} \quad \cup_{c \in G(CS,lo)}\, c.RK \subseteq CS \;\cup\; \cup_{c \in G(CS,lo)}\, c.AK$$

then it is

$$d(CS, lo) = \sum_{c \in G(CS,lo)} c.effort$$

(In Fig. 4, $G(CS,lo)$={(red)-circled LCs} and $d(CS, lo)$=9).

The daring threshold should express a measure of how far from the APS the learner is to be reasonably allowed to reach, while undertaking the acquisition of new knowledge. Since such definition appears to be quite restrictive, we extend it, by possibly considering all the LOs in CS. In that, we weight such LOs by their certainty. Then the daring threshold for a given lo has a dependence on the overall (average) certainty of the LOs in CS that are used to reach it (i.e. those in $G(CS,lo)$).

So let us call *support set* in CS for lo, the minimal subset of CS made of LOs involved in $G(CS,lo)$ definition,

$$Supp(CS,lo) = CS \cap \cup_{c \in G(CS,lo)}\, c.RK$$

(In Fig. 4, $Supp(CS,lo)$ is the lighter (yellow) subset of CS, comprising the three LOs, with indicated certainty). An average certainty ($avgCert()$) for a set of LOs can be defined straightforwardly.

Then, the daring threshold is to be defined as a function $p()$, expecting the average certainty parameter and monotonically increasing with it.

So, the zone of proximal development is defined as the set of those LOs in the course (excluding the APS) that are within the maximum distance p from CS, as determined by the average certainty of the support set:

$$\text{ZPD} = \{lo \in \prod^{KD,C} \setminus \text{APS} \;/\; d(CS,lo) \le p(avgCert(Supp(CS,lo)))\}$$

VI. SOME IMPLEMENTATION DETAILS

The Pool of LCs and LOs defines an acyclic directed layered vertex-weighted AND-OR graph, with the LOs that correspond to OR nodes (as they can be acquired by more than one LC) and the LCs that correspond to AND nodes (as they demand their RK LOs). To compute the distance of an LO, lo, from the CS we compute the minimum subgraph for which lo is root and all leaves are in CS.

In Fig. 4, the distance of the rightmost darker (red) LO

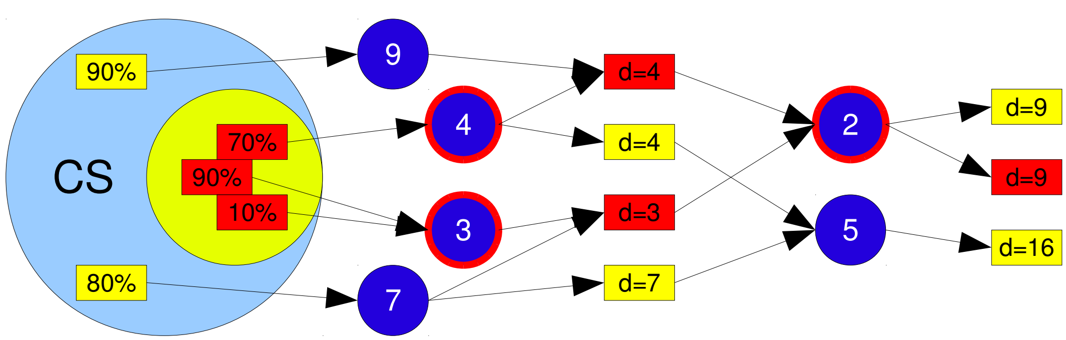


Figure 4 Computation of the distance d of an LO from the CS. (boxes represent LOs with their distance, circles represent LCs with their effort, arrows show the "acquired" relation, the highlighted set of LOs in CS is the support set of the red subgraph)

(d=9) is obtained by summing the efforts (3, 4, 2) of the minimum set of LC (the (red)-circled ones) that are necessary to reach the LO. In this process we consider only once the weight of any subgraph appearing on multiple paths.

Notice that, if the graph were a tree, the complexity of computing the distance of all LOs would be linear in the number of nodes, as we could apply bottom-up the following three cases: 1) the distance of an LO in CS is 0; 2) the distance of an LC is the sum of the distances of its required LOs plus its effort; 3) the distance of an LO outside CS is the minimum distance of the LCs that acquire it. In the case of a more general AND-OR graph, two different LOs could share part of their acquiring subgraphs, and thus the actual distance of an LC requiring them will be less than the one computed above. Solving this problem is equivalent to finding the minimum-cost abduction proof [22] of the LO, when we map the LOs in CS to unweighted assumable facts and we map LCs to weighted deduction rules.

VII. GROUP ACTIVITIES

In SOCIALX group activities are supported, and so they are going to be in the integration we are describing. A problem to address when selecting activities for a group is the determination of the overall group's CS (and accordingly the overall group's ZPD). Ideally, the group's common zone of proximal development should be as large as it is possible, in order to maximize the group members' gain coming from the common activities; on the other hand it should also avoid to present group members with activities that are, if not in their ZPD, at least not too far away, just in order to have that nobody is left behind. A first possible choice is to use the maximum common core of knowledge, of the group members (e.g. the intersection of their CSs). Unfortunately, this would limit each learner in the group to a ZPD basically defined over the weakest members: this would maximize the effect of outliers on the group ZPD, and probably reduce others' motivation; moreover this would deprive us of the possibility to leverage the support that could come from more experienced peers, which is a key feature in Vygotskij's model. Let us note that the dual choice, i.e. building the group's CS as the union of all members' CSs, would obtain similar negative results, mainly satisfying the brightest group members (them too being outliers), and leaving the others behind.

To satisfy both weaker and brighter students we may choose, therefore, an intermediate construction where we compute the group's CS as the union of the members' CSs, and where each LO has group-certainty assigned as the mean value of the certainty of that LO as in the members' CSs. This choice reflects more precisely the level of confidence in the possession of the LO by the group members. As a side effect, this model could motivate brighter students to help their weaker peers, in order to both improving the overall group's CS and letting the group ZPD grow, to comprise further, possibly more interesting, activities.

An alternative to such a direct construction of the group's CS is in following a reverse strategy and just defining implicitly the group ZPD, by establishing criteria of admissibility of activities for the group of learners. In this case two conditions are defined, by working on the group APSs (considered as "firm knowledge"), the group ZPDs, and the cognitive states of the group members.

For the first condition, we want to express requirements both on the group composition and on the selected activities. First, the group's member's (firm) starting points (APSs) must have some common intersection, and, second, at the same time such common starting point must allow the group to fulfil the activities prerequisites. So, given a group of students ST={s_1, …, s_{stud}} and a set of activities AC={lc_1, …, lc_{act}}, 1) *the group's members must share a common APS, and 2) the activities' prerequisites are well known by at least one of the members*: (APS_i is the APS of the i-th learner in ST)

$$\begin{gathered}(\bigcap_{s_i \in ST} APS_i \neq \phi) \wedge \\ \wedge (\bigcup_{lc_j \in AC} lc.RK \subseteq \bigcup_{s_i \in ST} APS_i)\end{gathered}$$

The second condition states that 3) students in a group ST must have some common proximal development, and 4) an activity lc_j in the group AC is admissible for ST iff, although it might be outside of the ZPDs of some members, it is *not too distant* from them, and it is comprised in the ZPD of at least one member (that could help the others reaching it): (ZPD_i is the ZPD of the i-th learner in ST; τ is a threshold to establish admissibility, for a learner, of an activity outside the learner's ZPD)

$$\begin{gathered}(\bigcap_{s_i \in ST} ZPD_i \neq \phi) \wedge \\ \wedge (\forall lc_j \in AC, \forall s_i \in ST : d(lc_j, ZPD_i) < \tau) \wedge \\ \wedge (\forall lc_j \in AC, \exists s_i \in ST : lc_j.AK \subseteq ZPD_i)\end{gathered}$$

Notice that if we used APS in place of ZPD, in the definition of the second condition, we would have some of the brighter members of the group left without anything new to learn in each one of the admissible activities, which could diminish their motivation.

VIII. CONCLUSIONS

Our long-term goal is the design of a Learning Community of Practice (LCoP) where the teacher plays the role of organizer and supervisor. Along this way, this work supports an adaptive design of the path of exercises that the learner follows during an adaptive course, and the management of an extended student model, gaining the feedbacks coming from the socio-collaborative activities supported by a reputation system. We addressed the problem of defining both an individual Zone of Proximal Development and a group one, in terms of the Learning Objectives stored in each individual student model. The determination of ZPD is done adaptively, according to the evolving student model; for that we have defined a cognitive distance between the state of knowledge of the learner and the learning objectives to acquire, and a ("daring") threshold function, dynamically stating the maximum distance, from the learner's CS, of the learning objectives in the ZPD, according to the certainty factors of the LO already in the student model. Cognitive distance and threshold function may deserve some further work, e.g. to consider meta-cognitive traits. Additional efforts will be directed at evaluating the correctness of the underlying assumptions, and longitudinal behaviour of the overall resulting framework.

ACKNOWLEDGMENTS

This work was partially funded by the Understand IT EU Life Long Learning Program project num. 2010-1-NO1-LEO05-01839.

REFERENCES

[1] Vygotskij, L. S. (1978). Mind in society: the development of higher psychological processes. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. and Trans.) Cambridge MA: Harvard University Press. (Original material published between 1930 and 1935).

[2] Vygotskij, L. S. (1981). The development of higher forms of attention in childhood. In J. V. Wertsch (Ed.), The concept of activity in Soviet psychology. Armonk, N.Y.: Sharpe.

[3] Quintana C, Carra A, Krajcik J, Elliot S (2001) Learner-centered design: reflections and new directions. In: Carroll(ed) Human-com-puter interaction in the new millennium. ACM Press, NY, 605–626

[4] Soloway E, Jackson SL, Kleim J, Quintana C, Reed J, Spitulnik J, Stratford SJ, Studer S, Eng J, Scala N (1996). Learning theory in practice: case studies in learner-centered design. In: Proc. SIGCHI Conf. on Human Factors in Comp. Systems (CHI'96), Vancouver, Canada, April 1996. ACM Press, New York, pp 189–196

[5] Kreijns, K. & Kirschner, P.A. & Jochems, W (2003). Identifying the pitfalls for social interaction in computer supported collaborative learning environments: a review of the research. Computers in Human Behavior, 19, 335-353.

[6] Cheng, Y. & Ku, H. (2009). An investigation of the effects of reciprocal peer tutoring. Computers in Human Behavior, 25.

[7] Siemens, G. (2005): Connectivism: A Learning Theory for the Digital Age, Int. Journal of Instructional Technology & Distance Learning, Vol. 2 No. 1 http://www.itdl.org/journal/jan_05/Jan_05.pdf#page=7

[8] Anderson, J.R., Reder, L.M., Simon, H.A. (1996). Situated Learning and Education. Educational Researcher, Vol. 25, No. 4, 5-11.

[9] Leblanc A., Abel M.-H. E-MEMORAe2.0: an e-learning environment as learners community siupport. International Journal of Computer Sciences & Applications, vol. 5 no.1, pp. 108-123. 2008

[10] Berger, T. (2009). Web 2.0 – Learning Culture and Organisational Change. In: Helen Keegan and Anne Fox (Eds.). Mentoring for 21st Century Skills – It's All About The Learning. University of Salford, UK.

[11] Nevis E., Di Bella A., Gould J. Understanding Organizations as Learning Systems. Sloan Management Review, Winter 1995, pp. 73-85, 1995.

[12] Wenger, E. (1998). Communities of practice: Learning, meaning, and identity. Cambridge Un. Press.

[13] Wenger, E. (2001). Supporting communities of practice a survey of community-oriented technologies. Available at www.ewenger.com/tech.

[14] Webb, I. (2003). Children, on-line learning and authentic teaching skills in primary education. Online at http://www.educ.utas.edu.au/users/ilwebb/Research.

[15] Sterbini, A., Temperini., M. 2009. "Adaptive Construction and Delivery of Web-Based Learning Paths". In Proc. 39th ASEE/IEEE Frontiers in Education , Oct.18-21 2009, San Antonio, TX, USA

[16] Fernandez, G., Sterbini, A., and Temperini, M."Learning Objects: A Metadata Approach" In IEEE Region 8 Eurocon 2007 Conference September 9-12, 2007, Warsaw, Poland

[17] Bloom, B.S. (Ed), 1964. Taxonomy of Educational Objectives, David McKay Company Inc., New York.

[18] Felder, R. M., and Silverman L. K. 1988. "Learning and teaching styles in engineering education", Eng. Education, 78(7), pp. 674-681.

[19] Sterbini, A., Temperini., M. 2009. Collaborative projects and self evaluation within a social reputation-based exercise-sharing system. IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology, Int.Wshop on Social and Personal Comp. for Web-Supported Learning Communities (SPeL 2009), IEEE CS P4174, ISBN 978-0-7695-3801-3

[20] Wei, W., Lee, J., and King, I. (2007). Measuring credibility of users in an e-learning environment. In Proc.16th Int. Conf. on World Wide Web, May 8-12, 2007, Banff, Alberta, Canada. ACM, New York, NY, pp. 1279-1280.

[21] Jin, F., Niu, Z., Zhang, Q., Lang, H., and Qin, K. 2008. A User Reputation Model for DLDE Learning 2.0 Community. In Proc. 11th Int. Conf. on Asian Digital Libraries: Universal and Ubiquitous Access To information, Dec. 02 - 05, 2008, Bali, Indonesia. LNCS 5362. Springer-Verlag, Berlin, Heidelberg, pp 61-70.

[22] M. E. Stickel, 1988. A Prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation. In International Computer Science Conference, Hong Kong, 343-350.

Sakai: Technology Transformation in an Open Source Community

Ian Dolphin
Executive Director
Sakai Foundation
Kingston upon Hull, UK / Ann Arbor, Michigan
iandolphin@sakaifoundation.org

***Abstract*— From its origins in 2003 as a grant funded project, Sakai has become a successful open source community, with over three hundred adoptions worldwide, and a growing ecosystem of community and commercial support. Over the same period, economic, policy and technology landscape changes, have dramatically modified the context against which an open source community rooted in Higher Education operates.**

The Sakai community is dealing with these challenges of this changed context on multiple levels:

The development of an innovative software environment, the Sakai Open Academic Environment, embodying an application of the experience of social networking to the online support of academic practice, and applying lessons of web-scale software development to that environment. Driven by community requirements, the project is providing a bridging layer from the established Sakai Collaboration and Learning Environment to the Open Academic Environment, thus allowing institutions to plan environmental transitions, or extend institutional capacity in support of online learning into social space.

The transformation of the Sakai Foundation, which is the organisational locus of Sakai community organisation, to sustainably support multiple software projects. This is taking place both internally, and by building new relationships with other open source communities.

I. INTRODUCTION

From its origins in 2003 as a grant funded project, Sakai has become a successful open source project, with over three hundred adoptions worldwide, and a growing ecosystem of community and commercial support. Over the same period, the landscape of online learning and academic collaboration has changed. Online learning has grown considerably in scale. According to Sloan-C, one in four college students in the US took at least one online course in 2008, a 17% increase on the previous year. Over the same period, there has been a dramatic transformation of the nature of the web, with an increased focus on social interaction, and the popularization of those technologies that have been labeled "Web 2.0".

Higher Education itself has seen dramatic changes, with reductions in funding in many national systems, and increased complexity from a range of key policy and economic drivers. These include the growth of remote and multi-campus institutions, moves to more flexible and internationally interoperable course and accreditation provision, the development of approaches to lifelong learning, and a significant and growing emphasis on interdisciplinary research.

These factors raise a number of challenges for an open source community rooted in sustaining innovation in the service of Higher Education. The Sakai Community is approaching these challenges in the following ways:

- The development of a new environment to support online academic collaboration, applying the lessons of social networking to this domain, and drawing on technology approaches which are designed to operate at web scale.
- Placing the Sakai community of educators at the heart of the design and development process.
- Transforming the locus of Sakai community organisation, the Sakai Foundation, to better support multiple, rather than single, software "products", and growing the Foundation by joining with other communities working in this space.

II. OPEN ACADEMIC ENVIRONMENT – BACKGROUND AND DESIGN CRITERIA

The Sakai community formed around the recognition that an environment to support academic practice and collaboration should not place artificial barriers between the component parts of that practice; learning, teaching and research. Thus, the existing Sakai online environment is termed a 'Collaboration and Learning Environment' (CLE), rather than a Learning Management System (LMS), Virtual Learning Environment (VLE), or Virtual Research Environment (VRE). Whilst the current Sakai environment is widely deployed as an LMS, and contains a toolset appropriate to that deployment, the environment has other features, such as support for project activity and JSR168 Portlets, which are more commonly found in Virtual Research Environments and portal frameworks. The current widespread use of Sakai for the support of learning should not obscure this broader use. Institutions using Sakai for purposes of research collaboration include a cluster of Australian universities; Melbourne, Monash, and the Australian

National University, together with a number of universities in the United Kingdom, including Cambridge, Lancaster, Bath and Newcastle.

In the eight years since the initial design of the Sakai CLE much has changed. Policy and economic imperatives have driven changing patterns of student recruitment in many countries, significantly increasing the scale of online or blended learning, the range of types of students participating in online courses, and their geographical distribution. The same period has seen much early experimentation with online and distributed research collaboration in Virtual Research Environments, although little consensus has emerged on their general characteristics beyond that elaborated by one of the coiners of the term [1]. Above all, the nature of the web has also been transformed by the considerable growth of social networking software, and what some have termed "Web 2.0" during the same period.

III. The Sakai OpenAcademic Environment Design Goals

The key design goals of the new Sakai environment, The Open Academic Environment (OAE), reflect these environmental changes, and the experience of the community itself. Several broad considerations inform the design goals for the new environment. These include:

A fundamental premise of the OAE design process is that education and research are inherently collaborative endeavors that should lend themselves naturally to the social web. Education and research are not, however, a perfect fit for today's consumer web social tools. Educational institutions have obligations to protect privacy; simply suggesting the use of Facebook or similar tools, is, as a consequence, potentially fraught with legal and ethical issues. Similarly, scholars are not so much concerned with discovering "friends", as establishing and building relationships with collaborators in the same or related disciplines, or parallel areas of interdisciplinary interest. The project to build the Sakai OAE has therefore set out to establish the use of social networking-like capabilities for academic purposes, and in an academic context.

The current conventional Learning Management System tends to be constructed around the centrality of the course site or virtual classroom. In contrast to the constraints of these "virtual classroom walls", Sakai OAE sets out to build around the centrality of individuals collaborating in a group, rather than the centrality of a site or course.

The typical Learning Management Systems builds on the extension of the course or classroom into virtual space by replicating a typical lecture and tutorial model. This tends towards a very hierarchical perspective of content within such an environment. Sakai OAE inverts this approach by assuming by default that every participant is an author, collaborating to create content, rather than viewing the educator as the principle or only content creator.

The rejection of a "walled garden" approach in an online environment speaks to a requirement that the environment more easily allows content to flow between that environment and the open web. Sakai OAE therefore seeks to be selectively and securely permeable, allowing content and activities from the outside world in, as well as from inside the system out to the rest of the web.

Sakai OAE Design Goals[2]
Sakai OAE is a scholarly space for research, teaching and learning. Designed by educators for educators, the Sakai OAE's design goals are:

A user-friendly experience informed by the needs of learners, teachers and researchers.
Powerful capabilities for content authoring, sharing, and reuse.
Group collaboration and social networking within an academic context and around learning, teaching and research interests.
Learner, group and activity focused. Not site centric.
Ability to create and deliver cohesive learning experiences and portfolio processes, and promote student engagement.
Highly configurable and scalable at both small and enterprise levels.
Ease of software development and flexibility for integration.
Agile Sakai code base built atop other reliable open source projects.
Open source license. Open educational resources.
Interoperability with other systems.

The core functionality of the Sakai OAE have been broadly defined and mapped by the Sakai teaching and learning community in a series of teaching and learning lenses

IV. Technology Perspectives

The underlying technology approach of the project is to attempt to re-use components from open source communities outwith education, or which span education and other sectors. This represents an attempt, which has yet to be fully validated, to focus software developers employed by Higher Education institutions squarely on academic requirements, rather than on the unnecessary wheel-re-invention of general features. Such component re-use includes the Apache Felix[3] implementation of OSGi[4], and Apache Shindig[5] implementation of the OpenSocial[6] Specification. Apache JackRabbit[7] was examined from the perspective of content storage, but proved unsuitable for an environment based on high-demand, secure authoring with multiple authors. In February 2011, the project consequently switched to a sparse approach, providing bindings to Apache Cassandra[8]. Further drivers are available for those wishing to deploy against a more conventional relational database using JDBC, such as MySQL, Derby and Oracle. Apache Solr[9] provides the OAE full-text indexing and search capabilities, introducing interesting parallels to systems providing access to materials in the repository or library world.

V. Transition between Environments

Transition between Learning Management Systems, or even versions of the same system, is time consuming and potentially disruptive to the core activity of an academic institution, affecting, as it does, significant numbers of both academic staff and students. It is not untypical for such transitions to take twelve to twenty-four months or more. As has been noted, the Sakai OAE is, in many respects, a radical departure from the conventional LMS, thus compounding the issues surrounding transition for institution and users.

The Sakai Community has evolved an approach to this problem space that rests on enabling an institution to run both Sakai CLE and Sakai OAE side by side, with a significant measure of user-facing integration. This approach is termed Sakai OAE “Hybrid Mode”. Hybrid Mode uses the IMS Global Basic Learning Tools Interoperability (LTI) Specification [10] to surface Sakai CLE worksites and tools in meaningful ways, determined by the institution or user, within the Sakai OAE, which is designed to be more composable, and to facilitate a greater degree of personalization. By these means, an institution may choose to extend its online capability by combining previous course or site-centered approaches with those based around the newer paradigms outlined above. An institution can therefore chart a forward course between environments with a minimum of disruption, and under its own direction at a pace determined by its own strategic imperatives.

VI. Community, Governance, Foundation

The development of the Open Academic Environment is driving significant change within the Sakai Community. The Sakai OAE Project, established in July 2010, is a collaboration between Cambridge University in the UK, and Charles Sturt University in Australia, together with New York University, the University of Michigan, the University of California at Berkeley, Indiana University, Georgia Tech and the American Academy of Religion. The project conducts its work in the open [11], publishes software under an OSI-approved [12] open source license, the Educational Community License[13], under the guidance of the Sakai Licensing Working Group[14].

The project to create the Open Academic Environment is initially a managed project, with direct financial and human resource contributed from the participating institutions to the project via the Sakai Foundation. Financial resource has been used to hire contributors in a number of scarce areas, including User Experience design. Project structure and practice is modeled on an application of agile techniques to highly distributed international teams. Teams are organised across institutions, a break with the former practice of the Sakai community, which has tended towards institutional “ownership” of specific tools or tool clusters. At each point, from design to realization, a User Reference Group (URG) validates the approach taken. The URG is made up of practicing academics and learning technologists from investing institutions.

Sakai Open Academic Environment Managed Project governance is constituted with a Steering Group comprising representatives of investing institutions, which is jointly chaired by an institutional representative and the Sakai Foundation Executive Director.

The Sakai Community is moving from the support of one software environment to two environments, at very different stages of a product lifecycle. This movement is generating further consideration of the role of the Sakai Foundation itself. There is a tendency to move the Foundation back from direct involvement in specific projects, and refocus it as a community facilitator, providing human and technical infrastructure, licensing and legal expertise, and brokerage in enabling community-driven projects. The latter concern is of particular significance for a community with a resource pool that is growing, but not growing in proportion to the work it seeks to undertake.

The latter consideration has sparked an extended conversation with another organisation supporting a number of Higher Education focused software projects and products, the Jasig[15]. Both organisations have charted a course towards organisational fusion, and have begun to map a common value proposition [16] for a common foundation nurturing software development projects supporting the academic mission. The new organisation that results from the joining of the Sakai Foundation and Jasig has the potential to be a significant factor in both innovation, and software sustainability in the service of education.

References

[1] Fraser, M; Virtual Research Environments: Overview and Activity, Ariadne Issue 44 July 2005, http://www.ariadne.ac.uk/issue44/fraser/

[2] https://confluence.sakaiproject.org/display/3AK/Design+Goals

[3] http://felix.apache.org/site/index.html

[4] http://www.osgi.org/Main/HomePage

[5] http://shindig.apache.org/

[6] http://www.opensocial.org/

[7] http://jackrabbit.apache.org/

[8] http://cassandra.apache.org/

[9] http://lucene.apache.org/solr/

[10] http://www.imsglobal.org/toolsinteroperability2.cfm

[11] https://confluence.sakaiproject.org/display/3AK/Project+overview

[12] http://www.opensource.org/

[13] http://www.opensource.org/licenses/ecl2.php

[14] https://confluence.sakaiproject.org/display/LIC/Sakai+Licensing

[15] http://www.jasig.org/

[16] http://www.sakaiproject.org/news/jasig-sakai-common-foundation-value-proposition

Mod-late: an innovative model for the construction of web templates for CMSs

G. Dimauro, P. Quintavalle, S. de Nichilo
Dipartimento di Informatica - Università di Bari, Italy

Abstract – **In this paper we propose an innovative model called Mod-late, whose main aim is to automate the process of web template creation and customization of the overall structure of the web pages.**
Mod-late describes how to produce fully customized web templates by standardizing the construction tasks.
These templates, both simple and complex, are created in order to be exported in packets which can be interpreted by the CMSs on the web or ready to be stored in a fully modifiable format.
Our model was designed in open mode and allows full compatibility with next-generation CMSs and E-Learning platforms.

Key words: Content Management System, Web Template, E-learning

I. INTRODUCTION

In the latest years the Internet has constantly changed and it does not stop so that there is someone who talks already about the Web 3.0 and who tries to put the base for the web of the future, based on the semantic and the natural language processing [1].
Despite the growing use of CMSs their maturation has not reached a turning point and the cause must be sought in the gap between the skills required for their use and the knowledge of the average users.
The area concerned is the Web 2.0 where the user is no longer just reader but a producer of networked information. “You can see the Web 2.0 as a set of principles and practices that tie together a veritable solar system of sites that demonstrate some or all of those principles, at a variable distance from that core” [2].
In this solar system several CMSs present the same problem: the only way to structurally interact with their web template is through the technical changes as well as with the E-Learning platforms.
As it is known, the Content Management System is a tool designed to publish content to the Web without knowledge of HTML in a controlled environment that ensures a consistent look and feel across all the pages of a web site [3]. A CMS is not really a product or a technology. It is a catch-all term that covers a wide set of processes that will underpin the "Next Generation" large-scale web site [4].
The state of the art about the context of automated web templates creation for CMSs and E-Learning platforms is not so advanced. “Artisteer” is the only Web design automation product that instantly creates fantastic looking Website and Blog templates. It is a powerful software that makes the user a professional Web designer of Websites [5].
Starting from a basic theme, it is possible to create visually appealing templates even if they are not structurally very customizable.

II. AN INNOVATIVE MODEL

The model proposed in this work aims to automate the process of creating and editing web templates. This way the user can save time creating web sites using a CMS as well as creating and editing the E-Learning platforms view.
Additionally the implementation of a software system based on our model, and its subsequent use, leads to standardize the processes and results in the creation of web templates, increasing the percentage of compliant templates.
The distance between the web users and the CMS and E-Learning platforms is very important. These systems are common on the internet and in order to open them new doors the ideal solution would fit an application between the users and the structure of the web application they prefer to use.
The purpose of Mod-late is to describe how to allow the users to build web templates and to export them towards the desired platform. Mod-late is fully compatible with such platforms: according to the guidelines dictated by the model is possible to produce web templates for every web application.
The model describes a particular generator and operator of HTML, PHP and CSS which is expandable by adding logics in order to import / export web templates for the system that the user wants to look out.
For a web programmer the model describes the techniques to quickly produce web portals. For a common user the module describes a powerful tool for creating web templates associated with the publishing tool that he prefers.
The model is expandable by simply integrating the modules which describe how to import and export web templates.

III. MODEL DESIGN

The model is structured on a three-level architecture:

- The presentation level, designed to handle the interaction of the model with the outside world;
- The domain level, which includes a set of business rules;
- The data level, which deals with the continuing entity treated.

The term “entity” means a set of objects used by the model to build up web templates.
Three type of entity are described:

- An entity type which can contain any other entity (Container);
- An entity type which can horizontally divide the other entity types (H-Divisor);
- An entity type that divides vertically any entity (V-divisor).

The domain level is divided into independent modules used for particular functions.
Each module must have a low level of coupling with other modules and a high internal cohesion while respecting the principle of modularity. Modularity is required to allow the developer to easily implement a system based on the model. It substantially simplifies the design process by substantially simplifying the individual components to be built [6].
At the edge of domain level two managers must be placed to enhance the modularity between the levels of the proposed model.
These managers may be called “StructureManager” and “PresentationManager”.
The "StructureManager" puts in communication the the data level with the domain level while the "PresentationManager" must have as its purpose the management of the information exchanges between the domain level and the presentation level.
The core modules of the domain level are the *Hierarchy module* and the *Representation module*.
The *Hierarchy module* describes how to create the hierarchy of entities of a web template. This module requires that the root of the hierarchy is a container entity. All entities should be included in the root container. Every incorporated entity must be adapted to its container entity; it can just keep distance from the limit of the container. Entities that are on the same level of the hierarchy cannot overlap.
The *Representation module* outlines how the entities of the model can be represented.
The main aim is to enable the creation of web templates, so the geometric shape that could represent the entities of the model is the rectangle.
The location of the figure must be well defined on the Cartesian plane. There are no restrictions on the use of rectangles to represent entities
The developer who implements the model may decide to adopt any kind of geometric figure to represent the model entities. The only condition required is that the chosen figure must be genuinely representative.
Instead the presentation level is not particularly descriptive: it establishes only that a good presentation must include all the necessary elements to display the hierarchy of entities.
This view must respect the hierarchy built in the *Hierarchy module* and must additionally respect the type of representation chosen in *Representation module.*

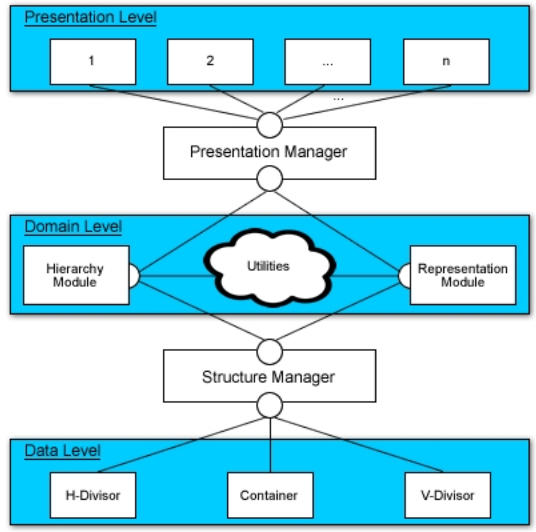


Figura 1 - Model Architecture

IV. MODEL IMPLEMENTATION

The proposed model was applied in the development of an application of web templates creation.
This basic application is developed following the object-oriented paradigm and it works primarily by using the entities described in the data level of the proposed model and with the implementation of the described modules. This application is able to produce web templates and export them in complete interpretable formats.
One of the most used CMS today is *Joomla!*. It is a CMS developed using PHP and open source published under “GPL v.2” license. It is at the forefront of one of the most dynamic trends in open source software: the growth of open source content management” [7].
Because of its characteristics "Joomla!" was used as an example for the development of the system that implements the model proposed in this paper.

V. SYSTEM DATA LEVEL

The main aim of the data layer is to manage the system entities. These entities result from the model description as follows:

- Container, entity used to hold every other entity;
- Band, which takes care of splitting horizontally the other entities;
- Sidebar, entity that creates vertical subdivisions in the entities.

During the design phase of the model-based system the entities have been further extended.
The band has been specialized into *FreeBand* and *ModulesBand*: the *FreeBand* can contain each type of entity while the *ModulesBand* may contain only a limited number of adjacent containers.

During the export phase inside of these containers the chosen *Joomla!* modules will be placed.
The Sidebar instead has been specialized in *LeftSidebar* and *RightSidebar* to simplify its positioning.
Every entity described by the model, within those entity extensions, was carried out in order to have all the properties to be represented on the web.
Among the properties it is important to underline: width, height, margin, padding, background image, background color and borders.

VI. System Domain Level

The domain level is the business services environment which allows the presentation layer to communicate with the data layer.
In the development of the domain level two basic modules have been implemented: the *Hierarchy module* and *Representation module* which are used by the managers to communicate among themselves and with the presentation layer.
In order to implement the hierarchy in the relative module a tree data structure has been used.
With regard to the *Representation module* the developers have decided to use rectangular shapes for the representation of the entities.
During the analysis phase some needs have come up, needs that led to the creation of various modules in the cloud of the *Utilities*.
First, the need to translate the system in multiple languages has led to the creation of the Localization module through a series of parameter files (*.ini) communicates the text data translated to the presentation layer.
To manage application-level events and to configure the system two additional modules have been implemented: the *Application module* and the *Configuration module*.
The modules implemented inside the utilities cloud communicate with each other as much as they can communicate with adjacent modules (Hierarchy and Representation) only through the operators.

VII. System Presentation Level

The presentation layer deals with the representation of the entities created by the hierarchy module using the Hierarchy Level Domain and the rules of representation given by the Representation of the same level form of architecture.
The presentation layer is completely independent from the domain level in order to respect the principle of modularity on which the model is based. This level is not divided into modules such as the domain level but has several interfaces that allow the management and system configuration and the creation of web templates.
The essence of the presentation layer is the preview screen that is useful to show the user what is actually being built.
In order to make the work more easy for the experienced users a property window has been placed. Such a window shows the properties of the entity the user selects in the preview so that it is possible to interact regularly with them without navigating menus.
It was finally introduced a view of the hierarchy created. From this hierarchy it is also possible to insert or delete these entities while respecting the rules of the model.

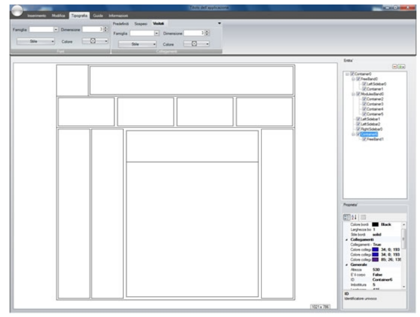


Figura 2 - Possible layout

The division of web templates, as shown in Figure 3, in separate areas is useful for two reasons:

- first, the good structure of the web site or portal that you want to build;
- each area can also be set as an area for placement of the contents of the CMS.

The user can then decide on the exact locations where the content has to appear on its website, menus, forms and more. He becomes a master of the CMS to which exports as able to interact with the framework of the CMS itself.

VIII. Test Phase

The test phase was carried out concerning the creation of templates for web sites and portals built with *Joomla!*.
To prepare these tests were collected from web various web templates already constructed and sample images of the web templates to be created from scratch. It was not adopted any policy for the collection of research material to make the results independent from the conditions that constrain the choice of web templates.
Several experiments have been performed to modify existing web template and creating new web templates. From these experiments the following results were obtained.
To edit a web template, it is needed for an advanced user from 10 to 30 hours, which should be added to the other if the user is willing to radically change the positions in which information is exposed.

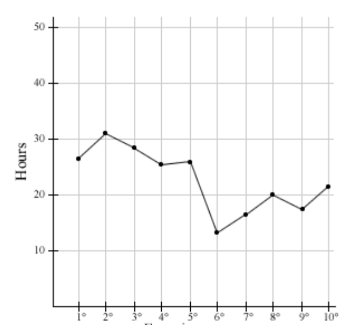


Figura 3 - Joomla! Template Edit Trend

The creation of a web template from scratch however, requires knowledge of HTML, CSS, PHP and frameworks of "Joomla"; in this case about 20 hours to achieve the desired result are necessary.
The necessary time for the creation of web templates in the experiments does not include the creation of graphics, but only the formatting of text, images and structure.

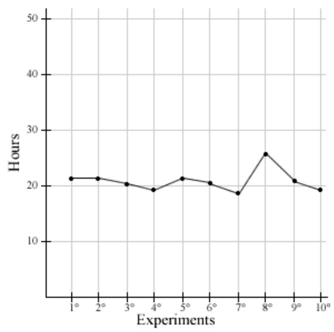


Figura 4 - Joomla! Template Creation Trend

After collecting the data about the creation and the modification of web templates for *Joomla!* several experiments were carried out on the application which implements the model proposed in this work.
These experiments are aimed at creating web templates and do not include the processing of graphics in order to make true comparisons with results in the creation of web templates from scratch, shown above. The experiments were also carried out on several people including an experienced user and an average user.
The results of the experiments performed with the system based on the following model are very interesting.

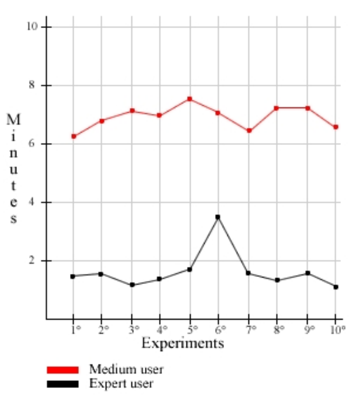


Figura 5 - Template creation through model based system

Action requie	Mean time
Building layout with header, left sidebar, and footer area readers.	40 seconds
Builder with several specialized areas and colors.	60 seconds
Creating web templates with margin settings, padding and colors.	200 seconds

IX. COMMENTS AND CONCLUSION

The model brings a great innovation in the development of web sites and portals.
Unlike the main concepts of “Artiseer” application, the center of the state of the concerned application domain, Mod-late makes significantly faster the creation and customization of the layout.
The Mod-late innovation leaves behind the state of the art so that it can be implemented by a new class of applications, not existing yet.
It is important to underline that the time savings for the production of a basic *layout* in the experiments performed is interesting. A new structure can be created in few minutes and do not require long learning time as the operations to achieve the desired outcome are few and simple.
Allowing the creation of web templates very quickly and so easily it is possible to eliminate much of the barrier between the Web and average users who could be transformed into real protagonists of the design.
This approach could mean an advancement of the use of CMSs as well as the use of E-Learning platforms and the reduction of the costs in a market where such systems abound.
Taking the study of CMSs and web templates to a more advanced stage it could be possible to make the model more precise giving more details about the description of the hierarchy creation. It could be eventually possible to look for commonalities between several systems in order to describe an additional part of the model dedicated to an abstract import / export policy for web templates.
The next step will be the implementation of a complete software system based on Mod-late that will result useful to test the power of the model and to give more precise results.

X. REFERENCES

[1] Prati G., *Web 2.0 Internet è cambiato*, UNI Service, Trento 2007.
[2] O’Reilly T., *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*, O'Reilly Media, Sebastopol (CA) USA 2007.
[3] Budziak S., *Strategies for managing your Web content*, July 2003.
[4] Browning P. & Lowndes M., *JISC TechWatch Report: Content Management Systems*, September 2001.
[5] Artisteer Web design revolution, *Artisteer User Manual*.
[6] Bryson J.J., *Intelligence by Design: Principles of Modularity and Coordination for Engineering Complex Adaptive Agents*, Massachusetts Institute of Technology, September 2001.
[7] Shreves R., *Joomla Bible*, Wiley Publishing, Inc., Indianapolis, Indiana, 2010.

Tool and Method for Evaluating Students Working on E-Learning Platforms

Luigi Romano
CASE, Faculty of Computer Science
Free University of Bolzano-Bozen
Bolzano, Italy
e-mail: luigi.romano@stud-inf.unibz.it

***Abstract*—**This document describes an ongoing project for defining a methodology and building artifacts that will help tutors in the evaluation process of students that work intensively on an E-Learning platform.

I. Introduction

In the last twenty years it has been a change in the educational scenario, such as, the diffusion of the Constructivist Didactical Approach [1][2][3][4][5][6][7]. The non-Constructivist, traditional didactical style is based on a "one way" interaction among teacher and students. The new role of a teacher (called tutor, in Constructivist terminology) is to support students in their experimentations and discussions about real problems, and this is the main way for building new knowledge as claimed in [1][2]. Wenger [16] claims that creation of a Community of Practice (CoP), i.e. a group of people that works together on a shared goal, is a necessary requirement for the creation of new knowledge. The Tutor's new role will be to foster the creation of a CoP in class. "Fig. 1" represents this shift in the teacher's role.

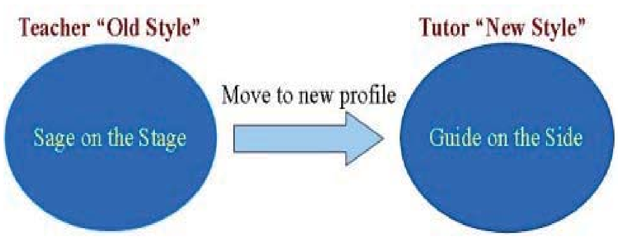


Fig. 1 New Didactical Role of the Teacher

The modern E-Learning Platforms (ELP) do not create obstacles to the Constructivist Methodology (CM), rather, but they offer to both students and tutors the opportunities of interacting and sharing concepts and comment them. The tutor has to assess students' performance referring to: the level of collaborative activities (CA), the level of new knowledge built by students, and their results in tests.

In this paper I will describe a research project to validate a methodology, and then, to build software that could help Tutors evaluating a community of students that works both in classroom and at home using an ELP. The paper shall explain in Section II the Goal and Metrics that it will be used as theoretical bases for the evaluation methodology. Then in Section III I will briefly describe a Constructivist Approach to a real problem of Solid Euclidean Geometry. In Section IV I will describe the evaluation methodology, in Section V I summarize the contents of the paper and I briefly describe the future developments.

II. Goals and Metrics

My research project is focused on building tools and methods for evaluating students working in class as well as on an ELP. I plan to assess my research by means of an experience that shall last for a relatively long time. Students' activity shall be monitored for one entire class of 30 students, aged fifteen to sixteen, over three terms, starting from the second term of 2nd year of Liceo Scientifico (Italian high school with specific reference to Mathematics and Physics) till the end of their 3rd year of study. The subject area involved shall be Mathematics. In details, students will work on Euclidean Geometry for the first part of the activity, and on Analytical Geometry for the second part of the activity. Results achieved by students will be part of the final evaluation. Outside the school, students shall be given homework, as it is customary in Italian schools. In such home-based activity, students shall work on an online open source platform named DIEL [8] with a 3D graphical interface. I decided to use DIEL because its graphical interface is usually well accepted by today's students, who are digital natives [28], and because its features help the tutor to monitor students' activities. Metrics for assessing students will be integrating both *in itinere* and final evaluation data from three sources:

- The Social Network Analysis (SNA) [10][11][12][13][14] of students' collaboration patterns, as they are emerging from students exchanges of information.
- Results achieved by students on short tests, passed to them without a previous advice. These tests are common practice for preserving students' attention, and as intermediate evaluation. They are known as Pop Quizzes (PQ) [25].

- Results achieved by students at the term examination (final test), done in the regular class time.

PQ and final test are a quite standard assessment mechanism for students not needing further description. On the other hand, the adoption of SNA in high schools is not widespread, so a few considerations should be added. SNA is a methodology used for monitoring the behavior and the relationships established inside a community of people. One of the most common structures, described by SNA, is the "clique", that is, a group of people that has strong relationships with one another. The rationale behind the use of the SNA in assessing students' CA is to check for the presence of cliques inside their working groups. The strong relationship between CoP and clique and the idea of the creation of new knowledge inside a clique is supported by several authors like [16][17][18][19][20][24], quoting [20] "… Cohesion is a primary network structure that contributes to the creation of Knowledge: shared beliefs and behaviors. Cohesion is manifested by the existence of cliques of participants who are connected internally more than externally. Members of a clique tend to create knowledge by virtue of their strong intra-responsiveness relations …".

Another important value is the Degree of Centrality. SNA is directly related to the number of ties between someone and other people in the Network. A high number of connections mean a high Centrality degree, as claimed in [11][15].

III. A Case Study of Collaborative Working Activities

A Socio-Constructivist approach requires that students face real problems and work collaboratively to overcome them for improving their skills. This team-working could be achieved more effectively using a "road map" to support students. Polya [21] lists four main steps of his Problem Solving Strategy (the road map):

1. Understanding the problem
2. Devising a solution plan
3. Carrying out the solution plan
4. Looking back to the results achieved

To this respect, any particular choice of problems that are proposed to students would do, provided that they work in such a sequence of steps, and that the tutor facilitates them to proceed along this road map. A typical problem to be proposed to students for their CA is the following.

Example: Problem of Architecture and plane Euclidean Geometry

"Fig. 2" shows a Cubba (or Cuba), a typical Byzantine architectonic element very common in Sicilian Churches. Sometimes the upper part of the Church is a perfect half-sphere or a spherical sector.

Question: How could we find the center of the sphere in both cases, i.e. the half-sphere and the spherical sector?

Realistically, the problem to be discussed in class will be more complex. The process that students will put in practice is described in "Fig. 3".

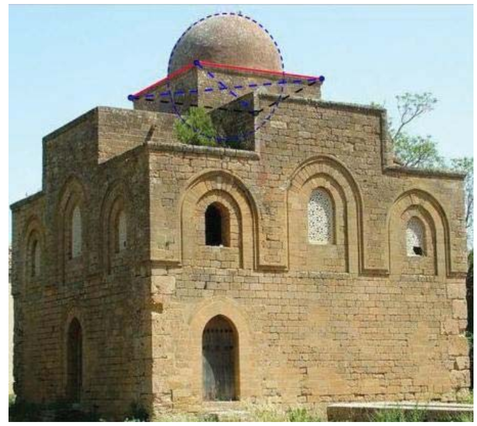


Fig. 2 Example of Collaborative Work using Euclidean Geometry: The Cubba on the Church of Holy Trinity of Delia in Castelvetrano (Sicily, Italy)

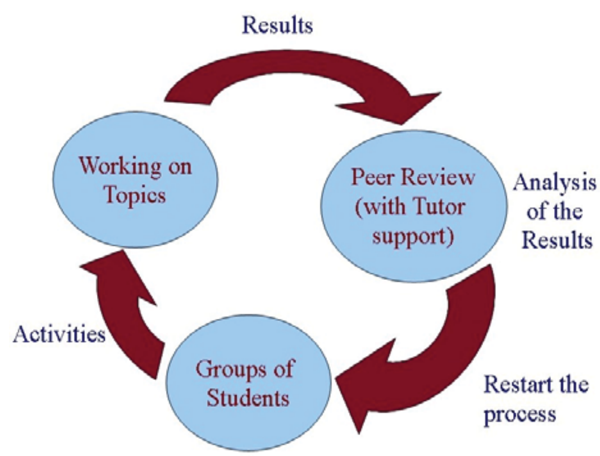


Fig. 3 Diagram of Collaborative Knowledge Construction process made by Students using the "roadmap".

The Solution

Students will apply the suggestions coming from each step of Polya's road map.

Concerning Geometry, students should discuss about how to find the center of the sphere having only the information on its spherical sector. This means inferring geometric properties from the Hypotheses. Reaching consensus about what has to be proved should involve the whole team of students. The complete description of each single step and action is outside the focus of this paper. Let us focus just on some aspects inside this scenario:

- Groups of students are working together.
- All activities start in the classroom, students must keep on working as homework using the on-line platform DIEL.

- Students shall be pushed to follow the road map by suggestions coming from the tutor, whenever they are diverging, in order to get an effective collaborative work.
- Interactions with tutor are available both in class and on the on-line platform.
- At the end of the solution process, all students will be rethinking their solution, in order to understand if the result might have been achieved in a different (possibly better) way.

To complete the case study, let us consider a sample Pop Quiz related to the above topic:

"*Describe the properties of chord, what about the perpendicular drawn from the center of the circumference to the chord itself? Is it useful for a sphere?*"

Distributing PQ to students could be achieved either during class time, or while they are on-line, by a random selection performed by DIEL.

IV. THE EVALUATION METHODOLOGY

During the three terms of the experience, our Evaluation Methodology will be used not just for proposing a final grade, but – more interestingly – as a set of guidelines that help the teacher to understand the progress of the individuals and of the entire class. Such guidelines foresee the evaluation of the following quantities (measures):

- Did each student work proficiently inside his/her team? This will be named Level Of Collaborative Work (LOCW).
- Did each student work proficiently on the proposed topic? Data gathered from POP-Quizzes and from the Final test will be used in the evaluation. This will be named Level Of Results (LOR).
- Did each team create a CoP? This result will be measured as CCOP (Creation of Community of Practice)
- Did the whole classroom create a CoP? This result will be measured as GCCOP (Global Creation of Community of Practice).

While each teacher has very clear how to collect and assess data in (b), results from SNA will be used for answering to questions (a), (c) and (d), mixing data about the creation of cliques, the degree of students' centrality etc. Authors like [31][32][33] and others, have suggested a content analysis on written material present in the ELP, but I prefer to avoid using it because there is not a universally agreed upon method for assessing results, as stated in [27].

The Evaluation Tool is an artifact that will suggest to the Tutor a graphical representation of ¨how well¨ the entire class is progressing. The representation is sketched in "Fig. 4", where the tool helps the Tutor to have a glance at understanding the students' performance. The Diagram in "Fig. 4" is interpreted in this way: The tool plots a point in the graph showing the positioning of each student with respect to his evaluation metrics, and the goal of the teacher should be ideally to have the entire class falling into boxes labeled 5-6-8 and 9.

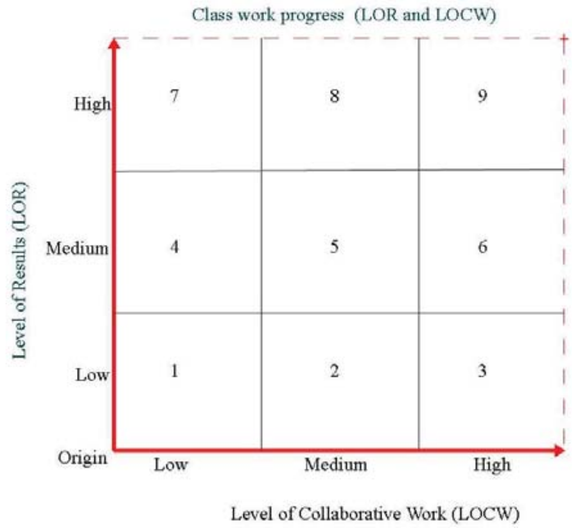


Fig. 4 Students' performance evaluation graph

Boxes (1) (and (9)) represent clearly identified, extreme situations. Students listed in such boxes are very poor (resp. excellent) both in their cooperative activities (CA) and personal understanding of topics. Boxes (3) and (7) are showing students in intermediate situations: excellent in CA, but really poor in understanding of the topics, and the opposite. Similar considerations will hold for the analysis of the behavior of students falling in the remaining boxes.

How to fill in automatically (by means of the tool which we shall develop) the above table is still under discussion. Intuitively, CCOP and GCOP will be used to ¨move up and to the right¨ the position of each student, with respect to the position he-she gained with PQ and final test, but only if his-her team of students and the whole class created a CoP. In the beginning data about SNA, PQ results and Final Test will be inserted manually in the Evaluation tool, and when the whole environment will be consolidated, data will be loaded automatically from the ELP. Given the fact that three term evaluations shall take place during my experience, I shall perform assessment of activities mostly manually for the first term, trying to progressively consolidate the methodology and to automate them in an increasing way.

V. CONCLUSIONS

In this article I described a research project aimed at building a methodology for evaluating students working on an ELP. The emphasis of my project is to support an intuitive graphical representation of class progress in a given topic, which is visualized by a 3x3 matrix. Dimensions of this matrix are based both on the outcome of individual study and on the degree of participation to CA, in line with constructivist approaches. In parallel with development and assessment of the methodology, I will build software for automating the creation of such an artifact, so that the tutor is constantly aware of students' on-line activities, and may take corrective actions to support the entire class and individual students. The metrics employed to build such an artifact are based on data gathered by SNA, Pop-Quizzes and Final Tests. In this paper I have described the rationale beyond the use of SNA for understanding the level of CA made by students. Today most high school students are spending several hours in school, plus do some homework at home, so they can be considered working in a blended learning environment. The use of an ELP seems not to be best suited for such a situation. Yet we believe that the effectiveness of the approach we propose can be profitably used to gradually shift the focus of study activities from strict tutor control (in class) regulated learning pattern (on the ELP), a skill that will be a must in tomorrow's society. Only students who learned to work collaboratively will face successfully tomorrow's life long learning perspective. Even today, as noted by [30], students that live far from the school or that suffers for a long illness are mostly working alone, and they usually feel a sense of isolation from the community of students and tutors. The use of the ELP in a similar scenario could improve their achievements and help to reduce scholastic renunciation.

REFERENCES

[1] J. Bruner, "Going beyond the Information Given", New York: Norton, 1973.

[2] J. Bruner, D. S. Wood, G. Ross, "The role of tutoring in problem solving", in Journal of Child Psychology and Psychiatry, volume 17, Pergamon Press, 1976.

[3] D., H., Jonassen, "Thinking Technology: Toward a Constructivist Design Model", Educational Technology, Vol. 34, 4, pp. 34-37, 1994.

[4] J. Piaget, "The grasp of consciousness: Action and concept in the young child", Cambridge, MA: Harvard University Press, 1976.

[5] L. Vygotsky, "Mind and Society", Cambridge, Harward University Press, 1978.

[6] L. Vygotsky, "Thought and Language", Cambridge, MA: MIT Press, 1962.

[7] S. R. Hitz, "Collaborative Learning in Asynchronous Learning Networks: Building Learning Communities", Invited Address at Web 98, Orlando, FL. 1998

[8] F. Di Cerbo, G. Dodero, P. Forcheri, V. Gianuzzi, M. G. Ierardi, "Hybrid Learning Experiences with a Collaborative Open Source Environment", Proc. of ICHL 2nd Int. Conf. On Hybrid Learning 2009, F.L. Wang, J. Fong, V.K.S. Lee (eds.), Lecture Notes in Computer Science, vol. 5685, Springer 2009, pp. 45-54.

[9] http://moodle.org/

[10] S. P. Borgatti, A. Mehra, D.J. Brass, G. Labianca, "Network Analysis in the Social Sciences Science", 323, 5916, pp. 892 – 895, 2009.

[11] S. Bonacich, "Power and Centrality", American Journal of Sociology, vol. 5, pp. 1170–1182, 1987.

[12] L. C. Freeman, "Centrality in Social Networks Conceptual Clarification", Social Networks vol. 1, pp. 215–239, Elsevier Sequoia (eds.), 1978.

[13] J. Scott, "Social Network Analysis: A Handbook", 2nd ed., London, Sage (eds.), 2001.

[14] S. Wasserman, K. Faust, "Social Network Analysis", Cambridge, MA: Cambridge University Press, 1994.

[15] R. A. Hanneman, M. Riddle, "Introduction to social network methods", Riverside, CA: University of California, Riverside, 2005

[16] E. Wenger, "Communities of practice: learning, meaning, and identity", Cambridge University Press, 1998.

[17] A. Reuven, E. Zippy, R. Gilad, "Cohesion and Roles: Network Analysis of CSCL Communities", Proc. of the 3rd IEEE International Conference on Advanced Learning Technologies (ICALT'03), 2003.

[18] A. Reuven, E. Zippy, R. Gilad, A. Geva, "Network Analysis of Knowledge Construction in Asynchronous Learning networks", JALN Volume 7, Issue 3, 2003.

[19] C. Haythornwaite, "Networks of Information Sharing among Computer-Supported Distance Learners", Proc. of CSCL '99: The 3rd Int. Conf. on Computer Support for Collaborative Learning, pp 218-222, Mahwah, NJ: Erlbaum, C. Hoadley (Eds.), 1999.

[20] R. S. Burt, "Structure, A General Purpose Network Analysis Program. Reference Manual", New York: Columbia University, 1991.

[21] G. Polya, "How To Solve It", Princeton University Press, 1945.

[22] D. D. Suthers, R. Vatrapu, R. Medina, S. Joseph, N. Dwyer, "Conceptual representations enhance knowledge construction in Asynchronous Collaboration". In Computer Supported Collaborative Learning 2007

[23] J. D. Novak, D. B. Gowin, "Learning How to Learn", Cambridge University Press, (Cambridge), 1984

[24] C. Reffay, T. Chanier, Social Network Analysis Used for Modeling Collaboration in Distance Learning Groups, S.A. Cerri, G. Gouardères, and F. Paraguaçu (Eds.): ITS 2002, LNCS 2363, 31–40, Springer-Verlag, Berlin, Heidelberg, 2002.

[25] V. A. Cicirello, "On the Role and Effectiveness of Pop Quizzes in CS1", Proc. of the 40th ACM, SIGCSE Technical Symposium on Computer Science Education, Vol. 41, Issue 1, pp. 286 – 290, 2009.

[26] http://www.analytictech.com/ucinet/

[27] B. De Wever, T. Schellens, M. Valcke, H. Van Keer, "Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review", Computers and Education Vol. 46, pp. 6 – 28, 2006

[28] M. Prensky, "Digital Natives, Digital Immigrants", On the Orizon, MCB University Press, vol 9, issue 5, pp. 1 – 6 , 2010.

[29] K. Kreijns, A. P. Kirschner, W. Jochems, "The Sociability of Computer-Supported Collaborative Learning Environments", Educational Technology & Society 5(1), ISSN, pp. 1436 – 4522, 2002.

[30] K. Kreijns, A. P. Kirschner, W. Jochems, "Identifying the pitfalls for social interaction in Computer Supported Collaborative Learning environments: a review of the research", Computers in Human Behavior, vol. 19, pp. 335 – 353, 2003.

[31] C. N. Gunawardena, C. A. Lowe, T. Anderson "Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing", Journal of Educational Computing Research, Vol. 17, pp. 397 – 431, 1997

[32] J. B. Pena-Shaff, C. Nicholls, "Analyzing student interactions and meaning construction" in Computer Bulletin Board Discussions. Computers & Education, vol. 42, pp. 243 – 265.

[33] A. Weinberger, F. Fischer, "A framework to analyze argumentative knowledge construction in computer supported collaborative learning", Computers & Education, Vol. 46, Issue 1, pp. 71 – 95, 2005

The TIE Project: Agile Development of a Virtual World Serious Game on Waste Disposal

Ida Bifulco, Rita Francese, Marco Lettieri, Luca Liscio, Ignazio Passero, Genny Tortora
Università degli Studi di Salerno, via Ponte don Melillo 1, Fisciano (SA), Italy

ibifulco@unisa.it, francese@unisa.it, mlettieri@unisa.it, lliscio@unisa.it, ipassero@unisa.it, tortora@unisa.it

Abstract

Virtual Worlds are largely adopted for creating educational games. In this paper, we describe the development process of a Virtual World-based game supporting a learning-by-playing approach to educate children on sustainable waste management principles. Because of the high uncertainty level of the project, a Scrum-based development process was adopted.

1. Introduction

Nowadays, there is a strong need of enforcing environmental education, by planning strategies and designing tools for environmental communication that effectively lead to changes in the citizen behavior. These activities, to be successful, have to involve the youngest citizens. The best way to capture the attention of this kind of learners is to adopt an edutainment approach. In particular, a serious game can be suitable, because it proposes an enjoyable experience, while training and educating users [23]. In particular, gaming environments can provide the means for a constructivist approach to learning. Constructivist approaches suggest that children should acquire knowledge through experience [13].

The development of a game suffers of the same problems of traditional software development due to the sequential approach and, in particular, to the performing of requirements analysis and detailed design before implementation. As in traditional software, also in the case of the development of games it is possible to follow agile software development processes, such as Scrum, FDD, Extreme Programming [10].

This paper describes the agile development of a serious game proposed in an ongoing project founded by the Regione Campania (Italy), named TIE project, aiming at presenting to primary school students the main concepts related to waste collection and garbage recycle. Learning is inducted through the exploration of an ad-hoc developed 3D virtual environment. During the exploration, specific educative games are proposed. The environment also supports collaboration and competition to better involve the students to use it in the time. Because of the high level of uncertainty, the game was developed following a Scrum-based development process.

The rest of the paper is organized as follows: Section 2 discusses related work; Section 3 describes the objective of the TIE project, while Section 4 illustrates the adopted development approach; Section 5, finally, concludes.

2. Related Work

City games are largely diffused. As an example, the Sims and SimCity games allow the users to play by creating and managing simulated communities and worlds, using an isometric representation. In particular, SimCity does not utilize individual game avatars as representations of the player and the environment is uninhabited. An example of simulation-based learning game is Math-City, supporting the students in the learning of mathematical concepts [14]. The progresses in mathematics are adopted by the students to grow their city. Recycle City [15] proposes a point-and-click web-based interface to teach environmental education. In our approach, we use a 3D world that can be explored by the user avatar and social mechanisms of collaboration and competition, adopting a learning by doing approach aiming at training students on correct environmental behaviors.

The adoption of agile methodologies to support game development has been investigated in literature, see, for example, [8][9][10][11]. In this paper, we applied the Scrum methodology to the development of a learning game based on a 3D Virtual World. The main characteristic of our project was the adoption of several opensource tools to create and manage the game and the high level of incertitude, due to the difficulty of their integration.

3. The TIE Project

The project TIE was founded by Regione Campania (Italy) and aimed at creating on-line learning activities related to environmental education for primary school children. This project is designed to assist school teachers in the education on sustainable waste management principles. In this way, even if the children are not yet involved in waste collection, they can contribute to influence their families and the broader community. In addition, behaviors and skills acquired in this period may have a long-lasting impact in later life. Thus, early childhood education has a relevant impact on the efforts to bring about sustainable development [21].

The educational program is based on the Waste Hierarchy [22]:

- *Avoidance*, including actions to reduce the amount of the generated waste;
- *Resource Recovery,* including reuse, reprocessing, recycling and energy recovery;
- *Disposal*, managing all options in the most environmentally responsible manner.

In particular, the project consisted in the development of:

1. multimedia lectures on all the three topics and the related auto assessment tests;
2. an on-line Virtual Reality serious game enforcing the acquisition of these concepts by involving the students in learning-by-doing activities.

The authors of the paper were responsible of the second activity, described in this work, while the first was in charge of the 3Dart partner, a communication and 3D consulting organization. Currently, the project is, at its final stages.

4. The development process

The project, as most game projects, has several uncertainty levels [9]:

- find the fun is not easy. Several iterations and experimentation may be needed;
- the results of the new technology adoption are often uncertain. In this project we plan to experiment a novel 3D game engine and several open source technologies;
- team talent is highly variable.

For this reason, we decided to adopt agile methodologies that were created for use in cycles of product development characterized by a high level of uncertainty.

The selected agile development methodology was Scrum, used for project management of software development since the early 1990s. According to Schwaber and Beedle [17], Scrum adopts the ideas of industrial process control theory in the software development. It is a flexible, adaptable, empirical and productive method, appropriate for small size projects with less than ten engineers and supports an iterative, rapid development style. Scrum concentrates on how the team members should work in order to flexibly develop the system in a constantly changing environment.

The Scrum methodology is composed of three phases: Pre-Game, Development and Post-Game.

4.1 Pre-Game Development

This phase consists in two sub-phases: Planning and Architecture/High level design.

4.1.1 Planning

Our project, as the majority of game projects, is characterized by a cross-functional team, composed of highly skilled individuals. In particular, the project team consisted of: a Scrum Master, three software developers, an ecologist and a primary school teacher, as domain expert, two primary school students (as final users), a graphic and a multimedia expert. In a greenfield engineering project, as ours, Scrum methodology suggests to work several days to build an initial Product Backlog, consisting of business functionalities and technology requirements.

The graphic and multimedia experts belonging to the 3Dart team had in charge the development of the cycle of multimedia lessons on the selected topic and participated to the Product Backlog activities to join the work of the two teams. Indeed, the scene and the learning experience of the traditional and the game approaches had to be related.

In particular, our Product Backlog was composed as follows:

- Definition of the game objective

The project presented to the Regione Campania did not deeply detail the game objective. During this phase, we better specified the aim and the scope of the game.

The game belongs to the Serious Game category. It is a simplified version of a Real Time Strategy (RTS) game. In this kind of games, players progress in a virtual environment, where resources are scattered on a map. A RTS game generally consists of three main phases: harvesting resources, building structures and competing with opponents.

The game, named *Pappi World*, is a 3D Virtual World game addressed to the 8-10 primary school kids to promote the practice of waste collection and garbage recycle. The activities are proposed to spread the didactic message and to reward participants who more closely adhere to the principles taught by the play actions and the simulations. Participant involvement is stimulated by adopting rankings mechanisms, game level progression and competition. Pappi World consists in the simulation of an imaginary virtual planet, where each school is represented by an archipelago, composed of islands associated to each class. Each student controls a village in his island (class). The game aims at stimulating the students to advance their village, their island and their archipelago by contributing to the improvement of the planet quality of life. The behavior of the students, according to the principles taught, determines to succeed in the game.

- Definition of the Educational Content

The game and the lessons are concentrated on waste collection and recycling. After a first iteration, a concept map representing the interaction among the contents to propose was prepared, involving all the team members and in particular the teacher and the ecologist.

- Definition of the learning approach

The game adopts a collaborative approach, when resource are exchanged with the other users, a competitive approach, when a user can collect the garbage of the others, visiting their village. Learning-by-doing is also stimulated by the individual games, where, as an example, users learn to properly separate the garbage and put it in the appropriate bin. The usage of own garbage enables the city to grow and increases the progress speed.

The main idea is to provide the concept that garbage is a resource and to effectively understand how correctly performing the garbage collection.

- Definition of the Virtual Reality Environment Requirements

These requirements deeply influence the selection of the technology and of the open source software to be adopted. In particular, the Virtual World should satisfy the following requirements [2][4]:

- *Synthetic*. It has to be generated in real time;
- *Tri-dimensional*. A 3D representation should be adopted to enforce the perception of realism, immersion and presence. The user should move in the environment, perceiving its depth.
- *Multi-sensory*. Moving in the environment should solicit several user senses, as vision, sound, space sense, depth.
- *Immersive*. The interface has to provide the perception to be really in the environment. This means that the user has to forget to look at the computer monitor.
- *Interactive*. The user inputs have to be detected in real-time and the virtual world should be accordly modified.
- *Realistic*. The elements of the worlds (trees, buildings, roads, etc.) should have a precision (number of triangles) that enables an efficient rendering and real time interaction for a user with a high band connection. This requirement is very uncertain. Only with the experimentation of the selected technology can be defined the trade-off between the performance and the resolution of the objects.
- *With presence*. It is the sensation to be part of the virtual environment, feeling of ‘‘being there’’ [3]. This expression denotes the perception of being in the place specified by the virtual environment, rather than just watching images depicting that place. The more this sensation is strong, the more the experience is meaningful. Thus, presence and learning are strongly related: increasing presence also increases learning and performance.
- *multi-users*. A student owning a village should visit the other student villages and they can meet.
- *metaphors.* The archipelago metaphor has to be adopted to represent a school, but the students can move only in a world represented by a village. The recyclable garbage is associated to a resource that enables the game level up, while the black bin is considered a danger.

- Definition the game functionalities

The Pappi World is structured in five levels:

- Level 0: the Pappi World,
- Level 1: the Archipelago (the school),
- Level 2: the Island (the Class),
- Level 3: the Village (the student land),
- Level 4: the Mini Games.

Each learner can act on Levels 3 and 4. When he registers, he has at his disposal a village that has a minimum set of street furniture, some houses and a playground, with the same appearance of the ones represented in the 3Dart video lessons.

Waste collection. By collecting and recycling the garbage the user will get some credits and special rewards that he can use for growing the village, by modernizing it and building new houses, community buildings like post offices and schools, add plants and decorations.

Neighbor Villages. A player can visit neighbors' villages and collect their garbage, obtaining precious resources for growing his village.

- Definition of the interaction modalities

The individuated interaction approach is the traditional point and click for navigating in world. A traditional 2D GUI should be adopted for enabling resource exchanges.

- Definition of the collaboration modalities

To create a new building or to upgrade an existing one, the player has to accumulate material to be recycled of several types (i.e., ten sacs of plastic, three of organic and four of paper to create a building of two floors). If he does not have the required material, but he has, for example, ten more sacs of plastic, he can try to exchange the material with his neighbors. In this way, the user learns that recycled garbage can be a source of income. It is also possible to visit the village of the other players and collect their garbage. This provides competition and incentives the users to clean their village to avoid losing precious material. In addition, this mechanism provides incentive to be often on-line and to collect not only own garbage, but also that of the others. To this aim, an appropriate mechanism has to be defined to propose the material needed for growing the village, using, as an example AI approaches.

It should be possible to write notes for the visitors and communicate with the other players using a chat or asynchronous communication.

4.1.2 Architecture/High level design

During this phase, we specified all the technologies we were going to adopt to implement the game and provided the design of the preliminary system architecture.

In particular, in this phase the following open source components of the architecture were individuated.

- Game Server

The game server initially selected was Darkstar, that was discontinued upon the close of the Oracle-Sun acquisition. In the next iteration, the selected server was the official community fork for Project Darkstar, RedDwarf [16]. It is a completely multi-threaded, event-driven transactional system that provides automatic data persistence. It is largely adopted for online games, virtual worlds, and social networking applications.

- Game Engine
  Due to the team member experience on Java programming, the selected game engine was the jMonkey Engine. Thanks to JavaVM, this engine satisfies the true cross-platform requirement, including a rendering engine for 3D graphics, a physics engine or collision detection (and collision response), sound, scripting and animation. The team had no experience in using this tool, thus its adoption could be reconsidered in the following.
- Development environment
  The selected development environment was NetBeans [12], because the Darkstar Project was originally proposed by Sun and both, JME engine and it, are well integrated in this development environment.
- The graphic elaboration pipeline
  Several graphic tools have been individuated, such as Google sketch-up, Blender, 3D Studio Max and Maya. We were conditioned in the adoption of tools managing the same format produced by 3DArt.
- The configuration management tool
  The selected tool was the Apache Subversion (often abbreviated as SVN) project, an open source version control system, founded in 2000 [1].
- Software architecture

After the first iteration, the software architecture was constituted by the following nine components:

- the RedDwarf server component;
- the Persistence Server, that wraps the MySQL database and complements the pure gaming persistence support provided by RedDwarf server;
- the World Editor component, for the maps assembly and generation;
- a web portal adopted as front-end for the administration and for accessing the application accessible via Java Web Start Technology [6].
- a generic component, named Common, containing all the functionalities that are shared by the other components and constituting the communication core of the system;
- the Client component;
- a simulator component, useful to stress testing the maximum charge supported by the server;
- a distribution server for building and deploying, on the supporting web portal, the Java Web Start application with all the required libraries and resources.

Once prepared a first description of the software architecture, a first draft of the database schema was created to have a preliminary idea of the involved entities and of their relationships.

At the end of the Pre-Game Development phase, the team participated to a design review meeting, on which are based the proposals for the implementation and decisions. Preliminary plans describing the release content were also defined.

4.2 Development

During this phase, the environmental and technical Scrum variables, such as time frame, quality, requirements, resources, implementation technologies and tools, and even development methods, are constantly monitored in order to be able to flexibly adapt to changes.

In the Scrum methodology the development phase is structured in *Sprints*, iterative cycles which last from one week to one month. A Sprint includes the traditional development phases (requirements, analysis, design, evolution and delivery phases). During each sprint, the system design and the architecture evolve.

The project team participated at the beginning of the Sprint to a planning meeting, where the functionality to be developed in the Sprint were established. In a second phase, only the Scrum team members met and discussed on how the product increment should be implemented.

The team member worked in the same laboratory and met every day in a 15 minute meeting.

In several occasions pair programming has been naturally adopted, not because the Scrum Master explicitly required the application of this practice, but because of a decision of the team members for reducing the error rate and for exploiting the skills of the pair.

4.2.1 First Iteration

There were some troubles at the beginning, because the team members were not practice with the Scrum approach and there was a lot of incertitude on the selection of the technology. As planned, during the first iteration, that lasted two weeks, the selected tools were installed and experimented and, by using the world editor, a first prototype of the virtual world was provided, together with some simple objects rendered in it. Figure 1 shows a screendump of the first village prototype after the first iteration.

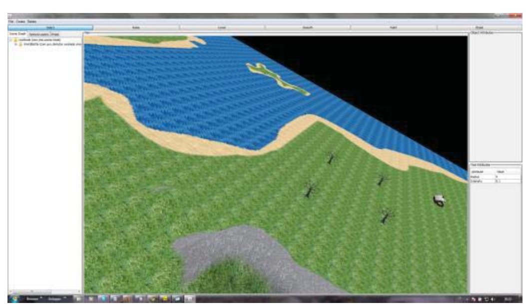


Figure 1. The first village prototype

The main objective of the First Iteration was the validation of the technologies individuated in the previous phase. Indeed, considering the graphical objects, we needed to import in the world game the objects produced by 3DArt using the proprietary Collada format of 3D Studio Max. To this aim, we opened and compressed the objects using MAYA, reducing the number of triangles. When necessary, the objects where further compressed using Blender. A great effort was devoted to understand how importing the objects, due to the poor documentation of opensource projects and the differences existing among formats and their versions. As an example, the mapping among the object faces and their textures, or the light reflection direction could be erroneous. In addition, a Darkstar project has been individuated as reference point, the Snowman project [19]. This project is a demo, thus several aspects are only sketched. During this sprint the development team experimented how to manage multiple game instances, communication between users, the game persistency and so on.

4.2.2 Second Iteration

During the Second Iteration, the Virtual World was further improved, importing several objects and characters produced by 3Dart. The solution adopted caused several problems when importing animated objects. The rendering of the animated objects supports MD5 mesh model, in a specific version with specific settings. If these specific settings are not identified, the world items were visualized as amorphous objects. No character with sense appeared for two weeks in our world. Thus, the iteration was delayed and lasted four weeks. At the end, when we were at the point of abandoning the technology solution, without being confident in other alternative options, after a long serious of trials and errors, a member of the team specialized in graphics solved the problem, identifying a plug-in for Blender enabling to correctly import the objects in the world. After this choice, the development proceeded fast, and all the developers were happy with the selected technologies. Figure 2 shows how the village appears with the imported building and street furniture, rendered with the appropriate resolution.

Figure 2. The Village

4.2.3 Third Iteration

The mechanisms for growing the village and resource exchanging were defined and developed. In Figure 3 an island representing a class is shown. Let us note that the different buildings represent the different levels reached by each student. Clicking on a building the student can access to the selected village.

4.2.4 Fourth Iteration

While the game details were being further implemented, the appearance of the village was improved with other objects created by 3Dart in the last weeks. The male character appearance was perfected and female one was designed.

The school archipelago was also designed. An archipelago is shown in Figure 4, where each island represents a class of the school.

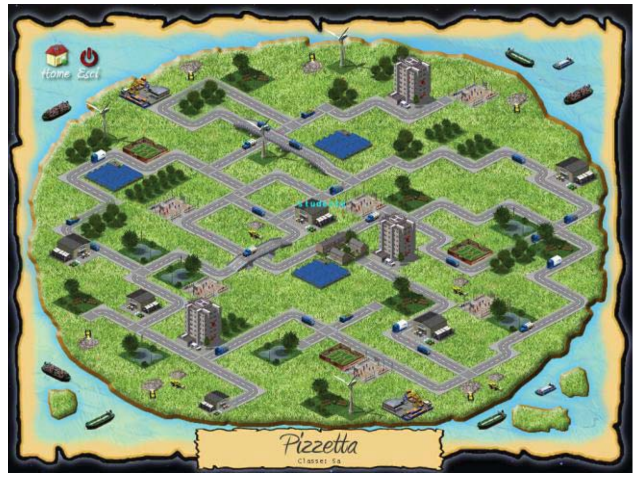


Figure 3. A Class Island

4.2.5 Fifth Iteration

During this phase, the Mini Games were better defined and developed. In particular, well known games such as Packman or Puzzle-Bobble have been adapted to learn how correctly performing the garbage disposal. Figure 5 shows a screenshot of the PackRecycle game, embedding the recycle metaphors and correct behaviors.

Figure 4. An example of school archipelago

4.2.6 Sixth Iteration

The development of the administration features supporting the management of the game, such as creating Archipelagos (schools), Islands (classes), and villages (student accounts) has been delayed because the adoption of the prototyping approach. In addition, these featureswere not critical because they are realized with consolidated traditional web technology.

In addition, five models of islands were defined, differently assigned to the various classes of a school.

4.3 Post-Game development

After the integration of all the components and the testing of the system, this phase is concluded with the release of the game and it is, at the moment, still in progress.

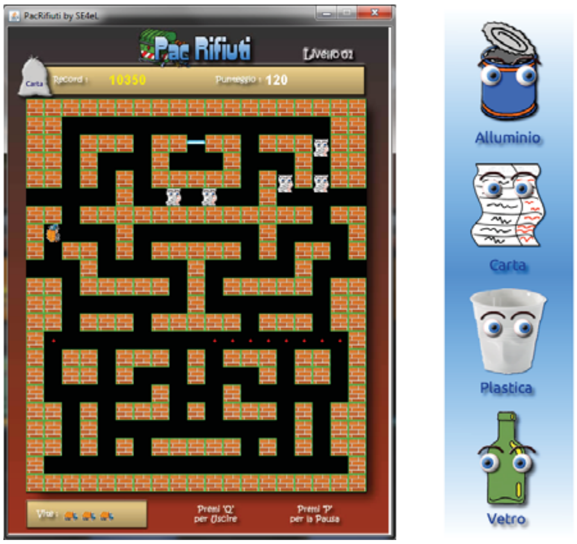


Figure 5. The PackRecycle game and waste disposal icons

5. Conclusion

In this paper we described how a Scrum based development process was applied to the creation of a serious game to support environmental learning. Some problems that can be encountered during the development of a game based on Virtual World technologies have been reported, such as the application complexity and its non functional requirements. The advantages of the adoption of short iterative phases that always monitor uncertainness are evident and enable the adaptation of the requirements and of the selected technologies to the encountered difficulties. Also the integration of open-source components is well supported by this evolutionary approach.

In the future, we plan to evolve the game to present to students the main concepts related to the energy saving. In this occasion we will collect the fun of the final user while the game is evolving.

Acknowledgments

The authors would like to thank Regione Campania that founded the TIE project and the experts of 3DArt for their precious collaboration.

References

[1]. Apache Subversion, http://subversion.apache.org/

[2]. De Lucia, A., Francese, R., Passero, I., Tortora, G. (2009) Development and evaluation of a virtual campus on Second Life: The case of SecondDMI, *Computers & Education* 52(1): 220-233 .

[3]. Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments,* Vol. 1, Issue 2, pp. 262–271, Cambridge, MA: MIT Press.

[4]. R. Kalawsky, The Science of Virtual Reality and Virtual Environments. Addison-Wesley, Reading, Massachusetts, 1993.

[5]. Petrillo, F., Pimenta, M. (2010) Is Agility out there? Agile Practices in Game Development, *SIGDOC 2010*.

[6]. Java Web Start Technology http://www.oracle.com/technetwork/java/javase/tech/index-jsp-136112.html

[7]. Kanode, C.M.; Haddad, H.M. (2009) Software Engineering Challenges in Game Development, Sixth International Conference on Information Technology: New Generations,. ITNG '09., vol., no., pp.260-265.

[8]. C. Keith, "Scrum Rising: Agile Development Could Save Your Studio", in Game Developer. 2007. p. 21-26.

[9]. C. Keith, "Scrum and Long Term Project Planning for Video Games". Gamasutra07 Dec.18, 2007; Available from:http://www.gamasutra.com/view/feature/3142/Scrum_and_long_term_ project.php.

[10]. C. Keith, "Agile Methodology in Game Development", in GDC06. 2006.

[11]. C. Keith, "Beyond Scrum: Lean and Kanban for Game Developers". Gamasutra08 Nov.11, 2008; Available from: http://www.gamasutra.com/view/feature/3847/beyond_Scrum_lean _and_kanban_for_php.

[12]. NeatBeans, http://netbeans.org/

[13]. Papert, S. (1980) Mindstorms: Children, Computers and Powerful Ideas, *Basic Books*, New York.

Apache Subversion, http://subversion.apache.org/

[14]. Polycarpou, I., Krausea, J., Rader, C., Kembel, C., Poupore, C., Chiu, E. (2010) Math-City: an educational game for K-12 mathematics, Procedia - Social and Behavioral Sciences, World Conference on Learning, Teaching and Administration Papers, Vol. 9, pp. 845-850.

[15]. Recycle City, retrieved from http://www.epa.gov/recyclecity/

[16]. RedDwarf Server, http://www.reddwarfserver.org/

[17]. Schwaber, K; & Beedle; M. (2002) Agile Software Development with SCRUM. *Prentice Hall*.

[18]. Sharma, M., Holmes, M., Santamaria, J., Irani, Isbell, ., C., Ram, A. (2007) Transfer Learning in Real-Time Strategy Game Using Hybrid CBR/RL, IJCAI-07

[19]. Snowman, project, http://sourceforge.net/projects/rd-snowman/

[20]. Thompson C. W. (2011) Next-Generation Virtual Worlds: Architecture, Status, and Directions, *IEEE Internet Computing*, Volume 15 Issue 1, pp.60-65.

[21]. Unesco (2008) The contribution of early childhood education to a sustainable society http://unesdoc.unesco.org/images/0015/001593/159355e.pdf

[22]. Waste Hierarchy http://en.wikipedia.org/wiki/Waste_hierarchy

[23]. Wikipedia (2011) Serious Game, retrieved from http://en.wikipedia.org/wiki/Serious_game

[24]. Xu, S. (2006) Empirical Validation of Test-Driven Pair Programming in Game Development, Proceedings of the 5th IEEE/ACIS International Conference on Computer and Information Science

Seminars in Second Life: Teacher and Student Views

Telmo Zarraonandia*, Rita Francese°, Ignazio Passero°, Ignacio Aedo*, Genny Tortora°, Paloma Diaz*
*Computer Science Department, Universidad Carlos III de Madrid, Leganés (Madrid)
°Università degli Studi di Salerno,via Ponte don Melillo 1, Fisciano (SA), Italy

tzarraon@inf.uc3m.es, francese@unisa.it, ipassero@unisa.it, aedo@ aedo@ia.uc3m.es, tortora@unisa.it, pdp@inf.uc3m.es

***Abstract*— Virtual Worlds are largely adopted as places where students can perform normal class activities, including lectures. In this paper we describe a didactical experience conducted in Second Life during an Italian-Spanish collaboration which occurred in spring 2010. We proposed to a group of university students that they make use of a virtual world for carrying out the presentations of the projects of one the courses they were enrolled on. The collected results revealed that Virtual Worlds can effectively support distance lectures, both from the perspective of the audience and from the one of the lecturer, although it is advisable to equip the environment with features for carrying out the lectures in a more natural way.**

I. INTRODUCTION

Over the last few years Virtual Worlds have been largely adopted as a platform for education, see for example [1][4][9], and to support collaboration and meeting in general [3] [12]. In [3] Gandhi affirmed that this kind of environment has been very much appreciated by IBM for conducting virtual meetings and discussions. The main reason is that Virtual Worlds are particularly suitable for communicating because they enhance the users' perception of presence, awareness, communication and their sense of belonging to a community [1]. These peculiarities of Virtual Worlds make them very appealing and, according to [12], in the future we might see a convergence of Web and Virtual World technologies into a 3D Web.

In this paper we describe the practical usage of Second Life (SL) to support several seminars having "Learning Technologies" as their subject. This 3D environment constituted an international place [6] [8] where students and staff from Spain and Italy met. During the seminars students performed both the roles of lecturer and audience. The main purpose of the experience was to investigate if virtual worlds could effectively support this type of lecturing exercise. We were also interested in analyzing the experience of the students who played the role of lecturers, and we tried to identify if the use of virtual worlds reported any benefits in terms of relaxation and reduction of fear to speak in public, as a side effect.

The paper is organized as follows: Section 2 discusses related work, while Section 3 summarizes the main characteristics of Second Life and details the virtual environment used for supporting the experience. In Section 4 we describe the didactical experience, and analyze the results and opinions gathered from its participants. Finally, in Section 5 we present some conclusions and lessons learned

II. RELATED WORKS

Although it seems likely that virtual worlds will have a large impact on the future Web [12], and, in particular, in teaching and learning, there is still a limited comprehension of the effect of their usage for didactical purposes. Virtual Worlds have been largely adopted as a learning environment in the last decade, see for example [1][2][7][9]. However, there are still some unresolved issues such as difficulties in navigation or in using 3D interfaces [13].

In [1] De Lucia et al. evaluated the effectiveness of performing synchronous distance lectures in Second Life. In addition to the usability of the environment, the evaluation focused on investigating four different factors: the users perception of belonging to a learning community, awareness, presence and communication. The feedback obtained on these terms from the participants experience was mainly positive, even from those of them who did not have specific technological skills.

In [7] Monahan et al. examined the design and usability of a virtual world named CLEV-R for supporting various learning tasks. Also in this case, the subjects were in favor of adopting CLEV-R for online education. However, it was also reported that users with no experience in 3D computer games expressed difficulty in navigating in the virtual environment.

Jarmon et al. [5] proposed a study aimed at evaluating the instructional effectiveness of Second Life as an experiential learning environment. Several information sources were considered: journal content analysis, surveys, focus group, and virtual world snapshots and video. Results revealed that a project-based approach favors experiential development of interdisciplinary communication awareness and strategies.

III. SECOND LIFE

The technology we selected to support the proposed learning experience is Second Life (SL) [11], one of the most popular 3D online Virtual Worlds. SL is based on the archipelago metaphor, where space is organized in terms of islands, which are inter-connected via teleportation links, bridges, and roads. The main characteristic of this environment is the persistency. This enables the user to consider the learning environment as a "place" where they can meet, not only for didactical purposes, but also for communicating and socializing [8]. Users access the online system with a client browser and

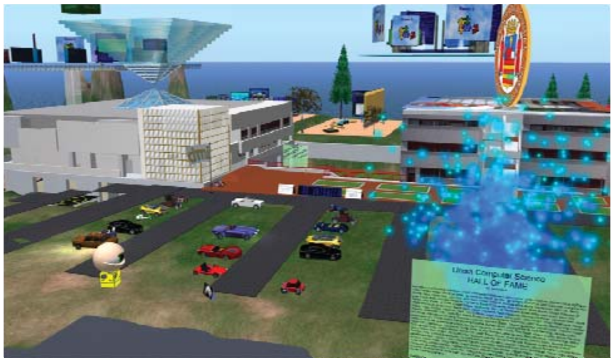


Figure 1. The SecondDMI Island.

interact with content and other "residents", represented by avatars, a user representation in the virtual world.

In this environment it is possible to create 3D objects, using a scripting language provided by LindenLab. The main interaction pattern between avatars and objects is the touch action, which is raised when pointing and left clicking on the object. By adopting the SL scripting language, it is possible to programmatically associate an active behavior to objects, and, in particular, to interactive content. The main difference that exists between a learning experience in Second Life and a two dimensional distance learning system is that SL enables students to attend lectures in a simulated classroom with the other fellows [14]. According to the results provided in [1], this kind of environment enhances the user perception of belonging to a learning community, as well as to the perception of awareness, presence and communication. Non-verbal communication is also supported by inter-avatar communication features, such as animations and gestures. These features are a way of increasing face to face communication. Context awareness is also well supported: the user is clearly able to locate himself, the other users (situational awareness, "who is there") and to understand what each one is doing (action awareness). SL 3D world is a *stigmergic environment*, which means that users can modify it and the modifications will remain visible from one user to another.

In addition, SL allows for using web, video, audio streaming, slide presentations projected as images, and VOIP. People can chat on an open channel both privately as well as publicly.

A. The SecondDMI island

The authors from Salerno designed and developed the SecondDMI island on Second Life, a fantasy version of the Department of Mathematics and Information Technology (DMI) building of the University of Salerno [1], shown in Figure 1.

Some of the DMI classrooms were reproduced with a high level of detail. Avatars could open the doors of the classrooms and sit on the chairs, as they would do in the real campus. Users were also provided with a gesture bar to facilitate the control of the avatar. In particular, we designed the *participant gesture bar*, which animates the avatar with applause, yes, no, and hand raising gestures. However, different experiences carried out suggested that for didactical purposes open environments were more suitable than virtual classrooms. On the one hand, non-expert users could experience frustration having to open and close doors and moving their avatars through a narrow environment. On the other hand, they allow expert users to exploit the *fly movement* of their avatars.

Figure 2. The seminar "place".

For this reason, we simplified the didactic setting, providing a more simple setting which included only a few objects. Among them the slide board, depicted in the rightmost part of Figure 2, supported the showing of the presentation slides. Users could change the current slide both by using the touch action on the board surface or using a button interface exposing the same functionalities of the presentation software.

The setting was also equipped with a booking list board, depicted in the lower rightmost part of Figure 3, whose aim was to help managing the interventions during the lectures. In particular, when a participant touches the Question Block, s/he is scheduled for an intervention and added to the booking list. The list of the interventions can be recorded on the supporting server.

Finally, security issues were also considered when setting up the environment. Accordingly, it was possible to restrict the access to the seminar zone and the control of the seminar boards only to selected participants.

IV. The Didactical Experience

The Computer Supported Education course is one of the elective subjects in which students from the Computer Science degree of the University of Carlos III de Madrid can enroll. During the course, students work in groups on different projects and, at the end of each project, each of the teams has to designate a spokesperson who presents the project results to the rest of the class following a lecture modality scheme. The aim of this exercise is to help students get used confronting an audience and to develop their communication skills, but every year many of them express their reluctance to carry out this labour as they do not feel comfortable speaking in public.

In this scenario, we proposed to the students that they carry out their lectures in a virtual world. On the one hand, the use of virtual worlds in education was one of the course topics, and therefore this experience could be a good hands-on exercise for the students. On the other hand it would allow us to explore two interesting research questions:

(1) Could the lecturing exercise be successfully implemented using virtual worlds in a future distance modality version of the course?

(2) Do students feel less intimidated carrying out a lecture in a virtual environment than in a real classroom?

With the aim of exploring these research questions two research groups from the University of Salerno and from the University Carlos III of Madrid, respectively, started to collaborate in the spring of 2010. Next, we describe the set up of the experience and the results obtained.

A. Experience description

The learning environment chosen to carry out the experience was located on the SecondDMI Island of Second Life. The staff supporting the virtual learning activities was Italian, while the teacher responsible for the seminar organization and the tutors who supervised the experience in the laboratories were Spanish. 20 students enrolled on the course and worked together in 8 different project teams. The presentation of the project was scheduled for the last week of the semester. This aimed to ensure that students would have gained the required experience in using virtual worlds, as by that time they would have got to complete other projects of the course related to the subject. However, a practical session in the SecondDMI Island was also arranged the week before the presentation.

The experiment was performed in two separate laboratories: one for the students and one for the lecturers. In order to simulate the remoteness of the experience the lecturers moved to a separate room to make their presentations. Each student who performed the role of audience had a computer at his disposal and was able to interact with the others and with the teacher only using the SL textual chat. Meanwhile, in the case of the lecturer, he could use the vocal chat for lecturing and answering. This kind of communication was chosen to better control the student interventions: participants booked a question by touching the Question Block, and wrote a question on the text chat when their intervention was selected by the teacher acting on the Booking List. The lecturer, then, answered using the voice chat. In the laboratories, two Spanish tutors provided assistance to the participants in the experience and ensured that the students communicated only with the textual chat. In Figure 3, the virtual classroom, the avatars of the students and the lecturer are shown.

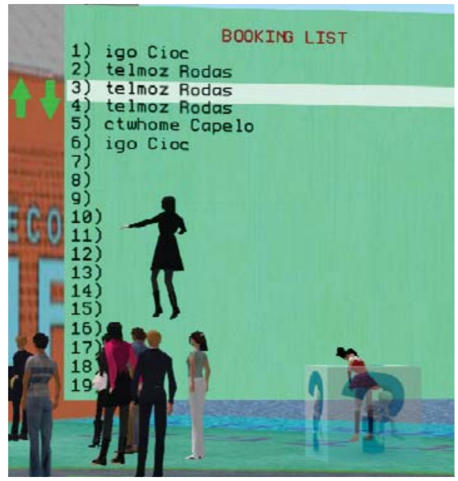


Figure 3. The Booking List and Question Block.

B. Survey description

At the end of the experience, each student was handed with a different survey according to the role s/he had played in the lecture: lecturer or audience. Tables 1 and 2, show the questionnaires used for the evaluation. Both surveys were organized in two different sections. The first one aimed to gather feedback on user satisfaction with the technology, its degree of acceptance and willingness to use it. This way, questions from Q1 to Q3 evaluated the lecture experience itself (and, as a consequence, also the support provided by Virtual World technology) in terms of ease of use and communication effectiveness. Questions from Q4 to Q6 aimed at collecting opinions on the environment qualities and design, while Q7 evaluated if the avatar controlling mechanism interfere with users' actions. In addition, the second section of the questionnaire (questions Q8 to Q15) addressed the user impressions of the experience, and aimed to evaluate the direct effects of the virtual environment on the lecture exposition in terms of relaxedness, user involvement, communication quality and difficulties using the technology. The final question Q15 directly addresses user opinion on the virtual world lecture experience.

Students provided the answers to the survey's questions using a 5 points Likert scale of level of agreement/disagreement. The questionnaires were also complemented with an open question to allow students to explain their ratings and provide comments, and which teachers encouraged the students to use. As Table 1 and 2 show, both questionnaires were similarly organized and addressed the evaluation of the same features and characteristics using questions slightly adapted from each perspective. The right-hand side column of the tables shows the mean scores obtained from the responses for each of the questions. In the following section we analyze these results and try to answer the two research questions tackled in the experience.

C. Research question 1: Discussion

All the mean scores for questions in Group 1, which were the ones that more directly addressed the research question 1, score above 3. Let us note that this is a lower threshold, corroborating the idea that participants were quite satisfied with the virtual environment provided to carry out the lectures, both from the lecturer as well as from the audience perspective. In particular, as shown in the left hand side of Figure 4, that depicts the scores distribution for questions from Q1 to Q6, most of the students who performed the lecturer role have perceived lecturing in the virtual setting as easy. Furthermore,

TABLE I. THE TEACHER VIEW QUESTIONNAIRE

Question Group 1. Describe your experience as a lecturer in a virtual environment		
Q1	It was easy to carry out my lecture.	4.38
Q2	Communication with students was effective.	3.63
Q3	Transferring my lecture to the other participants was easy.	3.87
Q4	The components, needed to have a lecture, were easy to locate	3.87
Q5	The object metaphors were intuitive.	3.87
Q6	Objects reacted in an consistent way to selection and manipulation.	4
Q7	It was easy controlling my avatar.	4
Question Group 2. Would you prefer to present your next exposition in SL o in a traditional classroom?		
Q8	Virtual Worlds provide a more relaxing environment than traditional classroom	3.25
Q9	Virtual Worlds provide a more engaging and attractive environment	3.62
Q10	Perceive the audience through their avatars results less intimidating	3.38
Q11	Perceive the audience through their avatars results distractive	2.62
Q12	Not being able to perceive audience real gestures makes difficult to carry out my presentation	3.75
Q13	The use of additional technology makes more complicated to carry out my presentation.	2.37
Q14	Having to control my avatar makes more complicated to carry out my presentation	3.25
Q15	In general, I feel more comfortable carrying out a presentation in a virtual environment rather than in a traditional one	2.25

TABLE II. THE STUDENT VIEW QUESTIONNAIRE

Question Group 1. Describe your experience as a student in a virtual environment		
Q16	It was easy to follow the lecture	4.1
Q17	It was easy to present my questions to the lecturer	3.25
Q18	It was easy to communicate with other students	3.45
Q19	It was easy controlling my avatar.	3.45
Q20	The system provided a good support for the understanding of the concepts covered.	3.4
Q21	The system encouraged student participation.	3.05
Question Group 2. Would you prefer to follow the next expositions in SL o in a traditional classroom?		
Q22	Virtual Worlds provide a more relaxing environment than traditional classroom	3.6
Q23	Virtual Worlds provide a more engaging and attractive environment	3.75
Q24	Perceive the lecturer through their avatars results less intimidating and encourages me to participate	3.15
Q25	Perceive the lecture participants through their avatars results distractive	3.5
Q26	Not being able to perceive lecturer real gestures makes difficult to follow the exposition	3.9
Q27	The use of additional technology makes more complicated to follow the presentation	2.7
Q28	Having to control my avatar makes more complicated to attend to the presentation	2.55
Q29	In general, I feel more comfortable attending a presentation in a virtual environment rather than in a traditional one	2.55

all the different lectures were carried out smoothly, according to the schedule time, and without registering any problem related to the technology used. This confirms that the SecondDMI Island on Second Life could provide an effective support for carrying on the lecturing exercise in a distance modality of the course.

It should be noted however that the students' comments and the answers they provided to the second group of questions suggest that there is still room for improving the experience. For instance, despite the positive general opinion about the communication support provided (Q2 and Q18), not being able to perceive the real gestures of the participants was reported as a major drawback both by lecturers (Q12) and by the audience (Q26). In the case of the audience this could be mainly due to most of the lecturers tending to leave their avatar still when carrying out their presentation. It is expected that the impact of this problem will be reduced as students get used to the new technology and the tools provided for expressing gestures. However, it is necessary that these tools can be used in a natural way, without interfering in the normal flow and pace of the lecture. As one of the students who performed the role of lecture stated:

"It's difficult to control your presentation, watch the audience avatar, perform gestures and carry out your speech all at the same time."

This is interesting as student's opinions are generally positive when considering individually the means provided for performing the task (Q13 and Q27), and for controlling their avatars (Q14 and Q28).

D. Research question 2: Discussion

With regards to the second research question, results of questions Q8, Q10 and Q13 lead to a moderately positive answer: students perceive the virtual world as a slightly more relaxing environment than a traditional classroom. However, according with the values obtained they are not very enthusiastic in this regard. Two major factors may explain these results. On the one hand, as Figure 5 depicts, the answers to questions Q9 and Q23 show that most of the students perceived the experience as engaging and fun, and therefore less stressful and more relaxing. On the other hand, due to the high level of presence felt by students in this environment, they are conscious of the fact that their identity is known by the other participants. Therefore, they know that their behavior and

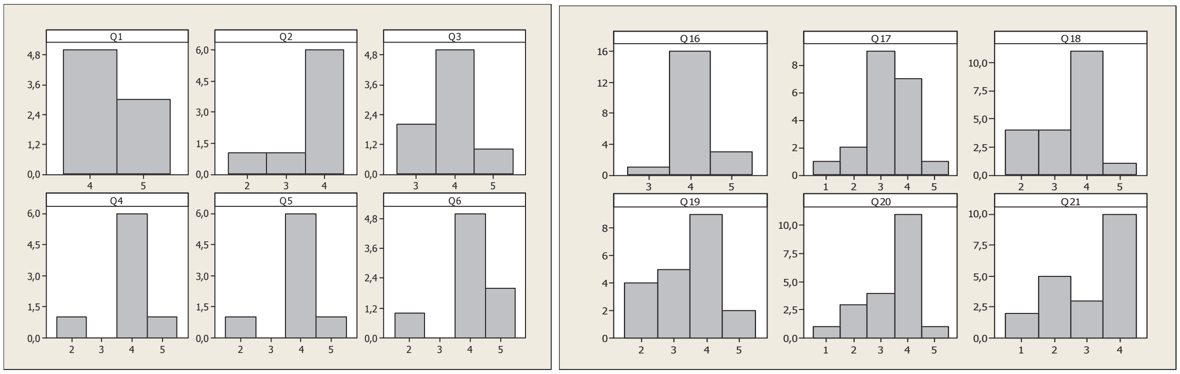


Figure 4. Experience Evaluation Results from the Lecturer (left) and Student (right) Perspectives

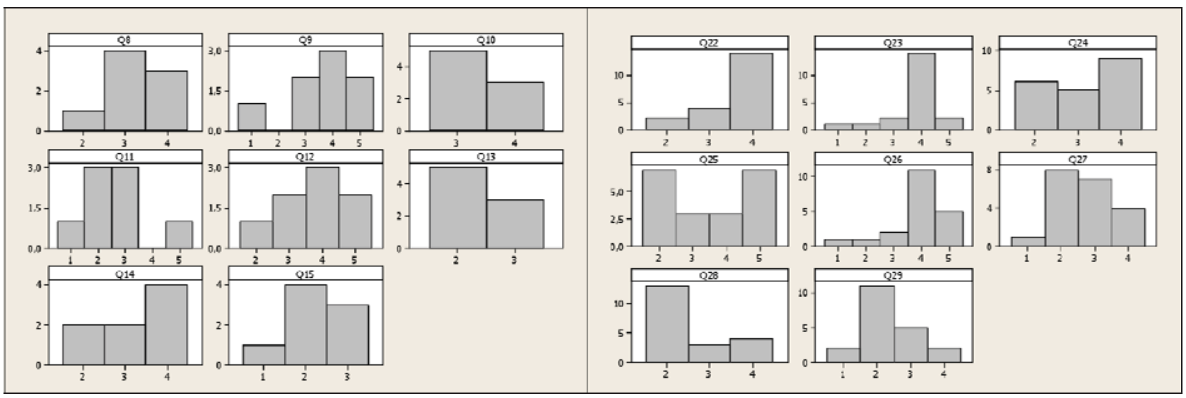


Figure 5. Opinions and Preferences from the Lecturers (left) and Students (right)

their work will be judged by the staff and by their colleagues the same way as in the real classroom and, analogously, some of them feel intimidated (Q10).

In addition, the answers provided in relation to the impossibility of perceiving the real expressions of the participants are also interesting. As mentioned before, neither the results obtained from the lecturer perspective (Q12 mean value 3.75) nor from the audience perspective (Q6 mean value 3.9) are encouraging. However, not all the students perceived this feature as a drawback, and according to one of the students comments it can also help to reduce anxiety and stress:

"A nice feature is that when a student asked me a difficult question, no one in the virtual world was able to perceive my embarrassment by the expression of my face."

It is also necessary to note that perceiving the rest of the participants through their avatar may distract some of the students from the lecture (Q25, Figure 5). However, this problem is not considered significantly important for the case of the lecturer (Q11, Figure 5).

Finally, when directly addressing students to compare the experience of performing a lecture in a virtual and a traditional classroom, there doesn't seem to be significant differences in the responses based on the role performed during the lecture (Figure 8). As stated before, students seem to feel more relaxed in a virtual world than in a traditional classroom, probably due the resemblance of the virtual world to a videogame, but the mean values obtained for each role are close (3.25 and 3.6) and the distributions are similar (Figure 8). In the same way, with regard to their level of comfort they favor the real classroom with respect to the virtual environment. This is reasonable, as this is the way they are used to performing this task. However, once again, the role played during the exercise does not seem to affect their level of preference as there are not great differences in the mean values obtained for each role (2.25 and 2.55), or in the distribution of the answers (Figure 6).

V. Conclusion

The results of the experience seem to corroborate that virtual worlds can effectively be used for supporting distance lectures. The lecture exercises could be successfully carried out, and students showed a positive attitude towards the adoption of this technology. However, in order to fully exploit all the possibilities it offers, it is necessary to provide the users with the means to carry out the lectures in a natural way. With regards to the second research question tackled by the experience, we can conclude that the benefits of carrying the lecture in a virtual environment in terms of reducing the stress of the lecturer are limited. Regardless of how graphically imaginative the representation of the audience is, the participants of the lecture are still aware that they are speaking to the same people as in the real classroom, and therefore and analogously, some of them feel intimidated.

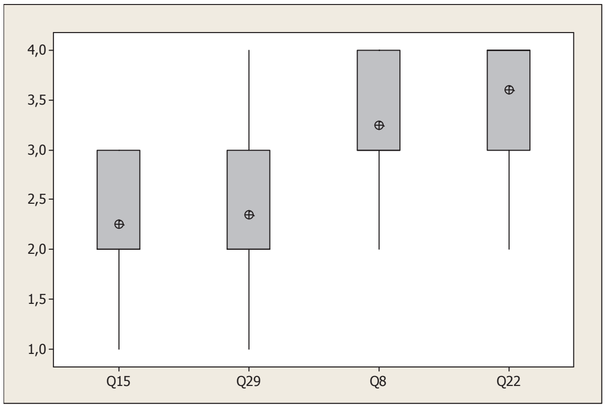


Figure 6. A direct comparison between scores assigned to teaching and attending a lecture: feeling of comfort (Q15-Q29) and relax (Q8-Q22).

ACKNOWLEGMENT

This work has been supported by a mobility grant of the research program of the University Carlos III de Madrid, and by the Project TIPEx funded by the Spanish Ministry of Science and Innovation (TIN2010-1859-C03-01).

REFERENCES

[1] De Lucia, A., Francese, R., Passero, I., Tortora, G. (2009) Development and evaluation of a virtual campus on Second Life: The case of SecondDMI, Computers & Education 52(1): 220-233 .

[2] De Lucia, A., Francese, R., Passero, I., Tortora, G. (2009) Development and evaluation of a system enhancing Second Life to support synchronous role-based collaborative learning, Softw., Pract. Exper. 39(12): 1025-1054

[3] Gandhi, S. (2010) IBM dives into Second Life: Meeting, collaborating, and brainstorming in a virtual world, http://www.ibm.com/developerworks/opensource/library/os-social-secondlife/index.html

[4] Harvard University. <http://slurl.com/secondlife/Berkman/105/74/35> .

[5] Jarmon, L., Traphagan, T., Mayrath, M., Trivedi, A. (2009) Virtual world teaching, experiential learning, and assessment: an interdisciplinary communication course in Second Life, Computers & Education, Vol. 53, pp. 169-182.

[6] Jones, Q. and Grandhi, S.A. (2005) P3 Systems: Putting the Place Back into Social Networks, IEEE Internet Computing, 2005, pp. 38-46

[7] Monahan, T., McArdle, G., & Bertolotto, M. (2008). Virtual reality for collaborative e-learning. Computers and Education, 50(4), 1339–1353.

[8] Prasolova-Forland, E.; Sourin, A.; Sourina, O. (2005) Place metaphors in educational cyberworlds: a virtual campus case study, IEEE International Conference on Cyberworlds.

[9] Proceedings of the Second Life Education Workshop. (2007). Part of the second life community convention. Chicago Hilton, 24–26th August 2007.

[10] Rymaszewski, M., Au, W. J., Wallace, M., Winters, C., Ondrejka, C., & Batstone-Cunningham, B., (2007). Residents from around the world. Second life: The office guide (p. 4). Wiley Press.

[11] Second Life. http://secondlife.com..

[12] Thompson C. W. (2011) Next-Generation Virtual Worlds: Architecture, Status, and Directions, IEEE Internet Computing, Volume 15 Issue 1, pp.60-65.

[13] Youngblut, C. (1998). Educational uses of virtual reality technology. Technical Report No. IDA Document D-2128, Institute for Defence Analyses.

[14] Zhu, Qing; Wang, Tao; Jia, Yufu, (2007). Second Life: A New Platform for Education. In the Proceedings of First IEEE International Symposium on Information Technologies and Applications in Education, pp. 201 – 20.

The SAMAL Model for Affective Learning:

A multidimensional model incorporating the body, mind and emotion in learning

Horace Ho-Shing Ip
AIMtech Centre and Department of Computer Science
City University of Hong Kong
Hong Kong
cship@cityu.edu.hk

Julia Byrne
AIMtech Centre and Department of Computer Science
City University of Hong Kong
Hong Kong
jbyrne@cityu.edu.hk

Shuk-Han Cheng
Department of Biology and Chemistry
City University of Hong Kong
Hong Kong
bhcheng@cityu.edu.hk

Ron Chi-Wai Kwok
Department of Information Systems
City University of Hong Kong
Hong Kong
isron@cityu.edu.hk

***Abstract*—In this study we propose a new model for affective learning, SAMAL Model (Smart Ambience for Affective Learning Model) which illustrates the interplay between the body, mind and emotion in learning. This model is based upon experiential theories and the use of the body moving within the virtual reality space. The SAMAL project investigates the application of immersive interactive media and virtual reality as a tool in education to enhance learners' motivation to learn, and make a bridge between affect, cognition and learning. The players actively engage with the interactive media experiencing the feelings associated with the challenges and successes of interacting with the virtual learning scenarios. SAMAL learning activities give opportunity for students to feel the many facets of animal survival. Our initial findings revealed that action based trial by error experiential process offered through the SAMAL allowed the students to experience the challenges of survival for birds from an affect sense thus increasing students' learning motivation. Our data also showed that there was also a higher correlation between affect and greater learning for players than was for the watchers, who actively observed.**

Keywords- affective learning; smart ambience learning; immersive learning, experiential learning scenario, action based Trial and and Error Experiential Learning , kinesthetic

I. Introduction

In this paper, we will present an Affective Learning Model based upon a smart ambience learning environment which we called the Smart Ambience for Affective Learning Model or SAMAL model. The SAMAL environment [13] allows students to step into specially designed virtual reality scenarios, and through a series of interactive and immersive learning activities, to motivate and facilitate the learning of specific topics, and thus provides the learner with a unique learning experience, not found in traditional classroom environments.

Centered around experiential learning theories and the use of the body moving within the virtual reality space, the Smart Ambience for Affective Learning (SAMAL) project investigates the application of immersive interactive media and virtual reality (VR) as a tool in education to enhance learners' motivation to learn. SAMAL experiential learning activities are designed to provide the learner in tertiary and secondary education with multimodal learning experience and aim to bridge between affect, cognition and learning.

One key element that makes SAMAL a unique learning platform is the emphasis of the use of the whole body movement in interacting with the virtual scenario. Learner's interaction with immersive and affectively evocative virtual scenarios stimulates a release of energy expelled through the body and is referred to as kinesthetic body release. This release is experienced in the body, taps into the senses and perception such as the sensory-perceptual level of self. While it can tap into the feeling states, it can equally tap into the cognitive states of the learner [1]. SAMAL introduces the learner's body motion as a new component in learning. Through the kinesthetic physical interaction between the body and the virtual learning scenario, SAMAL provides an effective platform for us to investigate and develop a conceptual model that links the relationship between affect, cognition and learning.

A. The concept of movement and sensation associated with learning

The concept of movement and sensation, associated with learning is one unique attribute that immersive virtual reality offers to education. Active participation, along with being immersed through the senses and psychological states are features of VR that can promote learning [2]. VR allows the user to 'get into' the scene and interact with the virtual characters/ environment actively. This kind of inside participation, along with full immersion facilitates an integration of the user in the virtual environment [3].

Winn and Bricken [4] postulated that the users experience of internalizing the virtual world is an authentic, but different hue of reality; this phenomenon enables them to interact (actively participate) with the virtual world in a unique way, and can be more convincing than passively watching a play or (3D) movie. "The immersion allows the user to experience, interact with and discover digital knowledge first hand while the manipulation of the verity of the virtual world allows the learner to establish visual and/or kinesthetic relationships to help understanding of the real world or concepts related to it ([3], pg. 40)".

In this project, we focus on the learning of the concepts of animal survival. Students may cognitively understand survival needs through traditional learning modes such as teacher prepared written material or watching a wildlife survival documentary. However an immersive VR educational programme offers to the student an opportunity to fully immerse in survival tactics through action gives them a deeper understanding in terms of the associated affects and cognition of the desire and concept of survival. Through such real time interactive activities, the students can directly 'feel' the experience in their body, and feel the emotions, such as anxiety, of survival. The student can connect to the animal concerned, e.g. a bird, in the learning scenario, as he/she perceives oneself as that animal (full immersion) in the VR scenario. Results from recent studies carried out about immersion and learning show that there was a higher level of enjoyment and motivation to learn for those learners exposed to immersive conditions as opposed to non-immersive ones [3].

B. Affective Learning

Much has been written about how emotions affect cognition and memory [5], as well as how emotions attached to the success/failure of academic performance impact learning [6]; but little has been researched about the association between immersion in VR and emotions and learning. Emotions within the educational setting most often relate to achievement (success/failure) and the associated feelings such as pride/happiness and sadness/shame based upon the perceived controllability of the activity [6].

Shen, et al. [7] conducted a study that explored how emotion develops during the process of learning and how emotions can be tracked; giving feedback. They postulate that this emotion feedback can improve the learning experience for students. An affective e-Learning model was proposed, based upon concepts from Russell's circumplex model of affect and Kort's learning spiral model which merged learners' emotions with the e-Shanghai e-Learning platform [7].

The focus today on e-Learning is learner-centered, with an emphasis placed upon pervasive and all-encompassing, along with personalized learning technologies. E-Learning should 'better engage learners in the learning process; engaged learners are behaviorally, intellectually, and emotionally involved in their learning tasks [8]. Research points towards the link between positive moods and effective problem solving. Positive mood generates a more creative approach to problem solving with more thorough decisions made.

Kort, et al. [9] proposed a four quadrant learning spiral model in which emotions change when the learner moves through the quadrants and up the spiral. They also proposed five sets of emotions that may be relevant to learning. However, empirical evidence is needed to validate the learning spiral model and to confirm the effects that these emotions might have on learning.

II. The SAMAL Model for Affective Learning: A Multidimensional Model

In recent years, with the advent of e-learning technologies, learning models are being developed to help us understand the role of technologies in the learning process and how technology serves to enhance learning effectiveness. Such attempts aim to provide a structure to the learning experience and a context in which learning material can be presented, enabling the learner to move through the learning quadrants. In this paper, we focus on developing a model that provides a rationale that underpins the SAMAL learning environment and scenarios design. Based upon our trials with students trying out the SAMAL learning environment, we propose a model that begins to explain how the immersive element of the SAMAL design links all three parts of the learning domain: the body, the mind and emotions. This model is multi-dimensional and has been inspired by two earlier models: Kolb's model for experiential learning and Sundstrom's model [12] of the affective loop.

SAMAL offers a new platform whereby all three parts of the learning domain (the body, the mind and emotions) are stimulated in virtual space; this model integrates two key concepts, otherwise known as components:

- That of experiential learning, an active learning process based upon trial and error through the interaction with the virtual environment and engaging the body in the learning scenario.
- The concept of control and focus in relationship to learning.

Features of the SAMAL design are linked to how the students can actively participate and control their body movements within the virtual environment and experience the challenges of achieving certain goals, in this case animal survival. Success can lead to positive feelings resulting in greater learning about the topic from this unique learning media. In the SAMAL design, the focus is placed upon creating a learning environment where the student can learn through a process of action based trial and error, whereby kinesthetic physical interaction with the virtual characters and environment are simulated.

This kind of experiential learning is linked to Kolb's theory of the experiential learning cycle shaped by earlier theories of Dewey and Lewin and Piaget, who all stress that learning involves a process in which new concepts are formulated from, and ceaselessly re-modified by, the experience [9;10]). The SAMAL Model for Affective Learning incorporates aspects mentioned in the above.

A. The Experiential Learning Cycle Component

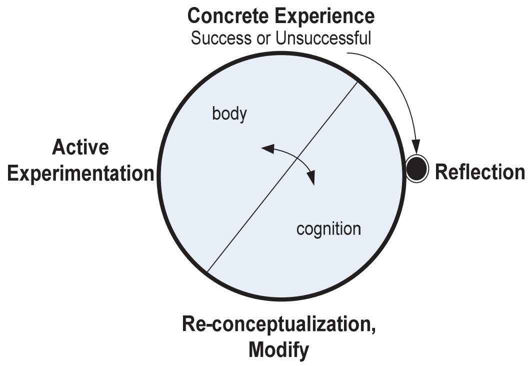


Figure 1. Experiential Learning Cycle Dimension of the SAMAL Model

With reference to Figure 1, the first component of the SAMAL Model for Affective Learning denotes a learning process based upon the concept of using a trial and error experiential approach set within the immersive virtual learning environment. This component of the SAMAL Model is shaped by Kolb's Experiential Learning Cycle and Basic Learning Styles Model. The theoretical underpinnings of Kolb's model postulates that learning occurs in four stages: concrete experience, reflection, abstract conceptualization, and then active experimentation, which then leads back to a new, more formulated concrete learning [10].

The experiential learning process signified in this dimension of the SAMAL model also takes place in stages; but extends the learning process past the newly devised concrete learning into a new deeper process of reflection, acquired by the student. This process is shaped by either the success or unsuccessful outcome of the concrete experience of the SAMAL activity.

In practice, SAMAL is conducted in a workshop format. For the animal survival topic, students experience two virtual education activities related to the topic. The objective of one of the activities, the Hummingbird Flying activity, is to experience how a bird needs to control flight navigation to reach his food target in order to survive which involves strategy design and implementation in order to achieve the goal of survival.

1) Concrete Experience: After introducing the learning objectives, the instructor will demonstrate the Hummingbird Flying activity, and explain how to operate the interactive devices, in this case a pressure sensor and a motion sensor. The instructor will encourage the students to try out the scenarios themselves. Players actively participate, handling the devices and directly controlling the virtual characters. Players will navigate and reach the flower for gaining sustenance in a 3D scene through shifting his/her body weight while flapping the arms at the same time (Figure 2).

Figure 2. A snapshot of a student (Player) playing the hummingbird flying scenerio

In the Hummingbird Flying Scenario, students will perceive themselves as the hummingbird, trying to overcome the difficulties in controlling flight and avoiding obstacles. The direction that the student shifting his/her weigh on pressure sensor in combination with kinesthetically flapping his/her arms is an important part of successfully reaching the target; in this case the flower, in order to suck the nectar and be nourished. It can be difficult at first for some students to combine the appropriate movements and also the right energy to exert in "flying", resulting in uncontrolled flight, unsuccessful outcome, and varying levels of frustration.

2) Reflection leading to re-conceptualization: At this time some form of intervention from the instructor is carried out. The student is encouraged to stop and reflect upon what may have gone wrong in the virtual process of flight impeding them to reach their target. After this reflection process, the instructor will work with the student (player) to facilitate him/her to re-formulate the strategy for reaching the target in the 3D space as if he/ she was the bird in flight. At this stage modifications to the flight/navigation are made.

Figure 3. Navigating towards the target

3) Active Experimentation Leading to a New Concrete Experience: This active experimentation leads to a new concrete experience; successful or unsuccessful. The student will try it again, putting into action new concepts and modifications of strategy; the right movements and exertion of appropriate power (energy) can enable the student to become

immersed in the virtual scene and achieve the goal successfully. If the student still encounters trouble then the trial and error process repeats again with a period of re-reflection and re-conceptualization/modification.

Through this experiential learning process the students can learn to control the flying direction and speed by carefully controlling the pressure of the legs and flapping the arm to control height and speed. The kinesthetic rapid shaking of the students arms echo the flapping of the wings of the hummingbird. The use of the body heightens the immersive experience of this Hummingbird Flying Activity. The process goal is met if the player succeeds in controlling the hummingbird to reach all the flowers (Figure 3).

B. Reflection of New Concrete Experience: Specifications

Successes can lead to positive feelings resulting in greater learning. As a way to solidify the new concrete experience the SAMAL Model cycle incorporates a post experience reflection process, enabling the student to reflect upon what learning elements lead to success and the learning achieved. It is proposed that this process can take place on two levels: 1. a brief verbal interview where the student reflects upon 'what went right' as a result of making modifications and 2. A questionnaire devised to assess the link between affect, motivation to engage in the educational activity, and perceived learning outcomes in this experiential learning activity.

C. The Variability Component of the SAMAL Model

In addition to the experiential learning cycle component which creates a framework for the SAMAL Instructional Plan, the SAMAL Model for Affective Learning incorporates another component, variability, which helps to explain the interface and interplay between focus, control and the resulting feelings in this kind of experiential learning process.

The development of this component looked to Sundstrom [12] who proposed an interesting concept known as an 'affective loop', which refers to an affective interaction process or cycle where emotion plays an important role in interaction involvement and evolution'. His 2-dimensional planar model looks at how to predict and interpret emotions on two axes: from much-control to little-control on the y-axis; and from high-focus to low-focus on the x-axis.

Taking aspects of Sundstrom's plane model, the second component of the SAMAL Model looks at the variability of how much control the student has in the learning activity, i.e. controlling his/her body and movement to interact with the virtual scene; and the level of focus the student has whilst engaging in the activity. Both axes follow a continuum ranging from high to low focus on the x-axis and little to much control on the y-axis. If the student is highly focused and has much control navigating in the virtual environment then especially positive feelings will result such as excited, happy, and amazed. On the other hand, if the student has some control and low focus, then boredom will often result. Continuing on with the activity with low focus will affect control in a negative way and, with little control, negative feelings such as irritated and frustrated will emerge. But at the same time these negative feelings will also emerge if the student is highly focused but has little control over the activity and if something does not change then the student will give up and cease the activity.

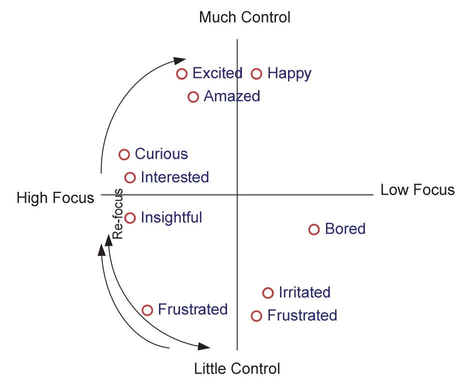


Figure 4. Variability Component of the SAMAL Model

This is where the concept of *re-focus* becomes paramount in this model. If the student can find a way to re-focus and make a change in how he/she approaches the activity/ or have a new concept (either gained from the student or though intervention by the facilitator/teacher), then he/she becomes more insightful, and can become curious and interested in engaging again. Applying this change with renewed high focus can result in greater control and with continued high control of the activity then excited and happy feelings related to achieving success will materialize. The movement towards having much more control over the learning activity stems from re-focusing and correlates to the reflection and re-conceptualization/ modification aspect of the experiential learning dimension of this model seen above. Contrary to this would be that the student would re-focus and still not gain greater control, resulting again in diminished control and negative feelings.

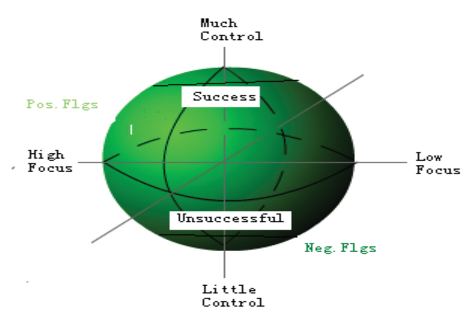


Figure 5. The SAMAL Model

When we combine the two components together we get a 3-dimensional model of the SAMAL model for affective learning, shown in Figure 5. Merging the experiential learning cycle component of the SAMAL Model with variability component serves to illustrate the interconnectivity of the body, mind and affects encompassed in the learning process of trial and error experiential learning.

III. THE EFFECTIVENESS OF SAMAL

Details of the SAMAL system design and the associated instructional design have been described in [13]. After participating in the SAMAL workshop, students are asked to complete a questionnaire devised to determine the effectiveness of the SAMAL activities and affective outcomes. The questionnaire asks the students to select all the feelings that they felt while engaging in the two experiential learning activities: Hummingbird Flying and Animal Jumping. There were 31 players in all and most players indicated between 2 to 3 feelings in all. The highest responses are shown below in Table I. For the players 61.29% responded that they were curious and also happy, as well as excited and interested (both at 54.84%). On the other hand 29.03% stated that they also felt frustrated while engaging and 6.45% felt irritated. This may point to the fact that for some students, they were not able to initially control flight navigation in the bird flying activity, and not achieving success became frustrated/ or irritated as a result.

Table II denotes players having positive feelings (happy, curious, excited, interested, awed/amazed, and insightful) and negative feelings (frustrated, irritated, and bored) while experiencing SAMAL. 21 out of 31 players (67.74%) responded that they had positive feelings during the activity, The 1 player who expressed negative feelings and did not want to further engage was due to having an unsuccessful experience. In summary, there was a high correlation between positive feelings and learning for players; as 19 out of the 21 players (90.48%) indicated that they had 'learned a lot'. 9 out of the 31 players, or 29.03%, experienced both positive and negative feelings but at the same time 7 out of the 9, or 77.78%, specified that they had also learned a lot. Most of the players went through a process of trial and error, where they needed to modify their movements and re-focus, and as a result gained more control leading to success.

TABLE I. FEELING STATES FOR PLAYERS

Players 31 in total		
Curious 19 (61.29%)	Happy 19 (61.29%)	Excited 17 (54.84%)
Interested 17 (54.84%)	Awed/Amazed 11 (35.48%)	Frustrated 9 (29.03%)
Insightful 6 (19.35%)	Irritated 2 (6.45%)	Bored 0 (0.00%)

TABLE II. CROSS CORRELATION DATA: FEELINGS ASSOCIATED WITH STIMULATION/MOTIVATION AND LEARNING FOR PLAYERS IMMERSED IN SAMAL

31 people in total	**Players**							
Feeling	**Responses**		**Engaged in flying**				**Learnt a lot**	
		% out of 31	**Want to engage**		**Don't want to engage**		**Learned a lot**	
Positive Feeling	21	67.74%	17	80.95%	4	19.05%	19	90.48%
Negative Feeling	1	3.23%	0	0.00%	1	100.00%	0	0.00%
Both Pos/Neg Feeling	9	29.03%	4	44.44%	5	55.56%	7	77.78%

IV. CONCLUSION

In this paper, a model for affective learning which we called the SAMAL Model (Smart Ambience for Affective Learning Model) has been proposed. The SAMAL model serves to help understand the interplay between the body, mind and emotion in a novel learning environment where a learner interacts with specially designed virtual scenarios in a VR space The SAMAL model is based upon experiential theories and is a multi-dimensional model that integrates two components of a learning process: an experiential learning cycle component based upon the concept of using a trial and error experiential approach within the immersive virtual learning environment and a variability component that models the concept of control and focus in relation to learning.

The SAMAL learning environment developed at City University of Hong Kong provides a platform for investigating the effect of immersive interaction through body movement of a learner with the virtual characters and scenarios as a tool in education. Through a set of specially designed interactive scenarios to help students to understand animal survival, the SAMAL environment offers students an opportunity to fully immerse in survival tactics giving them a deeper understanding of the concept of survival and to experience from an affect sense what it is like for animals to survive, thus increasing students' learning motivation, according to the data collected. The proposed affective learning model (the SAMAL model) helps us to begin to understand the relationships between affect, cognition and learning through an inter-linking process of control, focus (re-focus) and reflection.

ACKNOWLEDGMENT

This work is funded by a Teaching Development Grant of the City University of Hong Kong, No. 6000158

REFERENCES

[1] Kagin and Lusebrink (1990). The Expressive Therapies Continuum (ETC), Chicago: Bruner and Mazel Publications.

[2] Loeffler, C.E. & Anderson, T. (Eds.)(1994). The Virtual Reality Casebook, New York: Van Nostrand Reinhold

[3] Javidi, G. (1999). Virtual Reality in Education. Thesis (University of South Florida)

[4] Winn, W. & Bricken, W. (1995). The Virtual Reality Roving Vehicle Project, T.H.E Journal, 23 (5), 70-75.

[5] Bower, G. and Forgas, J. (2000). Affect, Memory, and Social Cognition. In M. Marschark (Ed.), Cognition and Emotion (pp. 87-168). Oxford: Oxford University Press.

[6] Pekrun, R., Frenzel, A., Goetz, T., Perry, R. (2007). The Control-value Theory of Achievement emotions: and Integrative Approach to emoitions in Education . In P. Schutz and R. Pekrun (Ed.), Emotion in Education (pp. 13-36). Amsterdam: Elsevier Academic Press.

[7] Shen, Ll, Wang, Ml, & Shen, R. (2009). Affective e-Learning: Using "Emotional" Data to Improve Learning in Pervasive Learning Environment, Educational Technology & Society, 12 (2), 176-189.

[8] Bangert-Drowns, R. & Pyke, C. (2001). A Taxonomy of Student Engagement with Educational Software: An Exploration of literate Thinking with Electronic Text, Journal of Educational Computing Research, 24 (3), 213-234.

[9] Kort, B., Reilly, R., & Picard, R. (2001). An Affective Model of Interplay between emotions and Learning: Reengineering Educational Pedagogy-building a Learning Companion, Proceedings of the IEEE International Conference on Advanced Learning Technologies, Los Alamitos: CA: IEEE Computer Society Press, 43-46.

[10] Kolb, D. (1984). Experiential Learning: Experience as the source of learning and development. Englewood Cliffs, NJ: Prentice-Hall

[11] Kolb, A. & Kolb, A. (2005). Learning Styles and Learning Spaces: Enhancing Experiential Learning in Higher Education, Academy of Management Learning & Education, 4 (2), 193-212.

[12] Sundstrom, P. (2005). Exploring the Affective Loop. Stockholm: Stockholm University.

[13] Horace Ho-Shing Ip, Julia Byrne, Shuk Han Cheng, Ron Chi-Wai Kwok and Maria Sau-Wai Lam, "Smart Ambience for Affective Learning (SAMAL): Instructional Design and Evaluation", Proc. of the 18th International Conference on Computers in Education (ICCE) - Workshop on The Design, Implementation and Evaluation of Game and Toy Enhanced Learning, Malaysia, 29 November – 3 December, 2010.

Using the SCORM Standard to Build Adaptive Content Packages in RELOAD

Pierpaolo Di Bitonto[1], Teresa Roselli[1], Veronica Rossano[1], Cristina De Serio
[1]Department of Computer Science,
University of Bari "Aldo Moro"
Bari, Italy
{dibitonto, roselli, rossano}@di.uniba.it

Abstract—**It is universally recognized that learning is more effective if the didactic content is tailored to the specific user's needs. Since the advent of Adaptive Hypermedia, research in e-learning has always investigated, from a theoretical point of view, the different dimensions involved, and from a technological point of view, the tools helping to integrate more adaptive contents. The problem is that these are often ad hoc solutions which are not interoperable at all. This paper presents a solution that aims at using the SCORM standard in order to build an adaptive content package. The strength of this work is that it applies the adaptation model using some API functions which affect only the client side of the package, without affecting the standard and the interoperability with the LMS. The paper also presents some additional functionalities for the RELOAD authoring environment intended to support non-expert users in building adaptive content packages.**

Standard SCORM; Cognitive Style; Adaptive Learning, RELOAD.

I. INTRODUCTION

The spread of e-learning has generated a profound modification of learning paradigms, opening out new pathways requiring new procedures and tools. One of the most interesting challenges in e-learning research is to define new models serving to adapt learning contents to the specific learner's requirements. It is universally recognised, in fact, that personalisation of the content improves the effectiveness of the learning process, because the resources proposed to the student will be in the form best suited to her /his requirements [2]. In particular, the most commonly used approach in adaptive learning research uses the student's learning styles as a basis on which to adapt the content [3, 4, 9, 10]. Moreover, the use of technologies in educational contexts allows the navigational path to be adapted during the user interaction, thus improving the student's learning experience and tailoring it more closely to each student's characteristics. Based on these claims the present research, conducted in collaboration with a research unit in cognitive psychology led by Maria Sinatra, professor at the University of Bari, has allowed us to define an adaptation model for multimedial learning content based on different students' cognitive styles, already presented in [8]. One of the main problems in defining an adaptation model is its interoperability. The majority of the proposed solutions, in fact, are ad hoc solutions that cannot easily be used in a different context. The main purposes of this research were, on one hand to define an adaptation model which allows the content to be tailored to the user's cognitive styles, and, on the other, to investigate whether it is possible to build a SCORM compliant adaptive content package. The paper closely examines the technical aspects faced when using the SCORM standard to build adaptive learning contents. The main aim of the research is to apply the previously defined theoretical adaptation model [8] using a standard technology. It is clear that this should not involve an extension of the specifics, that would hamper the interoperability of the standard, but would focus on an attempt to reuse them to achieve the adaptation process, according to the defined model. The research was then developed with the design of some solutions to be integrated in one of the authoring environments most commonly used to create adaptive content packages, RELOAD [12], in order to support non expert users. Thus, the research questions posed were: is it possible to build a SCORM compliant adaptive Content Package? Is it possible to facilitate the building process of adaptive content packages? Are there any authoring tools available that can do this? The paper addresses these questions in the following sections: in section II the adaptation model according to the student's cognitive styles is described; in section III the adaptive content package is presented and some limitations of SCORM are discussed; section IV describes the technical details of the SCORM standard; section V proposes some functionalities that could be added to one of the most popular authoring tools. Finally, some conclusions and future research directions are outlined.

II. THE ADAPTATION MODEL

It is widely recognized that the learning performance is improved if the learning content is adapted and tailored to the learner's cognitive style [3, 5, 13, 14, 15]. For this reason, in this work an adaptation model for multimedia didactic resources has been defined, in collaboration with Professor Maria Sinatra's research group on cognitive psychology. The starting point of the adaptation process, detailed in [8], is defining the different cognitive styles according to a process that aims at discovering which presentation modes the specific student prefers. The questionnaire used is the Italian Cognitive Styles Questionnaire (ICSQ) defined by De Beni, Moè, and Cornoldi [7]. The ICSQ consists of a series of questions that are posed to assess whether the subject prefers the "wholist" or the "analytic" style, or else the "verbalizer" or "imager" styles. The basic idea underlying the defined adaptation model is that each subject may prefer a particular presentation mode but that this preference can change during the interaction with the content. Thus, the result of the questionnaire is an ordered list of the student's cognitive style preferences that is employed during use of the didactic content to establish which presentation style should be used with the student at that time. During the interaction, in fact, some test points are set and if the student passes the test the same cognitive style will be used for the next content, whereas if the student fails the test reinforcement content will be supplied. This step aims at reducing the probability that failure of the test is due to a difficult content or to lack of attention by the student. After the reinforcement has been given, the student has to pass a second test: if she/he is successful this time the same cognitive style will be used, otherwise the next cognitive style will be read (from the student's profile) and used for the following didactic units. The model has been implemented using the learning content of the course on the Psychology of Communication and was experimented in two different degree courses: Informatics and Digital Communication, and the Humanities. The pilot study has demonstrated that the customization of learning paths according to students' different cognitive styles promotes academic success [8]. In addition, the model has been implemented in a mobile application, named MoMAt (Mobile Museum Adaptive tour), which is aimed at guiding a visitor on a virtual tour of a museum [6].

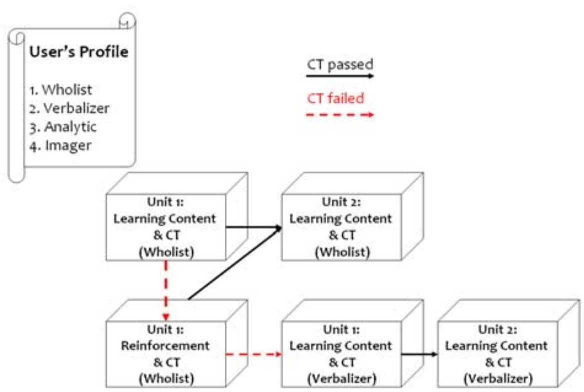


Figure 1. The adaptation model

III. The Adaptive Content Package

The paper is focused on the SCORM standard, that was chosen with the intent to build an interoperable LO that could be easily integrated into any e-learning environment. The key concept of the model is to offer a high level of personalization, but in order to obtain this a high level of granularity is needed. Therefore, the didactic objectives need to be divided into different topics (Figure 2). For each topic two kinds of SCOs are built: one containing the learning content and one containing the reinforcement. Both of them contain use presentation modes suited to each cognitive style, and for both of them there will be a SCO which will contain the Comprehension Test (CT) to assess the student's learning level. The CTs allow the content to be adapted according to the adaptation model described above. In order to make the model scalable and general, another type of SCO has been added to the theoretical model defined in the previous section. In fact, a cognitive style test SCO has been added, which contains the questionnaire serving to define the user's cognitive style. In this way, the adaptive content package is sufficiently general to be applied independently of the cognitive style questionnaire used. The SCOs are then organised using a tree aggregation form that represents the domain organisation and the adaptation model described.

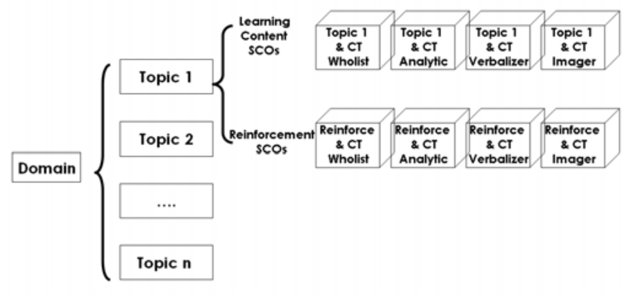


Figure 2. The tree organisation of the didactic content

A. *Limitations of the SCORM standard*

The next part of the research was focused on the SCORM standard, with the aim of defining an easy procedure for building an adaptive content package. At first sight, an easy assumption is that Sequencing and Navigation rules would be sufficient to implement the defined adaptation model. However, a deeper study of the standard pointed out some limitations, so special solutions had to be adopted.

First of all, the preconditions for each content, defined on the basis of the goal assessment, allow an item to be skipped or hidden. In the adaptive model, instead, it was necessary to modify the sequence of the activities and the item to be shown according to the preconditions. Moreover, even though the standard claims that for each item different conditions could be defined (from 1 to n), during the implementation of the model it was found that more than two conditions caused problems in their evaluation. Another limitation is that the SCORM is not able to store information in a global variable that is common to all the items in the whole learning path. The only way to do this is to store it as a note for the learner, but this does not allow the information to be processed in run time. Moreover, in order to allow non expert users to build a SCORM compliant content package, an authoring tool is needed. This paper addresses all these limitations.

IV. Revision of the SCORM Standard

The defined adaptation model has been developed using the SCORM 2004 standard 4th edition [1]. A study of the literature in this field revealed some researches aimed at adapting the functionalities offered by the SCORM [4, 11]. These researches aim at proposing extensions of the standard by creating a new version of the SCORM but this is not always compliant with the ADL version. As a consequence of this trend, the increasing number of personalised versions of the standard causes a loss of interoperability between LOs and LMSs. For the sake of preserving this interoperability our research was focused on the client side of the content package navigation, leaving the standard unaltered. This is the main strength of the present research work, that extends the functionalities of the standard without changing its interoperability.

A. SCORM implementation details

In a SCORM compliant LO it is possible to distinguish two main logical components: the content and the navigational path. The SCORM standard in the CAM (Content Aggregation Model) discusses how the content should be organised and the SN (Sequencing and Navigation) how the user may navigate it. Both information items have to be included in the content package file Manifest in order to allow the content to be launched by the LMS. The navigational path within a LO is defined according to the learner's activities, which are controlled by the RTE (Run Time Environment) and managed using Javascript API, which allows interaction between the LO and the LMS. To see how this research uses the SCORM standard to build an adaptive content package, while preserving the LO interoperability, the main SCORM elements need to be examined in more detail. Moreover, section II includes a detailed list of the API functions defined in the research in order to implement the adaptation model.

1) CAM Elements

The CAM elements describe the content organisation. The items serving in this research are:

- Organization: this contains the whole structure of the course. This element was important to define the control mode in the learning path;
- Item: this indicates that a node belongs to a specific Organization and indentifies a SCO. The attribute *identifier* allows the resource to be identified. The value is used to define which resource should be associated to the item during the user interaction.

2) Sequencing and Navigation Elements

The SN elements in the file Manifest used to define the navigational path in the content structure are:

- adlnav:hideLMSUI: this allows the common navigational buttons (next and previous) of the LMS user interface to be hidden. The element has been used in order to define the rules that inhibit some navigational functions and to implement navigational actions among the SCOs according to the user's interaction;
- imsss:sequencingRules: is a container of sequencing rules. Each rule describes the behavior that each single object should have. The standard establishes that each item can have an unlimited number of rules. However, using the ADL Learning Management System, usually used to test the content packages, it was found that it is not able to manage more than two rules for each item. For these reasons an API function has been implemented. Moreover, the element imsss:sequencing allows navigational rules among single SCO to be defined. In the case of adaptive content packages, rules have been defined in order to navigate among the nodes of the hierarchy according to the results of the Comprehension Tests;
- imsss:objectives: is a parent element that contains all the learning goals of a specified activity (imsss:objective). It includes both the primary objective and secondary objectives, if there are any. The primary objective element identifies the objective that contributes to the rollup associated with the current item, starting from the Sequencing rules of other items. The secondary objectives allow other items belonging to the same learning path to be evaluated in order to modify the path according to the pre and post condition verification. For example, let us suppose that after a first test to assess the learner's knowledge the learning path could go two different ways: SCO A (containing the first lesson of the course) and SCO B (containing the lesson on the requisite prior knowledge for the course). In this case, the item that contains the SCO A will control, through a secondary objective "test", that for the item with the primary objective "test" a sufficient score has been achieved to allow the SCO A to be supplied, otherwise SCO B will be supplied;
- imsss:MapInfo: defines the mapping of an activity's local objective information to and from a shared global objective. This information is important in order to be able to use a specific item as the reference for the learning goal of another SCO;
- imsss:PreConditionRule: are rules that allow a specific SCO to be visualised or hidden. Those preconditions are evaluated by the LMS before the Initialize instruction is called. In the adaptive Content Package these rules are important in order to visualize the right SCOs according to the student's learning style;
- imsss:ruleConditions: is the container for the set of conditions to be applied as pre-condition, post-condition and exit condition rules. In other words, it contains a set of elements named imsss:ruleCondition, which contain the conditions to be evaluated;
- imsss:PostConditionRule: these are rules applied when an attempt is made to terminate the activity. In the case of the adaptive package the post conditions are important in order to control the navigational path after the CT has been completed. In order to show an item in a unit that has just been finished (if the user failed the test) the post condition rule modifies the control flow, going back on the tree activities and proposing the same didactic unit in a different presentation mode. Otherwise, if the user passed the test then the control flow goes on with the tree activities;
- imsss:RuleAction: represents the intended action or behavior that the package should have after the evaluation of a set of specific condition sequencing rules. The attribute *action* represents the desired

sequencing behavior (skip, disabled, hiddenFromChoice, etc.) if the rule condition is evaluated as true;

- imsss:ControlMode: is the container for the sequencing control mode information including descriptions of the types of sequencing behaviors specified for an activity. In other words, it allows the content developer to define how the navigation requests should be applied to an item and how the item activities should be considered while processing sequencing requests. The attributes used are: *Choice*, that indicates that a choice of navigation request is permitted (True or False) to target the child of the activity; *choiceExit* indicates whether the user is permitted to terminate the activity (True or False) if a navigation request is processed; *Flow* indicates that the control flow can go on (True or False) to the child of the activity; *forwardOnly* indicates that backward targets (in terms of the Activity Tree) are not permitted (True or False) by the child of this activity. Moreover, it is important to notice that in the adaptive content package, the Previous button has been hidden in order to prevent the learner from navigating freely in the learning content;

3) Run-Time Enviroment elements

The elements that have to be implemented in order to build the adaptive content package through API functions are:

- *cmi.score.scaled*: this indicates the learner's performance. The value is used to define the navigational path in the tree activities;
- *cmi.completion_status*: indicates whether the learner has completed the SCO. Since the SCORM does not limit the ways to complete an activity, in the adaptive content package it will have different values (complete or not complete) according to the type of SCO (that can be CT, content, reinforcement or a cognitive style test)
- *cmi.success_status*: indicates whether the learner has mastered the SCO. Also in this case, the way to conclude the activity is defined on the basis of the type of SCO;
- *adl.nav.request*:. allows the SCO to indicate a navigation request for processing by the LMS when the SCO is terminated. In the adaptive content package this element has been used to allow the Exit button to terminate the activity and contemporarily store the learner's progress in the content package.

B. Extension of API functions

As pointed out in the previous sections, in order to allow the LMS to interact efficaciously with the adaptive content package, new API functions have been defined. The strengths of this work are that the API functions affect only the client side of the package, leaving the standard and the interoperability with the LMS unaltered. The defined API functions exploit the native functions of the standard to share data with the LMS.

The defined functions are:

- *ris_ga()*: a function that calculates the result of the first part of the ICSQ questionnaire in the cognitive style SCO. It evaluates the score for the first 9 questions, aimed at defining if the user's cognitive style is wholist or analytic.
- *ris_vv*: is the twin function of the *ris_ga,* which aims at calculating the score for the second part of the questionnaire aimed at defining if the user's cognitive style is verbalizer or imager. The results of both functions are stored as global variables that are used for the *cognitive_style()* function in order to update the user profile.
- *cognitive_style():* builds the user profile that contains the ordered list of the learner's preferred cognitive styles. According to the first cognitive style stored in the user profile, the value of cmi.score.scaled in the RTE is set. The cmi.score.scaled value is the measure threshold that allows a SCO to be activated. The values are chosen in these ranges: [0, 0.25] for the analytic style, [0.25, 0.50] for the global, [0.50, 0.75] for the verbalizer and [0.75, 1] the imager. These ranges represent a way to encode the cognitive styles in the content package. An in-depth study was made in order to assign this value to each user's cognitive style and thus promote an efficient adaptation of the content package. In other words, the cmi.score.scaled value is updated according to the result of the CT in order to allow the content package to use the same cognitive style (if the CT is passed) or a different one (if the CT is failed). Moreover, it stores the user's profile in a cookie so as to allow the course to be suspended and restarted at any time.
- *TestSco()*: calculates the result of the CT and updates the user's profile accordingly. In other words, using the function *Style_reminder()* and on the basis of the user profile, if the student pass the CT the next didactic unit is shown using the same cognitive style, otherwise the reinforcement content will be supplied. In order to distinguish if a content SCO or a reinforcement SCO should be supplied, the score.scaled value is used. If it is positive the content will be supplied as discussed in *cognitive_style(),* otherwise a negative value is assigned according to the above described ranges: [-1, -0.75] analytic style, [-0.75, -0.50] global style, [-0.50, -0.25] verbalizer style and [-0.25, 0] for the imager.
- *Style_reminder()*: updates the cookies and the user profile according to the user interactions.
- *The Cookie Functions* are a set of functions used to write, read and to verify if the browser allows the cookies to be saved.

V. THE AUTHORING ENVIRONMENT

In order to facilitate the building of an adaptive content package our research has also defined some interventions that could be made in one of the authoring environments used for building SCORM packages: RELOAD.

The RELOAD Project [12] is funded under the JISC (Joint Information Systems Committee) Exchange for Learning Programme (X4L) [16]. *The project focuses on the development of tools that are based on emerging learning technology interoperability specifications*. The main goals of the project are to facilitate the building, diffusion and reuse of LOs. Thus, the project offers a suite of software all of which is SCORM compliant.

The starting idea in our research was to propose the introduction of some facilitations into the RELOAD authoring environment that could support non expert SCORM users in building an adaptive content package. First of all, in order to modify the Organization of the package an option "Add adaptivity" could be added to its shortcut menu, which will be able to add XML code into the Manifest of the package as the child of `organization`.

```
<imsss:sequencing>
  <imsss:controlMode choice="false" choiceExit="true"
      flow="true" forwardOnly="false"
      useCurrentAttemptObjectiveInfo="true"
      useCurrentAttemptProgressInfo="true" />
  <imsss:rollupRules rollupObjectiveSatisfied="true"
      rollupProgressCompletion="true" objectiveMeasureWeight="1" />
</imsss:sequencing>
```

Moreover, the following XML code will also be generated and stored in the Sequencing Collection

```
<imsss:sequencingCollection>
  <imsss:sequencing ID="common_seq">
    <imsss:rollupRules rollupObjectiveSatisfied="true"
          rollupProgressCompletion="true" objectiveMeasureWeight="1" />
    <imsss:deliveryControls tracked="true" completionSetByContent="false"
          objectiveSetByContent="true" />
  </imsss:sequencing>
</imsss:sequencingCollection>
```

After this procedure, the user will add the adaptivity to all the items within the content package. Also in this case, a new option "Add adaptivity" will be added in the shortcut of the item. At this point, the user should choose the type of item that should be added to the content package (Figure 3): cognitive style test, content SCO, reinforcement SCO or Comprehension Test.

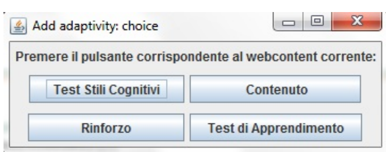


Figure 3. Choosing the type of SCO to be added to the adaptive content package (the labels of the buttons are cognitive style test, content SCO, reinforcement SCO, Comprehension Test)

A. *Cognitive style test*

If the user wishes to add an item containing the test to define the user's cognitive style in the tag <item> the following XML code will be generated.

```
<adlnav:presentation>
  <adlnav:navigationInterface>
    <adlnav:hideLMSUI>previous</adlnav:hideLMSUI>
  </adlnav:navigationInterface>
</adlnav:presentation>
<imsss:sequencing IDRef="common_seq">
  <imsss:objectives>
    <imsss:primaryObjective satisfiedByMeasure="false" objectiveID="obj_pretest">
      <imsss:mapInfo targetObjectiveID="map_pretest" readSatisfiedStatus="true"
      readNormalizedMeasure="true" writeSatisfiedStatus="true" writeNormalizedMeasure="true" />
    </imsss:primaryObjective>
  </imsss:objectives>
</imsss:sequencing>
```

B. *Content SCO*

If the user wishes to build a SCO content, it will be necessary to define the presentation mode used to build the content. In figure 4 the user should choose among the different styles Analytic, Wholist, Verbalizer and Imager according to the results of the ICSQ questionnaire.

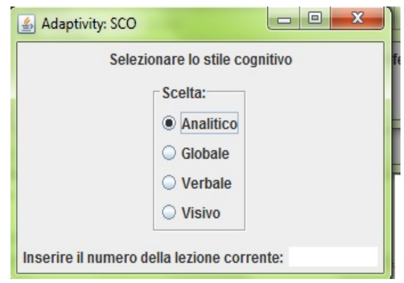


Figure 4. Selection of the cognitive style for the content and reinforcement SCOs (the option button labels are the different cognitive styles: Analytic, Wholist, Verbalizer and Imager)

The XML code corresponding to the Content SCO will be the following.

```
<adlnav:presentation>
  <adlnav:navigationInterface>
    <adlnav:hideLMSUI>previous</adlnav:hideLMSUI>
  </adlnav:navigationInterface>
</adlnav:presentation>
<imsss:sequencing IDRef="common_seq">
</imsss:sequencing>
```

After the selection of the cognitive style for the content (in the example analytic) the code inserted in the Manifest will be the following:

```
<imsss:sequencingRules>
  <imsss:preConditionRule>
    <imsss:ruleConditions conditionCombination="any">
      <imsss:ruleCondition operator="not" measureThreshold="0"
        referencedObjective="test" condition="objectiveStatusKnown" />
      <imsss:ruleCondition operator="noOp" measureThreshold="0"
        referencedObjective="test" condition="objectiveMeasureLessThan" />
    </imsss:ruleConditions>
    <imsss:ruleAction action="skip" />
  </imsss:preConditionRule>
  <imsss:preConditionRule>
    <imsss:ruleConditions conditionCombination="any">
      <imsss:ruleCondition operator="not" measureThreshold="0.25"
         referencedObjective="test" condition="objectiveStatusKnown" />
      <imsss:ruleCondition operator="noOp" measureThreshold="0.25"
         referencedObjective="test" condition="objectiveMeasureGreaterThan" />
    </imsss:ruleConditions>
    <imsss:ruleAction action="skip" />
  </imsss:preConditionRule>
</imsss:sequencingRules>
<imsss:objectives>
  <imsss:primaryObjective satisfiedByMeasure="false" objectiveID="obj_sco1_analitico">
    <imsss:mapInfo targetObjectiveID="map_sco1_analitico" readSatisfiedStatus="true"
      readNormalizedMeasure="true" writeSatisfiedStatus="true" writeNormalizedMeasure="true" />
  </imsss:primaryObjective>
  <imsss:objective satisfiedByMeasure="false" objectiveID="test">
    <imsss:mapInfo targetObjectiveID="map_test1" readSatisfiedStatus="true"
      readNormalizedMeasure="true" writeSatisfiedStatus="true" writeNormalizedMeasure="true" />
  </imsss:objective>
</imsss:objectives>
```

C. *Reinforcement SCO*

The code inserted is similar to that described for the content SCO. The main difference is in the name of `primaryObjective` and in the values and ranges used for the Pre-conditions rules that are defined according to the description in section IV.

D. *Comprehension Test*

In this case, the reference number to the didactic unit related to the comprehension test has to be added in the Manifest. This will allow the LMS to create a link between the content or the reinforcement SCO and the CT related to them.

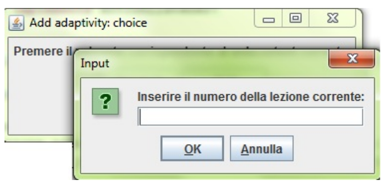


Figure 5. Introduction of the reference number to the didactic unit related to the CT

The code in the Manifest will be the following:

```
<adlnav:presentation>
  <adlnav:navigationInterface>
    <adlnav:hideLMSUI>previous</adlnav:hideLMSUI>
  </adlnav:navigationInterface>
</adlnav:presentation>
<imsss:sequencing>
  <imsss:sequencingRules>
    <imsss:postConditionRule>
      <imsss:ruleConditions conditionCombination="all">
        <imsss:ruleCondition operator="not" condition="completed" />
      </imsss:ruleConditions>
      <imsss:ruleAction action="previous" />
    </imsss:postConditionRule>
  </imsss:sequencingRules>
</imsss:sequencing>
```

And this will be the code serving to define the relationship between the content and the test.

```
<imsss:objectives>
  <imsss:primaryObjective satisfiedByMeasure="false" objectiveID="obj_test1">
    <imsss:mapInfo targetObjectiveID="map_test1" readSatisfiedStatus="true"
       readNormalizedMeasure="true" writeSatisfiedStatus="true" writeNormalizedMeasure="true" />
  </imsss:primaryObjective>
</imsss:objectives>
```

These functionalities have been designed, implemented and integrated in the RELOAD environment.

VI. CONCLUSIONS AND FUTURE WORKS

The paper describes a method for building an adaptive content package using the SCORM standard. The learning content will be adapted during the interaction according to the user's preferred cognitive style. The basic idea underlying the defined adaptive model is that an individual can code, organize, and run content using different cognitive styles. During the learning process, on account of the domain, the motivation, etc., the preferred cognitive style can change. The theoretical model was then applied to learning content using the SCORM standard in order to assure the interoperability of the LOs. The main novelty of this work is that the SCORM specifications were not modified at all, but their primitives were used in order to implement our adaptation model. Additionally, some API functions were defined to allow the content package and LMS to apply the adaptivity on the client side of both. Moreover, a set of interventions has been built to allow the adaptive content package to be built, even by a non – expert user, using one of the most popular authoring tools, RELOAD. The next step of our research will be to set up experimentation of both the adaptation model and the functionalities in the RELOAD environment.

ACKNOWLEDGMENT

This work was supported by Regione Puglia under grant DIPIS and by the University of Bari under grant ex 60%.

REFERENCES

[1] Advanced Distributed Learning Initiative, SCORM 2004 4th Edition Version 1.1 Overview. Available at http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/2004%204th%20Edition/Overview.aspx

[2] Brusilovsky P., 1996, Methods and techniques in adaptive hypermedia," User Modelling and User-Adapted Interaction, 6(2-3):87 129.

[3] Calcaterra A., Antonietti A., Underwood J., 2005, Cognitive style, hypermedia navigation and learning, Computers & Education, 44, 441–457.

[4] Chang H. P., Shih T. K., Li Q., Wang C., Wang T., Chao L. R., An Adaptive Caching Strategy for m-Learning Based on SCORM Sequencing and Navigation. Journal of World Wide Web, Volume 11 Issue 3, September 2008.

[5] Chou H. W., 2001, Influences of cognitive style and training method on training effectiveness, Computers & Education, 37, 11–25.

[6] Di Bitonto P., Roselli T., Rossano V., Monacis L, Sinatra M., MoMAt: a Mobile Museum Adaptive tour, In T. Chan, D. G. Sampson, U. Hoppe, R. Pea, C. C. Liu (Eds), IEEE International Conference on Wireless, Mobile and Ubiquitous Technologies in Education, Kaohsiung, Taiwan, Republic of China, April 12-16, 2010, pp. 156-160, IEEE Computer Society Press, CA: Los Alamitos

[7] De Beni R., Moè A., and Cornoldi C., AMOS. Abilità e motivazione allo studio: valutazione e orientamento. Questionario sugli stili cognitivi. Erickson: Trento, 2003.

[8] Monacis L., Finamore R., Sinatra M., Di Bitonto P., Roselli T., Rossano V., 2009, An Adaptive Learning Object SCORM standard. International Workshop on Distance Education Technologies (DET'2009), part of the 14th International Conference on Distributed Multimedia Systems, San Francisco Bay, USA, pp. 222-227.

[9] Notargiacomo Mustaro P., Silveira I. F., 2006, Learning Objects: Adaptive Retrieval through Learning Styles, Interdisciplinary Journal of Knowledge and Learning Objects, 2, 35-46.

[10] Rasmussen K. L., Davidson-Shivers G. V., 1998, Hypermedia and learning styles: can performance be influenced?, Journal of Educational Multimedia and Hypermedia, 7(4), 291-308.

[11] Rey-López M., Díaz-Redondo R. P., Fernández-Vilas A., Pazos-Arias J. J., López-Nores M., García-Duque J., Gil-Solla A., Ramos-Cabrer M., T-MAESTRO and its authoring tool: using adaptation to integrate entertainment into personalized t-learning. Journal of Multimedia Tools and Applications, Volume 40 Issue 3, December 2008, pp. 409–451.

[12] RELOAD project, http://www.reload.ac.uk/

[13] Riding R. J., and Ashmore I., Verbaliser-imager learning style and children's recall of information presented in pictorial versus written form, Educational Psychology, vol. 6, 1980, pp. 141-145.

[14] Riding R. J, and Mathias D., Cognitive Styles and preferred learning mode, reading Webster attainment and cognitive ability in 11 year-old-children Educational Psychology, vol. 11, 1991, pp. 383-393.

[15] Riding R. E., and Cheema I., Cognitive Styles: An Overview and Integration, Educational Psychology, vol. 11, 1991, pp. 193-215.

[16] http://www.x4l.org

Applying Case-Based Planning to Personalized E-learning

Antonio Garrido
Universitat Politècnica de València
Camino de Vera s/n, 46022
Valencia (Spain)
Email: agarridot@dsic.upv.es

Lluvia Morales
University of Granada
Cuesta del Hospicio s/n, 18070
Granada (Spain)
Email: lluviamorales@decsai.ugr.es

Ivan Serina
Free University of Bozen-Bolzano
viale Ratisbona 16, 39042
Brixen-Bressanone (Italy)
Email: ivan.serina@unibz.it

***Abstract*—Sequencing of Learning Objects (LOs) has been an important issue in the last decades. From a technical perspective, we can take advantage of Artificial Intelligence (AI) planning techniques that allow us to adapt these sequences to pedagogical and students' requirements. However, there is neither a standard way to represent and compile such LO knowledge into a planning model, nor an optimal way to deal with changes during the execution of the previously adapted learning sequences. In this paper, we propose a general and effective approach to automatically extract information from the LOs to create planning domains, which are then solved by case-based plan merging techniques that are also used as a recommender system. This way of proceeding allows the teacher to store, and reuse, the best learning routes for each student's profile and course objectives. When discrepancies on the student's profile or state are detected during the course execution, the system assists the teacher in readapting, repairing or improving the route to fit the new objectives.**

I. Introduction

E-learning is becoming a high-impact innovative topic as it offers a promising way to facilitate and enhance the learning process by combining LOs to create flexible courses. In particular, AI planning techniques have shown to be very adequate to generate fully tailored routes of LOs according to pedagogical theories [1]. More specifically, the planning objective in an e-learning setting is to provide a student-centered solution (i.e. personalized learning) by offering a flexible learning process where courses and LOs are tailored to the specific needs, learning styles, objectives, background and, in general, the profile of each student [1], [2], [3], [4].

We focus on the application of Case-Based Planning (CBP) techniques [5] for the definition, memorization, retrieval and adaptation of learning routes. However, defining a high quality learning route is difficult; it depends on many variables, involving the LOs, the available resources, the students, and their preferences/learning styles. Thus, our goal is to help teachers choose the most suitable learning route and (semi-)automatically adapt it in accordance with the students' goals and individual features. Hence, our approach contributes with:

1) A mapping of learning courses into planning domains by using an automated translation of e-learning templates [2] into PDDL durative actions [6].
2) A CBP repository for planning domain+file compilations that contains students' learning information, which are successively analyzed by our case-based planner [5].
3) A flexible way to make curriculum authoring easier because teachers can easily retrieve, improve and reuse courses previously generated.
4) A simple translation of the resulting sequence of LOs into another standard representation, namely learning design, that provides a usable manifest for on-line learning platforms, thus closing the e-learning cycle.

The paper is structured as follows. First, we introduce the role of planning for learning routes personalization, together with some basic background on planning and CBP techniques. Next, we present our general approach, describing its structure and main elements: i) required metadata information, ii) compilation of this metadata into a PDDL description, iii) planning personalized learning routes, iv) integration with Learning Management Systems (LMSs), and v) navigation and monitoring of the routes. Finally, we discuss the lessons learnt and raise the most important issues that are necessary for a full support of planning within an e-learning setting.

II. The role of AI planning in personalized e-learning

A. Using planning for personalization

The traditional mode of instruction (one-to-many lecturing, or one-to-one tutoring), which is adopted in conventional education, cannot fully accommodate the different learning and studying styles, strategies and preferences of diverse students. Consequently, offering personalized learning routes to individual students' needs and profiles is essential to promote better learning initiatives in e-learning.

Educational systems have always aimed at ensuring flexibility within course personalization. Many authors have addressed this flexibility using different techniques, such as adjacency matrices, integer programming models, neural networks, graph-based sequencing procedures and, more recently, intelligent planning techniques [4], [7], [8], [9], [10], [11]. We use planning (and scheduling) techniques to bridge the gap between the purely e-learning necessities and the accommodation of temporal+resource constraints of the environment, to make the course applicable in a real scenario, which is usually missing in previous approaches. We go beyond the traditional e-learning insights and give support not only to adaptation and LO sequencing, but also to scheduling constraints and multi-criteria optimization metrics. This raises a challenge for a successful integration with LMSs that facilitate the dynamic navigation of contents/LOs, monitor the students' progress when following their proposed learning routes, check whether some discrepancies appear and react to them to adapt the routes to the new necessities. And this monitoring part is also missing in most current approaches. In this monitoring line, [4] uses specific planning capabilities to: i) wait for specific information in critical points, ii) update the students' profile+state information, and iii) resume planning until the next critical point. Our approach also manages dynamics in e-learning, but when a discrepancy appears we use a case-based planner which, rather than creating a new route from scratch, retrieves the best element that fits the current requirements from a CBP repository and adapts it if necessary.

B. Some background on planning

AI planning is the task of finding a solution within a search space defined by the application of legal actions (grounded operators). Each action has preconditions and effects. Preconditions need to be satisfied before the action is applied, whereas effects describe the results (in terms of what changes) after the action is executed. More formally, a planning problem is defined as the tuple $\Pi = \langle \mathcal{F}, \mathcal{I}, \mathcal{G}, \mathcal{O}, \mathcal{M} \rangle$, where:

- $\mathcal{F}$ represents the fluents of the problem, defined as ground atomic propositional formulae or numeric variables;
- $\mathcal{I}$ represents the initial state;
- $\mathcal{G}$ represents the goal conditions to be achieved;
- $\mathcal{O}$ is a finite set of operators, where each operator is defined as a tuple $\langle Pre, dur, Effs \rangle$. Each precondition in Pre defines a propositional or numeric precondition that expresses a constraint as a tuple $\langle \texttt{f-exp1}, \texttt{binary-comp}, \texttt{f-exp2} \rangle$, where `f-exp1` and `f-exp2` represent functional expressions and $\texttt{binary-comp} \in \{<, \leq, =, >, \geq\}$ is a binary comparator. $dur \in \mathcal{R}^+$ is a positive duration. Finally, each effect in $Effs$ can be either a positive or negative propositional effect or a numeric effect as a tuple $\langle \texttt{f-head}, \texttt{assign-op}, \texttt{f-exp} \rangle$, where `f-head` represents a numeric variable, and $\texttt{assign-op} \in \{:=, +=, -=, *=, /=\}$ is an assignment operator which updates the value of `f-head` according to the functional expression `f-exp`.
- $\mathcal{M}$ is a multi-criteria metric to be minimized/maximized, thus addressing optimal planning.

The planning goal is to come up with a proper plan, i.e. a partially ordered sequence of actions, whose execution will transform the initial state $\mathcal{I}$ into a final state in which all goals $\mathcal{G}$ are satisfied, while also optimizing the metric $\mathcal{M}$.

C. Description of CBP techniques

In CBP, previously generated plans are stored as cases in memory and can be reused to solve similar planning problems in the future. CBP can save considerable time over planning from scratch, thus offering a potential (heuristic) mechanism for handling intractable problems. Similarly to other Case-Based Reasoning (CBR) systems, CBP is based on two assumptions on the nature of the world [12]. The first assumption is that the world is regular: similar problems have similar solutions. As a consequence, solutions for similar problems are a useful starting point for new problem-solving. The second assumption is that the types of problems an agent encounters tend to recur; hence future problems are likely to be similar to current problems.

In general, the following steps are executed when a new planning problem must be solved by a CBP system:

1) *Plan Retrieval* to retrieve cases from memory that are analogous to the current (*target*) problem.
2) *Plan Adaptation* to repair any faults found in the new plan.
3) *Plan Revision* to test the solution new plan π for success and repair it if a failure occurs during execution.
4) *Plan Storage* to eventually store π as a new case in the case base.

Following the formalization proposed by [13], a *planning case* is a pair $\langle \Pi_0, \pi_0 \rangle$, where Π_0 is a planning problem and π_0 is a plan for it, while a plan library is a set of cases $\{\langle \Pi_i, \pi_i \rangle | 1 \leq i \leq m\}$.

To the end of applying the reuse technique, it is necessary to provide a plan library from which "sufficiently similar" reuse candidates can be chosen. In this case, "sufficiently similar" means that reuse candidates have a large number of initial and goal facts in common with the new instance. However, one may also want to consider the reuse candidates that are similar to the new instance after their objects have been systematically renamed; this corresponds to identifying a mapping between the objects of the reuse candidate and the objects of the new instance such that the number of common goal facts is maximised and the additional planning effort to achieve the initial state of the plan library is minimised. This is extremely important in our context, where teachers could decide to reuse a course, or a part of a course, that has been adopted by their colleagues or by themselves previously. Obviously, the students will not be the same, but if their profile, their goals and the resources available are similar to the corresponding ones in the case base, our system will be able to propose a new high quality learning path to the teacher with a limited number of changes w.r.t. the one stored in the library. Our approach uses an approximate evaluation based on kernel functions [5] in order to compute an appropriate mapping between the students and the objects of the reuse candidate and the corresponding ones of the current instance, which can be computed in polynomial time.

The plan adaptation system consists in reusing and modifying previously generated plans to solve a new problem and overcome the limitations of planning from scratch. Moreover, any kind of planning system that works in dynamic environments has to take into account failures that may arise during plan generation and execution. In this respect, CBP is not an exception; this capability is called plan revision and it is divided into two subtasks: evaluation and repair. The evaluation step verifies the presence of failures that may occur during plan execution when the plan does not produce the expected result. When a failure is discovered, the system may react by looking for a repair or aborting the plan.

After finding the plan from the library and after repairing it with the plan adaptation techniques, the solution plan can be inserted into the library or be discarded.

In order to improve the efficiency of the system and reuse as much as possible parts of previously executed plans we have adopted *plan merging techniques* [14], which are based on the well-known *divide and conquer* strategy. In order to apply this strategy, our system must accomplish two further subtasks: *problem decomposition* and *plan merging*. The problem decomposition is performed identifying the set of actions and the initial facts needed for a single goal and storing them in the case base as a new problem instance (if not already present); moreover these new instances remain related to the original solution plan in order to maintain a statistic of their effective usage. The stored (sub)cases are then used in the merging phase in order to identify a single global plan that satisfies all goals. We progressively identify the unsatisfied goals and the corresponding (sub)cases that allow to satisfy them, giving the preference to the (sub)plans that allow us to improve the plan metric and that have been successful in a greater number of times in analogous situations.

III. Explanation of our approach

Figure 1 provides the overall schema of our approach, which involves several technological issues: i) use of common LO repositories and modeling tools, ii) compilation of planning domains, ii) algorithms for students' information acquisition, iii) application of CBP as solving techniques, iv) visualization of learning designs on LMSs, and v) monitoring students' progress when following the course to detect discrepancies. We now explain each of these issues in more detail.

A. LO metadata information and its use in modeling courses

Modeling an instructional course involves selecting LOs and defining how they are interrelated (i.e. their causal dependencies), which is done by the course designer. This highly relies on the metadata information defined in the LO itself. Metadata specification for LOs is usually specified in an XML standard format, such as LOM [15]. This specification has many useful entries for pedagogical theories, but only a few of them are really essential to support planning personalization. First, we need the technical platform requirements, seen as the particular resources for the LOs. Second, we need the educational information about the student's learning style (profile), difficulty of the LO and its typical learning time (i.e. duration). Finally, we need the relations as the content dependencies which comprise hierarchical structures and orderings among LOs. The hierarchical structures use the *IsPartOf* relationship, which represents an aggregation of LOs. For instance, in Figure 2 "LO4 *IsPartOf* LO1", which means that LO4 is compulsory in a learning route that includes the (higher level) LO1. There are also three types of ordering

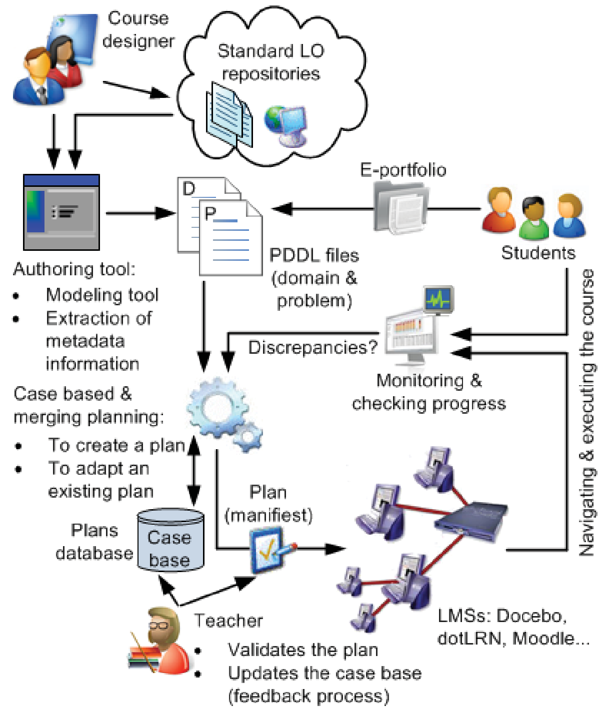


Figure 1. Overall structure of the system

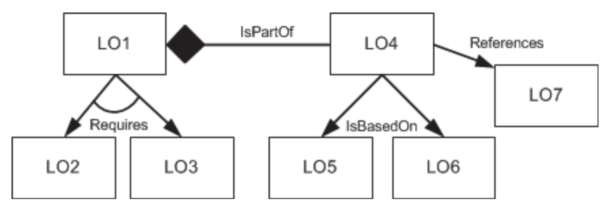


Figure 2. An example of LOs that use the LOM relations

relations to represent causal dependencies: i) *Requires*, as a conjunctive precondition; ii) *IsBasedOn*, as a disjunctive precondition; and iii) *References*, as a (soft) recommended precondition that may involve a kind of incentive or learning reward. In Figure 2, both "LO2 and LO3" are required by LO1 but only "LO5 or LO6" are required by LO4. On the contrary, the *References* relation does not denote a hard precondition but a soft one to complete a LO before proceeding with the next one (e.g. LO7 in Figure 2).

B. Planning compilation: from course+students' information to planning

The planning compilation is a knowledge-engineering task that is automated by using a mapping between e-learning metadata information and planning elements. This means to express LOs as PDDL operators and to model the students' profile and learning goals as part of a PDDL problem.

The general algorithm to compile a planning domain is to iterate all over the LOs and generate one PDDL planning operator per LO, as detailed in the mapping of Table I. This generation relies on a closed world assumption, and if new

LOs are to be used the domain must be recompiled. Anyway, this compilation is very efficient, as each operator comprises five entries which are automatically extracted from the values of the LO metadata specification:

- A unique name taken from the LO name.
- One parameterized student, which facilitates the application of the same operator to different students.
- The LO duration (*typicalLearningTime* field) to model a PDDL durative action [6].
- The preconditions to support all the dependency relations, according to the semantics of conjunctive (*Requires* and *IsPartOf*), and disjunctive (*IsBasedOn*) preconditions. A dummy precondition `(not (action-name ?s done))` is used to avoid planning the same action more than once. Furthermore, other educational requirements, such as the intended student's role, his/her learning style, the difficulty of the LO, the required language for the LO or multimedia requirements can be modeled both as strong (*Requires* or *IsBasedOn*) or soft (*References*) preconditions.
- The effects to represent the LO outcome, i.e. having it done. They also include a `reward` to offer a full support for LO adaptation to the students, that is, how the learning resource type of each LO fits the student's learning style, based on Honey-Alonso's, Felder's or any other classification. Optionally, the compilation can include numeric expressions or resource costs, as are common in planning and scheduling, and even a particular sequencing used to assist hierarchical decomposition.

According to this compilation, the PDDL durative action for LO4 of Figure 2 is:

```
(:durative-action LO4
 :parameters (?s - student)
 :duration (= ?duration 5) ;the typicalLearningTime
 :condition (and
  (at start (not (LO4 ?s done)))
  (or
    (at start (LO5 ?s done))
    (at start (LO6 ?s done)))
  (or
    (at start (LO7 ?s done))
    (at start (true))))
 :effect (and
    (at end (LO4 ?s done))
    (at end (increase (reward)
                      (reward-value ?s Equipment)))))
```

After the planning domain is generated, the planning problem is compiled by extracting the relevant student's characteristics from his/her e-portfolio. The planning problem includes the initial state, the goals and, if needed, the metric to optimize. The initial state represents the students' profile and background, the language of the course, and some other information (e.g. special equipment or resource availability). The goal is to pass the entire course or a part of it. An example of the planning problem for a `student1` that wants to attain both LO1 and LO5 outcomes in the shortest time is:

```
(:init
 ;Profile and learning style
  (learning-style student1 theoretical)
  (performance student1 high)
 ;Academic trajectory (background)
  (language-level es student1 high)
  (language-level en student1 high)
  (availability student1 much))
(:goal (and
  (LO1 student1 done)
  (LO5 student1 done)))
(:metric minimize (total-time)) ;shortest time plan
```

C. Application of plan merging techniques

Case-based plan merging techniques have been used in our approach to store learning routes (plans) previously validated by the teachers in the case base. These solution plans can be manually generated by the teachers or simply generated by a domain independent planner such as LPG, SGPlan or by our case-based planner itself, and then validated by the teachers. Moreover, in order to reuse as much as possible parts of previously executed plans, we decompose the solution plans into subparts that allow us to satisfy every single goal and we store them in the case base, if they are not already present.

When a new e-learning planning problem must be solved, we search in the case base if a plan exists that already solves all goals. If such a plan does not exist we apply, as previously exposed, plan merging techniques that progressively identify (sub)plans in the case base that can satisfy the goals. Intuitively, the adaptation consists in reusing parts of the retrieved plans to complete a new one that is similar to the original one. This allows the teachers to easily validate the proposed learning route since they can simply consider and analyze the parts of the learning route that differ from the elements stored in the case base and that have been introduced in order to satisfy, for example, new users' goals or prerequisites of LOs that do not fit with students' requirements, instead of reconsidering the whole learning route. Note that different criteria can guide the definition of a learning route. In our current version we do not only try to find good quality plans that best fit the students' requirements but also to minimize the number of LOs that have been introduced or removed w.r.t. the case base elements. The relative importance of plan quality w.r.t. plan stability can be chosen by the teacher when the case-based planner is executed.

D. Closing the cycle

The global solution plan is then provided to the LMS that identifies, for each student, the instructional design that must be followed by the students under the IMS-CP or SCORM specifications. Moreover, all of these (sub)plans or instructional designs, in addition to those plans of the same topic that can be recommended by the teachers, are additionally stored in the case base, if not already present.

Note that the LMS is fundamental not only to provide the students a sequence of LOs but also to visualize the learning plan to the teachers and allow them to monitor the students' progress, to easily detect significative discrepancies between the current situation and the scheduled activities, to select a single LO or sequences of LOs that can be integrated in

LO metadata item →	PDDL action entry
Identifier	action-name
-	:parameters (?s - student), to model the student
typicalLearningTime	:duration for temporal planning
dependency relations: type of relation: case (*Requires, IsPartOf*): conjunctive (and) case (*IsBasedOn*): disjunctive (or) case (*References*): recommendation the LO: entry	:condition (and (at start (not (action-name ?s done)) if and-precondition: (and ... else-if or-precondition: (or ... for each entry: (at start (action-entry ?s done) ... other optional preconditions for current profile)
model that the action has been done, and increase a reward/utility expression due to learningResourceType (profile adaptation) and the References relation	:effect (and (action-name ?s done), and (increase (reward ?s) (reward-value ?s LRT) ... other optional rewards and/or costs)

Table I
GENERAL LO METADATA MAPPING TO PDDL ACTIONS. IRRELEVANT INFORMATION HAS BEEN IGNORED.

the case base, to manually modify the LOs proposed by the planner and, finally, to validate the instructional design the students have to perform. Furthermore, in order to easily define a new course, the LMS allows the teachers to import whole courses from external repositories; their structure with the relations among their LOs and the corresponding pedagogical motivations are directly stored in the case base, while the LOs are added to the domain description allowing the planner to effectively identify their causal structure and reuse them in different parts of the course.

While the students are interacting with the LMS and performing the previously planned LOs, some changes can occur in his/her profile information, either by external or internal conditions. Some examples of external conditions are:

- The student has taken classroom courses on different languages and increased his/her language level.
- Some prerequisites of LOs of his/her instructional design have been satisfied in other courses.
- The student has new temporal constraints (a new job or gets sick), and now (s)he has less time to accomplish the goals of the course, being unable to perform some activities.
- The student has bought a new computer that has more equipment capabilities than the previous one.
- Some equipment is no longer available, for example the student's computer is broken or a laboratory is busy.

The changes in students' profiles caused by external conditions must be modified directly by the students themselves, or by the teachers using the LMS interface for this purpose. On the other hand, examples of some internal conditions are:

- During the course execution, some tests or questionnaires are applied to evaluate the students' comprehension of the objectives. If this comprehension is negatively evaluated by getting a low score, then it is inferred that the students' performance is decreasing.
- When it is detected that the student uses LOs with a recurrent learning resource type, this behavior indicates that his/her learning style orientation is changing.

These conditions can conventionally and automatically change the students' profile through a set of pedagogical rules. These rules can also be statically defined into the database as default rules that can be modified by the tutor (by using an intuitive interface).

When changes in the student's profile are detected, we simulate the execution of the remaining part of the learning plan in order to identify if it contains flaws, i.e. if the preconditions of LOs and the goals are no longer satisfied (this can be computed efficiently in polynomial time w.r.t. the number of actions in the remaining part of the plan). If an inconsistency is detected, it is highlighted to the teacher and (s)he can decide whether to repair it manually or to ask for a new plan to the case-based planner. If no inconsistency is detected, a new schedule of the remaining LOs is provided to the student in order to better satisfy his/her requirements and time availabilities; note that this new schedule can be simply computed in polynomial time w.r.t. the number of LOs and resource involved, and does not require any kind of validation by the teacher. Moreover, the student and the teacher can also ask the planner if a new plan of better quality, according to the new student's profile and the current resources, can be found. Anyway, when a new plan is computed by our system it must be always validated by the teacher before its execution. And the plan stability w.r.t. previously executed plans is of capital importance to reduce the teacher's overhead.

When the plan execution finishes and all the students' goals are satisfied, the corresponding plan is provided to the case-based system in order to be stored in the case base, if not already present, closing in this way the learning cycle.

IV. DISCUSSION AND FUTURE WORK

Considering the huge number of LO repositories and courses, implemented in different XML standards such as SCORM, LOM or IMS, available on the Web, it is a real challenge to offer a general approach to deal with personalized e-learning routes that fit the students' profiles and styles in a flexible way. But given the strong resemblance between planning and the creation of learning routes, we can apply planning and scheduling techniques to deal with i) course designs, ii) LOs and learning activities sequencing, and iii) adaptation to students' profiles together with pedagogical theories. By defining a versatile mapping (such as the one presented in Table I), we can automatically reason on the time a LO requires, its difficulty, the whole duration of the course, the available time each student has, his/her preferences and learning styles, the resources that are available, and even the level of cooperation among tutors and students. This provides many opportunities to support e-learning personalization and routes adaptation.

Previous features face particular aspects of the whole e-learning picture, but there are other aspects related to the LO metadata specification, course modeling, the role of the teacher once a learning route has been created and when this route is eventually executed by the student. All in all, what we have learnt is that the following issues are important in personalized e-learning settings and can fill, in themselves, their own research lines:

- LO metadata focuses on the specification of educational and pedagogical matters, but important entries to promote personalization are still missing; e.g. ontologies on the use of the relations, use of resources, collaboration on the use of a LO, and definition of temporal constraints.
- Planning techniques need a well-defined domain and problem specification, which is not always easy to generate. We have provided here a general, flexible mapping, but some aspects are still a bit complex to extract from the metadata description. For instance, it is not intuitive how to measure the incentive value, i.e. reward, when a recommendation (soft precondition) holds.
- Teachers are not always *happy* with the learning routes that are automatically generated and feel reluctant to abandon their traditional role of *human planners* when creating learning routes. Therefore, a mixed-initiative approach that allows teachers to make further changes seems the most adequate way of proceeding.
- LMSs need to be extended to provide capabilities for integration with monitoring and checking the students' progress when executing the learning route. So, it is not only a matter of plugging in a planner to find a route, but also to provide additional tools to find out discrepancies (differences between the expected and the real student's state) to adapt the route. And here CBP techniques show very useful to adapt existing routes, at little cost on the average, rather than creating them again from scratch.
- The evaluation of this approach is not simple as it involves many parameters —we refer the interested reader to [16] for further details. But it is important to note that the success of this approach cannot be directly assessed through the students' grades/scores; we mainly aim at improving the students' satisfaction when using personalized routes, which enhances personal instruction.

Our current work focuses on the planning aspects and on the integration with Moodle LMS. On the one hand, we are working on modifying the case base repositories by including the new plans that are progressively generated by our case-based planner [5]. Also, the teacher will be able to modify the standard LO repository, which can be now extended to be not only a repository of LOs but also a course (sequence of LOs) repository. On the other hand, we are trying to offer an integrated approach of planning and Moodle by means of Web services. Although our first attempt relies on Moodle and ILIAS, we also expect to create integration modules for other open sources LMSs, such as dotLRN and Claroline.

V. Acknowledgements

This paper was partially funded by the Consolider AT project CSD2007-0022 INGENIO 2010 of the Spanish Ministry of Science and Innovation, the MICINN project TIN2008-06701-C03-01, the Mexican National Council of Science and Technology, the Valencian Prometeo project 2008/051 and the BW5053 research project of the Free University of Bozen-Bolzano.

References

[1] P. Brusilovsky and J. Vassileva, "Course sequencing techniques for large-scale web-based education," *International Journal Continuing Engeenering Education and Lifelong Learning*, vol. 13, no. 1/2, pp. 75–94, 2003.

[2] A. Garrido, E. Onaindia, L. Morales, L. Castillo, F. S., and D. Borrajo, "Modeling e-learning activities in automated planning," in *Proceedings of the 3rd International Competition on Knowledge Engineering for Planning and Scheduling (ICKEPS-ICAPS 2009)*, 2009, pp. 18–27.

[3] D. Peachy and G. McCalla, "Using planning techniques in intelligent systems," *International Journal of Man-Machine Studies*, vol. 24, pp. 77–98, 1986.

[4] C. Ullrich and E. Melis, "Pedagogically founded courseware generation based on HTN-planning," *Expert Systems with Applications*, vol. 36, no. 5, pp. 9319–9332, 2009.

[5] I. Serina, "Kernel functions for case-based planning," *Artificial Intelligence*, vol. 174, pp. 1369–1406, 2010.

[6] M. Fox and D. Long, "PDDL2.1: an extension to PDDL for expressing temporal planning domains," *Journal of Artificial Intelligence Research*, vol. 20, pp. 61–124, 2003.

[7] L. Castillo, L. Morales, A. Gonzalez-Ferrer, J. Fdez-Olivares, D. Borrajo, and E. Onaindia, "Automatic generation of temporal planning domains for e-learning," *Journal of Scheduling*, vol. 13(4), pp. 347–362, 2010.

[8] A. Garrido and E. Onaindia, "On the application of planning and scheduling techniques to e-learning," in *Proc. 23rd Int. Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA-AIE 2010) - LNAI 6096*. Springer, 2010, pp. 244–253.

[9] N. Idris, N. Yusof, and P. Saad, "Adaptive course sequencing for personalization of learning path using neural network," *Int. J. Advance. Soft Comput. Appl.*, vol. 1(1), pp. 49–61, 2009.

[10] G. Knolmayer, "Decision support models for composing and navigating through e-learning objects," in *Proc. Annual Hawaii Int. Conference on System Sciences (HICSS'03)*, 2003, pp. 31–41.

[11] E. Kontopoulos, D. Vrakas, F. Kokkoras, N. Bassiliades, and I. Vlahavas, "An ontology-based planning system for e-course generation," *Expert Systems with Applications*, vol. 35(1-2), pp. 398–406, 2008.

[12] D. Leake, *Case-Based Reasoning*. Cambridge, Massachusetts: The MIT Press, 1996.

[13] P. Liberatore, "On the complexity of case-based planning," *Journal of of Experimental & Theoretical Artificial Intelligence*, vol. 17(3), pp. 283–295, 2005.

[14] Q. Yang, D. Nau, and J. Hendler, "Merging separately generated plans with restricted interactions," *Computational Intelligence*, vol. 8(4), pp. 648–676, 1992.

[15] LOM, "Draft standard for learning object metadata. IEEE. rev. 16 february 2005," 2002, available at http://ltsc.ieee.org/wg12/files/IEEE_1484_12_03_d8_submitted.pdf (accessed May 2011).

[16] L. Morales, A. Garrido, and I. Serina, "Planning and execution in a personalised e-learning setting," in *CAEPIA (in press)*, 2011.

Educational Concept Maps: a Knowledge Based Aid for Instructional Design

Giovanni Adorni, Mauro Coccoli
Department of Communication, Computer, and Systems Science - University of Genoa
Genoa, Italy
{adorni, mauro.coccoli}@unige.it

Giuliano Vivanet
Department of Pedagogical and Philosophical Sciences
University of Cagliari
Cagliari, Italy
giuliano.vivanet@unica.it

***Abstract*—This paper discusses a knowledge-based model for the design and development of units of learning and teaching aids. The idea behind this model originates from both the analysis of the open issues in instructional authoring systems, and the lack of a well-defined process able to merge pedagogical strategies with systems for the knowledge organization of the domain. In particular, it is presented the Educational Concept Map (ECM): a, pedagogically founded (derived from instructional design theories), abstract annotation system that was developed with the aim of guaranteeing the reusability of both teaching materials and knowledge structures. By means of ECMs, it is possible to design lessons and/or learning paths from an ontological structure characterized by the integration of hierarchical and associative relationships among the educational objectives. The paper also discusses how the ECMs can be implemented by means of the ISO/IEC 13250 Topic Maps standard. Based on the same model, it is also considered the possibility of visualizing, through a graphical model, and navigate, through an ontological browser, the knowledge structure and the relevant resources associated to them.**

Keywords- Instructional design, Knowledge representation, Instructional Authoring Systems, Topic Maps

I. INTRODUCTION

The recent evolution of Web-based technology has dramatically changed the learning processes and, consequently, e-learning. Nowadays, on the one hand learners can access information and knowledge sources that are practically unlimited, can share their learning experiences, and collaborate within dedicated social networks in order to better achieve their instructional objectives. On the other hand, teachers and instructional designers can access large repositories in order to retrieve and share learning assets, and can use authoring tools for designing Web-based educational paths. As a consequence, new models are required in order to support meaningful learning processes, based on the assimilation of new concepts and propositions into existing cognitive structures. The role of subject matter analysis and knowledge reusability has been long recognized by instructional designers with the aim of supporting meaningful learning processes through a careful knowledge selection, organization, and sequencing [1]. In this scenario, one of the most relevant problems concerns the fact that there are no canonical representations of knowledge structures and that a knowledge domain can be structured in different ways, starting from various points of view. As Ohlsson [2] highlighted, this fact has such relevant implications for authoring systems, that it should be stated as the "*Principle of Non-Equifinality of Learning*". According to this, "*The state of knowing the subject matter does not correspond to a single well-defined cognitive state. The target knowledge can always be represented in different ways, from different perspectives; hence, the process of acquiring the subject matter have many different, equally valid, end states*". Therefore it is necessary to re-think learning models and environments in order to enable users to better build represent and share their knowledge.

II. THE EDUCATIONAL CONCEPT MAP MODEL

For the formal representation of the subject matter structure in the generic context of learning environments, the Educational Concept Map (ECM) model (see Figure 1) is herein introduced. This model was developed by keeping into account the following pedagogical and technical requirements [3]: *i) pedagogical flexibility*: the model must be able to describe the structure of instructional contents regardless of a specific learning theory; *ii) learner centrality*: the instructional content design process must be based on students' profile and needs; *iii) centrality of learning objectives:* the instructional content design process must be based on a preliminary definition of the learners' pedagogical objectives; *iv) personalization*: the model must be able to design learning contents and resources in a flexible way, consistently with learners' profile; *v) domain-independence*: the model must be able to describe instructional content regardless of its disciplinary nature; *vi) reusability*: the model must allow to define and de-contextualize learning content structures and to reuse these in other contexts; *vii) interoperability*: the model must be language-independent, so that it can be implemented through different knowledge representation languages and exported in different learning environments; *viii) medium neutrality*: the instructional contents design process must be medium neutral, so that it can be used in different publication formats; *ix) compatibility*: the model must fit in with existing standards and specifications on learning resources; *x) formalization*: the model must describe instructional content according to a formalized model, so that automated processing is possible.

The ECM is a logical and abstract annotation model created with the aim of guaranteeing the reusability of both teaching materials, and knowledge structures. It shifts the generalization level from the contents to the definition of the relevant schema. It derives from fundamental theories of instructional design that can only be briefly sketched in the following. According to Merrill [4] and Gagné [5], once the learner profile is known, the instructional design process should start from the definition of a hierarchical organization of learning objectives describing what learners should know or be able to do at the end of the instruction. Such a hierarchy of learning goals provides the general framework for the selection of contents. The preliminary organization of learning objectives and contents into a hierarchical structure (thus making explicit the logical order of learning content) is also considered in Ausubel [6] works. He introduced the notion of *"advanced organizer"* as a way to provide the cognitive structure or, in other words, the mental scaffolding during meaningful learning processes. According to this theory, new concepts to be taught should be introduced by more general and inclusive concepts.

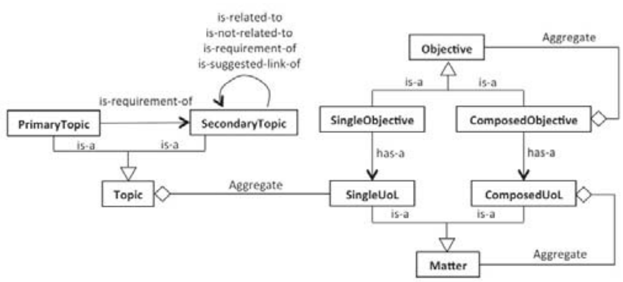


Figure 1 The ECM Model

Referring to such a theoretical framework, the ECM model has been developed by means of an ontological structure characterized by the integration of hierarchical and associative relationships. Firstly, it asks teachers and instructional designers to focus their attention on learners' objectives and on their profile, described in terms of educational background, as well as of learning and cognitive styles. Taking into account these elements, the model suggests how to identify, within the discipline's subject matter, the key concepts and their relationships so as to identify effective strategies of contents presentation and to support the activation of meaningful learning processes [7].

According to the ECM model, a profiled learner has a goal identified by a specific objective (single or multiple) that is achieved by a *Unit of Learning* (UoL), or by a composition of *UoLs*. The *Course Unit* (CU) is the indivisible union of an objective with its unit of learning and can be composed by creating the tree structure of the course (learning units, sub-learning units, etc.). The course units may be connected each other by means of the *Educational Associations* (EA) that may represent a link or a propaedeutic relationship the units have. In particular, four types of EA have been identified: *i) is-requirement-of*: identifying a transitive and propaedeutic association between two or more topics (e.g., it may be used with the aim of specifying the logical order of contents); *ii) is-related-to*: identifying a symmetric association among closely related topics (e.g., it may be used with the aim of creating learning paths without precedence constraints); *iii) is-not-related-to*: identifying a symmetric relation of indifference between two or more topics (e.g., it may be used with the aim of making explicit the absence of association among topics); *iv) is-suggested-link-of*: identifying not-closely related concepts (e.g., this relationship type may be used in order to suggest in-depth resources, internal or external to the contents repository).

These relation types have been selected with the aim of allowing teachers to create different learning paths (with or without precedence constraints among topics). The same types of relationship can be found between topics. The latter are the smaller granularity of the ECM model. They represent the concepts of the domain: any subjects a teacher may want to talk about. Moreover, the units of learning are connected to the topics through two relationships: (i) *has-primary-topic*: where a primary topic identifies the "prerequisites", in other words the concept that a student must know before attending a given unit of learning; (ii) *has-secondary-topic*: where secondary topic identifies the concepts that will be explained in the present unit of learning (this kind of topics will have specific learning materials associated).

More formally, given D the set of all topics in the ECM and U, sub set of D of topics $\{t_1, \ldots, t_n\}$ of the same UoL, one can obtain the set P of primary topics:

$$t \in P \Leftrightarrow \exists t_i \in U : is-requirement-of(t, t_i)$$

In a similar way, one can identify the set L ($L \subseteq U$) of "learning outcomes", i.e., the specific intentions of a unit of learning. Given $t_j \in U$:

$$t_j \in L \Leftrightarrow \forall \neg \exists t_i \in U : is-requirement-of(t_j, t_i)$$

with $i = 1, i-1, i+1, \ldots n$.

So, in the ECM model, a course unit contains an educational objective and a unit of learning. Connected to the *UoL* there are the topics of the conceptual map describing the domain of the course itself. These topics can be primary or secondary, depending on the context they are included in, within the unit of learning. Finally the secondary topics contain the material aid. Such resources, grouped in a unit of learning, enable to reach the objective connected to the *UoL* itself. The *CUs* allow the teachers to create complex and nested structures using the EA.

Furthermore, an ECM can be published on the Web and the relationships suggest the different navigation strategies of the underlying subject matter. More precisely, each Web page related to a topic contains i) one or more links to topics that are considered propaedeutic of this ECM (through the *is-requirement-of* relationship); ii) link to related topics (through the *is-related-to* relationship), iii) links to suggested readings or further topics (through the *is-suggested-link-of* relationship). Such a map can also be used to generate a linearized path, useful to produce a lesson or a document about a given subject matter. In this latter case, a *Suggested Path Strategy* is necessary, to be expressed by means of *is-requirement-of* relationships. Differently from some of the aforementioned EMLs, the ECM model enables authors to define alternative learning content presentation strategies, in order to provide personalized learning paths, depending on specific needs. The definition of this strategy, introduced also in reply to the *non-equifinality problem*, derives from the *Cognitive Flexibility Theory* [8], stating that learning processes (especially in ill-structured and complex domains) require a multiple representation of content and the *criss-crossing* of the conceptual landscape which should be revisited from different directions, with the aim of acquiring an advanced knowledge.

In order to fully exploit the potentiality of such a model, a formalization for the ECM is needed, to effectively represent the model. Among the many possible, the Topic Maps standard [9] has been adopted as the knowledge representation language for the implementation of the above model. Topic Maps (TM) are an ISO multi-part standard designed for encoding knowledge and connecting this encoded knowledge to relevant information resources. The standard defines a data model for representing knowledge structures and a specific XML-based interchange syntax, called XML Topic Maps (XTM) [10]. The main elements in the TM paradigm are: *topic*, *association* and *occurrence*. According to the ISO definition, a Topic is a symbol used within a TM to represent one -and only one- subject, in order to allow statements to be made about the subject. An Association represents a relationship among two or more topics. An Occurrence is a representation of a relationship between a subject and an information resource. In this vein, the TM standard is based on the simple paradigm of TAO: *Topic*, *Association* and *Occurrence* [11; 12]. Therefore, two layers can be identified into the TMs paradigm: (i) the *knowledge layer*: representing topics and their relationships, allowing to construct the ECM model; (ii) the *information layer*: describing information resources, to be attached to the ECM topics.

Each topic can be featured by any number of names and variants for each name, by any number of occurrences, and by its association role. All of these features are statements and they all have a scope, representing the context a statement is valid in. Then, using scopes it is possible to avoid ambiguity about topics; to provide different points of view on the same topic (for example, based on users' profile) and/or to modify each statement depending on users' language, etc. Therefore, to solve ambiguity issues, each subject, which is represented by a topic, is associated a subject identifier. This unambiguous identification of subjects is also used in TMs to merge topics that, through these identifiers, are known to have the same subject; in practice, two topics with the same subject are replaced by a new topic that has the union of the characteristics of the two originals.

III. THE ECM VISUALIZATION

As stated in the previous section, Topic Maps are a very powerful means of representation, but they may be difficult to use. Intuitive visual user interface may significantly reduce the cognitive load of teachers. In the recent years, different relevant initiatives have been carried out to formalize the graphical notation of topic maps, such as GTM (*Graphical Topic Maps notation*) and TMMN (*Topic Map Martian Notation*). The former, specified in ISO 13250-7 (the part 7 of the Topic Maps ISO Standard), is a graphical notation defining ontologies and representing TM instance data. The latter is an alternative simple graphical notation, currently under development, able to express the Topic Maps Data Model that can be adopted with the aim of representing both ontologies and instance data on whiteboard, a paper, as well as within a diagram-based software.

With reference to the ECM visualization requirements, the main goal is to help teachers to quickly navigate an ECM, to group the topics in a *Unit of Learning* and to compose the course structure in an easy and intuitive way. Thus, there are two types of requirements for TM visualization: topics representation and navigation. The first helps teachers to identify points of interest, while the latter allows accessing information rapidly.

First of all let us identify the different kinds of views that a teacher may need. These views reflect the different approaches and steps to the activity of course planning and development. During the *planning*, when the teacher is defining the macro-aspect of a course, he/she must be able to navigate the Course Units structure, create new units, and change the relevant associations. In the *design* of a unit of learning, a teacher needs to explore the domain of knowledge and to add topics to the *UoL*.

Therefore, two different layers of Topic Maps can be considered: the layer of Course Units and the layer of Topics. In the first, the Course Units and the associations among them are represented. In the second one, there are all the topics belonging to the domain. The two layers are strictly connected through the relationship *has-{primary,secondary}-topic* as described before. Two layers offer a better view of the problem and a better way of information retrieval; merging course units and topics in one only topic map increase the complexity and reduce the re-usability of the map in different contexts. Moreover a teacher needs an overall view of the two layers, so to understand them globally. There would be two different overviews (course and topic) that should reflect the main properties of the structure. However, the teacher should be able to focus on any single part of the TM.

In order to support the knowledge structure navigation, the use of XTM makes possible to explore the ECM. Several TM engines provide suited visualization tools. Most of them display lists or indexes, from which it is possible to select a topic and to see related information. This representation is very convenient when users' needs are clearly identified. The navigation is usually the same as on Web sites, i.e., users click on a link to open a new topic or association. An example of such visualization is the one provided by Omnigator (part of the Ontopia Knowledge Suite), a Web-based Topic Maps browser (with RDF support) allowing both teachers and learners to display in a meaningful way the learning contents. It also allows following different learning paths, according to the specific educational purposes that have been defined. Finally, the same browser can be exploited with the aim of: i) merging different ECMs, on the fly; ii) searching learning content within the knowledge structure; iii) exporting this latter to various syntaxes; iv) personalizing different views; v) filtering subsets of topics according to different scopes.

As an alternative solution, in order to facilitate the visualization and the understanding of ECMs, it is also possible to integrate specifically designed plug-in, such as the Vizigate, offering an intuitive graphic visualization of the learning content. In Vizigate topics are nodes and their type may be symbolized by different colors, shapes and textures. However, the number of different shapes, colors and icons is limited. Class hierarchies can be used to reduce topic types to a small number of "*super-types*": *Course Unit, Unit of Learning* and *Topic* in the ECM; in this case, we only need a specific shape and/or color for each super-type. In the same way, the *is-related-to*, the *is-requirement-of*, and the other Educational Associations could be drawn differently, so that they can be quickly recognized by the teachers. The EAs are binary relationships, so they could be arcs and their type may be symbolized by the style of lines (e.g., full line, dotted line, etc.). Currently, the topic map standard does not suggest a standardized way of specifying type hierarchies. However, this could be done in Vizigator, an application based on the Vizigate, by using a kind of a style-sheet mechanism. This would allow to specify which association types represent the *supertype/subtype* relationship for any given topic map.

In Vizigator a mouse left-click event can display topics and associations names and types when the cursor is over these elements. Moreover, a right-click on a topic displays several additional information and permits the user to expand, collapse or navigate all the elements of that type. The heart of Ontopia is the Ontopia Topic Maps Engine on which all the other components are based. The basic functionalities of this engine are in representing, storing, and making available topic maps to the rest of the product suite. In addition, the engine also provides a rich set of APIs against which TM-based applications can be built. Moreover, Ontopia includes a Query Engine, built upon the Tolog query language. For retrieving information from the topic maps, the query language is far best suited than the API. It makes the task much easier, while at the same time, enables far better performance, since the query engine can take shortcuts not available to applications working through the API. As an example, let us consider the idea of combining the Vizigator API, that is a good tool for Topic Maps navigation, with the Ontopia Core Engine API, which offers several methods for the manipulation of TM and uses the Tolog language for machine reasoning and course sequencing. Then, an effective visualization and enquiry tool can be made available as a mash-up application.

IV. Conclusions

The ECM allows creating learning paths and instructional resources more easily interoperable and reusable, owing to the fact that the concept representation is independent of its implementation. In addition, the same concept network can be used for the design of different courses. Such an approach is suited to deal with the problem of delivering a course whose content needs to be personalized according to different learners needs. Moreover, an ECM drawn with reference to a specific subject matter can be used also to publish directly on the Web the underlying subject matter with links to prerequisites and related or suggested topics, or it can be used to generate linearized paths useful to produce a lesson or a printable document [13].

Graph linearization is another relevant issue to be considered, to take into account technological constraints regarding, for example, the compliance with standards requirements (such as the ADL SCORM), and in order to support the learning process tracking by means of a Learning Management System. Relationships used in the previously discussed model can be automatically processed and linearized, with the aim of extracting different learning paths (thanks to the information topics). These constraints, however, do not restrict the possibilities for teachers and instructional designers to identify the most effective strategies of contents presentation in different educational scenarios, because it is always allowed to modify the topics sequence in all the steps of the learning design process.

References

[1] Merrill, M. D. (2002). Knowledge objects and mental models. In D. A. Wiley (Ed.), The Instructional Use of Learning Objects (pp. 261-280). Washington DC: Agency for Instructional Technology & Association for Educational Communications and Technology.

[2] Ohlsson, S. (1987). Some principles of intelligent tutoring. In Lawler, R.W., and Masoud Yazdani, M. (eds.), Artificial Intelligence and Education: Learning Environments and Tutorial Systems v. 1, Intellect Books.

[3] Koper, R. (2001). Modelling units of study from a pedagogical perspective: the pedagogical metamodel behind EML. Heerlen: Open Universiteit Nederland. Retrieved from http://dspace.learningnetworks.org/handle/1820/36?mode=simple

[4] Merrill, M. D. (2002a). First Principles of Instruction. ETR&D, Vol. 50(3), 43-49.

[5] Gagne, R. M. and Briggs, L. J. (1979). Principles of Instructional Design. New York: Holt, Rinehart and Winston.

[6] Ausubel D. P. (1963). The psychology of meaningful verbal learning. New York, Grune and Stratton.

[7] Adorni, G., Brondo, D., Vivanet, G. (2009). A formal instructional model based on Concept Maps. Journal of E-Learning and Knowledge Society, Vol. 5(3), 33-43.

[8] Spiro, R. J., Coulson, R. L., Feltovitch, P. J., and Anderson, J. K. (1988). Cognitive flexibility theory: advanced knowledge acquisition in ill-structured domains. Patel, V. (ed.) Proceedings of the 10th Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Lawrence Erlbaum.

[9] ISO13250 (2003). ISO/IEC 13250:2003: Information Technology - Document Description and Processing Languages - Topic Maps. International Organization for Standardization, Geneva, CH. Retrieved from http://www.y12.doe.gov/sgml/sc34/document/0322_files/iso13250-2nd-ed-v2.pdf

[10] Pepper, S. and Moore, G. (2001). XML Topic Maps (XTM) 1.0. Retrieved from http://www.topicmaps.org/xtm/1.0

[11] Pepper, S. (1999). Navigating Haystacks, Discovering Needles. Markup Languages: Theory and Practice, Vol. 1(4), MIT Press.

[12] Ontopia (2002), Learn more about topic maps. Retrieved from http://www.ontopia.net/topicmaps/learn_more.html

[13] Adorni, G., Coccoli, M., Vercelli, G., Vivanet, G. (2008). Semantic authoring of learning paths with Topic Maps. In Proceedings of DMS 2008, the 14th International Conference on Distributed Multimedia Systems, Knowledge Systems Institute Graduate School, ISBN/ISSN: 1-891706-23-3, Boston, Massachusetts, U.S.A., Sept. 4-6, 2008.

Design and development of multimedia interactive systems for digital learning

Francesco Epifania
Dip. Elettronica Informazione
Politecnico di Milano
Dip, Scienze Informazione
Università degli Studi di Milano
Milan, Italy
epifania@elet.polimi.it Francesco.epifani@unimi.it

Abstract— For some time now in Italy, new technologies have attracted the interest of young people, they are the primary users of new communication devices such as smart-phones, tablets, iPad and iPod, and socialization environments such as social networks, to an extent that they use them as main tool for studying traditional subjects, rather than textbooks. On the other hand, students' abilities in Italian High-schools and Universities are seriously decaying in the ranking of OCSE and in other international statistics while national instructional budget ranks at top.

This document describes the development of e-learning 2.0 support for both the university and secondary school, experience in which the students of the Digital Communication Degree of the University of Milan have had an active role. A community effort with the involvement of teachers, allowing the creation and maintenance of websites, forums and WebTV portals, also accessible through mobile phones, resources for teaching new digital technologies while breaking the traditional "time & place" bound of the classroom.

Lessons and homework become available through video uploaded to our server in Research Laboratory. Results achieved will be measured by reduced insufficient marks at the end of this Academic Year.

Keywords: - E-learning 2.0 - Digital learning - Social networking technologies - video hosting

I. Introduction and Project Motivation

The first traces of our project date back a few years when we started an innovative portal called Webcen http://wcem07.dsi.unimi.it [1], with the aim to develop a fully functional framework to support the interaction between students and professors, as well as distance learning for students who cannot attend frontal lessons [2-3]. This innovative e-learning project released by us with whom we were able to address specific issues:

- each student has academic support in order to reflect their background knowledge and link this with the requirements of the degree class
- providing learning material for those students who cannot regularly attend lectures in classrooms
- providing as a result, higher ICT skills among students, while helping them to obtain better grades.

Subsequently, we developed the UnimiTube platform (http://video.dsi.unimi.it) (fig 1) as a result of a research and analysis of the approach of young people to the communication means (new media, mobile supports, advanced use of Internet). This product increased student involvement, using mobile support, generating new teaching and information channels and opportunities, and of course reached a larger target audience.

The UnimiTube portal, described in this paper, is the main example for the project which originated with the previous consideration. It is designed to reach an improved effectiveness through the use of communication equipment such as mobile phones, iPad, notebooks; technologies that young people are usually familiar with, in conjunction with a PC and IPTV, which allows viewing videos on a laptop screen whenever a user desires, through an internet connection.

The project is addressed to students, with mainly two scopes: the first is to involve them in an entertaining e-learning process with topics that are related to their lifestyle and preferences; the second is to get them closer to university environment through cooperation with university students, a similar project is [4].

It should be noted that high school students are already familiar with documenting events through video recordings with the use of a mobile phone, which are subsequently made public by uploading to video hosting portals such as YouTube for reaching a world-wide audience or contributing articles to magazines/journals that are promoted by their school under a professor's supervision, [5]. This capability, can enforce the e-learning process and the participation of students to current events and to the learning process, bringing closer the relationship between the school and the outside world.

II. Video Hosting Portal for University Students

The UnimiTube platform is a free access portal that allows audio/video file sharing, similar to YouTube but specifically targeted to University students.

The main features are: registration for users, Flash Player enhanced behavior (timer, percentage of video downloaded, full screen view and adjustable audio volume), enhanced video conversion in high quality, portability on mobile devices (tested on Nokia phones like N95, E66, C7) and an administrative backend where the administrator can manage server configuration, users and media.

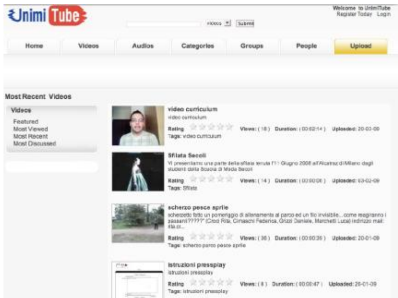


Fig.1

On the technical aspect, the project has been developed using the free Media Sharing CMS (Content Management System) PHPmotion representing the kernel of the website. The platform has been released with a Linux (Ubuntu) distribution for server "LAMP" (Linux – Apache – MySqll – Php) providing apache2 web server, PHP 5 (with CLI support), MySQL database server 5, LAME MP3 encoder, Libogg and Libvorbis, Mencoder and Mplayer, FFMpeg, GD Library, CGI-BIN, w32codecs (Fig 2).

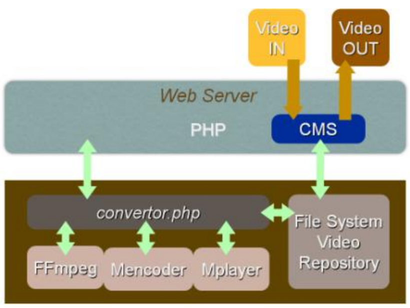


Fig 2

The website is divided into two main areas: the public section and the administrative section. In the public section the registered user can upload video, then the uploaded video is converted into Flash (flv file type) with different extensions (mpg, avi, divx, etc). The user just has to install the Flash Player plug-in in order to play all videos. He/she can edit the title, description and tags of the video, additionally he/she can delete the uploaded video or edit its properties: make it private or public, enable or disable comments and embedding for other websites. Users have also the option to group videos as Favourites, change the information of the user profile and link to other fellow users, thus creating a thematic social network.

As for the UnimiTube's Administrators, they watch, approve and delete every video, give them the "Featured" status, and control statistics concerning users, groups and comments. The Administrator has the faculty of banning a user that doesn't follow the guidelines about the ethic and privacy rules of the site.

UnimiTube stands out from YouTube with its custom template, exclusive audio section, own management for users/categories/groups/media and no advertising. The UnimiTube data are stored in a MySql database located on a server based on the CAD laboratory of research in the School of Computer Science of the University of Milano. The whole content is divided in sections and categories, containing audio and video. The interface has been carefully designed by the students in order to be user-friendly on every aspect. Maintenance and backup is provided following the standard MySql tools, and database exports is regularly performed to provide an efficient way to save all the content. The web site is accessible from a PC with every web browser and also from a mobile phone or device (Fig. 3). The user can customize the environment of the platform in order to single out the presentation of the information according to the type of request (multi-client approach), or to divide the interfaces in multiple versions (usually implemented within sub-domains), each of this versions communicate with a single client type (multi-site approach) (Fig.4).

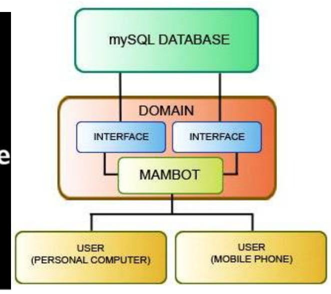


Fig 3 Fig 4

III. INTERACTIVE VIDEOJOURNAL

The Interactive VideoJournal project deals with video-based interaction as a tool designed to:

- improve and enrich the educational process, taking advantage of the wide availability of communication devices such as mobile phones, MP3, iPod and iPad;

- reduce "blackboard" dependency by employing technologies the students are familiar with, such as educational videos accessible with PC, tablets and mobile phones;

- emphasize the active role played by students, as they are capable to produce digital video-recordings. These video recordings can be available to the general public thanks to video hosting portals such as YouTube, or contributing material to magazines/journals promoted by their high schools or universities under a tutor's supervision.

The analysis of these new approaches were first described at their first stages in paper [6]. Subsequently we decided to explore other means to improve the efficiency of the learning practices and to reach a larger target audience.

We develop a new experiment called "Interactive VideoJournal", also known as VGG (Video Giornale Giovanile in Italian) described in [7–8], a free access portal that allows audio/video file sharing, a kind of "University students-targeted YouTube", in fact all video material contained in the VGG is hosted on the UnimiTube website. The VGG exploits all available communication possibilities on the World Wide Web which allow immediate one-to-one and many-to-many interaction.

Alongside the single video-still image selected to appear on the VGG homepage for each video article, there are also two buttons (Fig 5) that give additional possibilities:

- a "mail" icon, at bottom left, allows to send an e-mail to the author of that specific video article, thus allowing a personalized contact to the reader and further feedback
- a "printer" icon allows to print a one-page summary of the topic of the video article.

Fig.5

IV. THE EXPERIMENTATION WITH STUDENTS

We discovered certain advantages related to the launch of a collaborative learning process: on one side, the students of the course of HCI have put to practice what they learned in the classroom. On the other hand, the students have recognized a support for their own needs of learning (especially for university students coming from the same secondary institutions), putting to practice the qualities and competencies that belong to the university level [10].

A period of time has been dedicated to collecting information regarding the needs expressed by the teachers of the high schools. Former students of the course of HCI have been committed to such inquiries, they visited around fifteen high schools within the urban area of Milan and collected related data.

In a second phase, the groups of students will have assigned suitable spaces for the production of the VideoJournal, with the aim of supporting the didactic material coming from different schools. Each group will try to detect, based on the collected information related to the needs, the areas in which certain recovery and support courses should be activated for one or more classes. Didactic video material for these areas will be inserted, projected by university tutors in collaboration with school teachers. Further activities include the management of the spaces of interaction with the participation of diverse contributors for the project: the university students will undertake the duties of tutor and moderator, utilizing the instruments of collaborative learning already available.

In the current phase of the project video resources are in production by/for teachers and students of both universities and secondary schools, for the elaboration of contents and learning paths and for the use and sharing of materials as well [9]. The first phase of the experimentation – the cooperation between the University students of the course of "Fondamenti di Comunicazione Digitale" (Fundamentals of Digital Communication) and the very secondary schools they used to attend (through headmasters, teachers and students of the final year) – allowed the design of a map of matters and common contents for experiences and materials. As a result, tutors and students created video on basic mathematics, physics and information technologies, with the active cooperation of a few university teachers.

However, a quite obvious obstacle is the little motivation of students' and in this case we think that it could be useful to study an attractive system for the elaboration of content in secondary schools, for both the students and teachers/coordinators. There can be many advantages to take note, such as:

- solving the relationship issues between the university-level didactics and those of secondary schools (a better comprehension of their respective needs in terms of arguments, competencies, levels of knowledge, and the style of teaching contents);
- contributing to the improvement of the basic competencies in scientific matters;
- diffusion of technical competencies and knowledge, exploiting the availability of certain tools (UnimiTube – video hosting FIG1.) to produce and manage video by different users (University and Secondary School students and teachers, tutors, etc.);
- better mapping of matters and specific arguments shared at different levels by secondary schools and universities (such as Mathematics, Physics, Information Technologies, etc.);
- improvement of the diffusion of information about the University's didactic offerings, as a result of a better visibility of the specific contents of the University Degree courses, often not well-presented to the scholastic world by the usual channels, thus offering a better choice of academic studies.

V. SIMILAR PROJECTS AND DEVELOPMENTS ELSEWHERE IN THE WORLD

With the support of the CAD Lab, we explored the potentiality of the multimedia platforms extending the experiments also to certain fields away from the University. For example, the potential of Digital Video Interaction in the Tradeshow business is investigated in cooperation with UCIMU, the Italian Association of Mechanical Tool Machinery. The presentation of the International research in LAMBDA session at LAMIERA TradeShow, Bologna Fiere .

The VideoJournal of PIANETA GIOVANI presented at BI-MU, at MilanoFiera, produced in collaboration with the organizer Probest SpA. is available on the website http://www.polomeccanica.net/polo/index.php?option

=com_content&view=category&layout=blog&id=35&Itemid=46 (Fig. 6).

Fig. 6

Another important project is Megaupload, an online Hong Kong-based company established in 2005 which offer to the user the possibility of uploading and downloading files. It includes the Megavideo, MegaLive and MegaPix sites, specifically for video browsing. The basic service is free and allows users to upload files up to 2 GB of size. However, free users cannot download files larger than 1 GB. Free registered users are offered 200 GB of total file storage while Premium users have unlimited file storage. After a successful file uploading, the user is given a unique URL which allows others to download the file. Any file uploaded anonymously will expire if there are no downloads for at least 21 days. For free registered accounts, the file expiration period is 90 days. Premium accounts have no expiration date as long as the user remains a premium member. Non-registered users have to wait 45 seconds in the download queue and a certain amount of time between transfers after a certain number of megabytes had been downloaded. Users have to wait 25 seconds in the download queue. Premium members are able to hotlink [11] .

VI. CONCLUSION

In the current phase of the project we are creating a stock of audio and video educational resources for High School students and also teachers. As an alternate step to be applied in the Academic Year 2010/11, we have foreseen an eventual role for Digital Communication students as private teachers in the high schools they used to attend; with the implicit benefit of reusing VideoJournal-like tutorials and lectures as a source of additional pocket money for the students.

The next version of UnimiTube (now in development) will use the new tag audio/video in HTML5 standard. This new feature consist in playing audio and video files directly in HTML language in the source page without any other external plug-in, in other words it will not depend on the browser used. In fact we have multiple codec for audio and video files and as it's shown in the image below Ogg Theora and Ogg Vorbis are compatible with almost all browsers in their latest versions.

Tag Video	Chrome 10	Safari 5	Firefox 4	Opera 11	Ie 9
Mpeg-4	No	Si	No	No	No
H.264	Si	Si	No	No	Si
Ogg Theora	Si	Si	Si	Si	No
WebM	Si	No	Si	Si	No
Tag Audio	**Chrome 10**	**Safari 5**	**Firefox 4**	**Opera 11**	**Ie 9**
PCM	Si	Si	Si	Si	No
MP3	Si	Si	No	No	Si
AAC	Si	Si	No	No	Si
Ogg Vorbis	Si	Si	Si	Si	No
WebM	Si	No	Si	Si	No

Fig. 7

Meanwhile, Adobe is continuously developing new versions of the Flash player platform making it compatible with most of the latest devices (Apple iOS, Symbian and Windows Mobile) avoiding also battery decrease.

We decided that we need to maintain backward compatibily: the new site/portal of UnimiTube should automatically understand the technology that is in use so if the client has an old browser, the video will be displayed with Flash Player plug-in instead of the new browser with HTML5 video tag. The same process goes for mobile devices. Other features will be HD video and custom apps/templates for Blackberry, iOS,Symbian and Windows Mobile devices (Fig. 7).

REFERENCES

[1] Valle, G., & Folgieri, R. (2004). Continuous Learning Relaxing Time & Location Constraints in a Regular Curricula Degree Offered by University of Milan. World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004(1), 5166-5171.

[2] Loevinger L., Jurimetrics (1949) . The Next Step Forward, Minnesota.

[3] Downes S. 2002, La nuova economia dell'educazione: progettazione e riusabilità dei "learning objects" – Convegno "E-learning", Università degli Studi di Milano, , 11-12

[4] Nicoletta Di Blas, Paolo Paolini, Aldo Torrebruno: Innovative Technologies and Education - Two Successful Case-studies. CSEDU (2) 2010: 341-348

[5] Nicoletta Di Blas, Paolo Paolini, Amalia Sabiescu: Collective digital storytelling at school as a whole-class interaction. IDC 2010: 11-19

[6] Epifania, F. (2007). An experience in distance learning for university and secondary school. In G. Richards (Ed.), Proceedings of World Conference on E- Learning in Corporate, Government, Healthcare, and Higher Education 2007 (pp. 6786-6791). Chesapeake, VA: AACE.

[7] Valle, G. & Epifania, F. (2008). An Interactive VideoJournal to complement Learning & Training. In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008 (pp. 1164-1169). Chesapeake, VA: AACE.

[8] Jost, N. Nakanishi, Y. Tatsuta, L. & Gunji, K. Combining TV conferencing with BBS for Better Ways of Cross-culture Communication Proceedings Ed-Media 2004 Washington DC

[9] Muranishi, A., Shintani, K., Kono, A., Haga, H., & Kaneda, S. (2006). Proposed observational record and database scheme that can be applied developmental stages in children. Proceedings of 22nd Meeting of the Japan Society for Educational Technology. (pp. 287-88).Tokyo: JSET.

[10] Nagamori, M., Ando, M., Nagasawa, M., Songmuang, P., & Ueno, M. (2007). Collaborative e-Learning among Teachers Using a Web Database in Special Support Education. 7th IEEE International Conference on Advanced Learning Technologies Proceedings. (pp. 328-329). Los Alamitos: published by the IEEE Computer Society.

[11] http://en.wikipedia.org/wiki/Megaupload *Megaupload*

Educational games on a large multitouch screen

Carmelo Ardito, Paolo Buono, Maria Francesca Costabile, Rosa Lanzilotti
Dipartimento di Informatica
Università degli Studi di Bari "Aldo Moro"
via Orabona, 4, Bari, Italy
{ardito, buono, costabile, lanzilotti}@di.uniba.it

***Abstract*— Understanding how technology, coupled with skillful pedagogical solutions, can help to innovate and improve learning at school is the main goal of the "Learning for All" (L4A) research project. The combination of educational games and advanced technology has the potentiality of arousing pupils' attention, also engaging them in learning activities while having fun. This paper presents two educational games available through a large multitouch displays installed in the hall of a primary school. Such games aim at stimulating pupils to exercise their knowledge about history and geography. Field studies have been planned to study both educational and social aspects about the interaction with such games.**

Educational games; educational experience; multitouch displays; pupils

I. Introduction and Motivation

Capturing students' attention and engaging them in learning activities performed through electronic devices is very challenging. Gameplay is a promising solution [1, 2], especially if proposed through newfangled devices, which have the potentiality of arousing pupils' curiosity.

The research reported in this paper is part of "Learning for All" (L4A), a research project funded by the Italian Ministry. The main goal of L4A is to understand how technology, coupled with skillful pedagogical solutions, can help to innovate and improve learning at school. The project is carried out by seven partners, representing different institutions covering different fields of expertise: education and technology, mainly (further details can be found in [3]). The role of our research group in L4A is to design educational experiences based on advanced technology. In particular, we have proposed excursion-games on cell phones for supporting pupils visiting archaeological parks [4]. Field studies have demonstrated the learning effectiveness of excursion-games, which have been greatly appreciated by the teachers of the classes participating in the studies; they noticed that this game approach is able to engage pupils' interest in studying the historical site and that the notions learned seem to remain longer in their memory [2]. More recently, we are designing educational games on multitouch screens, which allow pupils to actively collaborate to solve the proposed challenges, manipulating by hand gestures the objects displayed. Thus, such games foster relational skills. Each pupil can carry out the activities s/he feels most congenial and, by working together on different parts, the whole group can solve the clues offered in the game and overcome difficulties thanks to their joint efforts. Thus, these games also foster the team spirit.

In [5], we have presented an early prototype of History-Puzzle, a game that asks participants to complete puzzles representing 3D models of historical monuments. History-Puzzle has been conceived to integrate the excursion-game experience of pupils visiting archaeological parks. Showing such 3D reconstructions it helps pupils to create a correct mental representation of the original environment, since in the park pupils are faced with ruins of ancient settlements that have lost their original image and whose current appearance no longer reflects their initial purpose. However, during the field studies, we have observed that pupils do not devote so much time to the 3D features, because they are engaged in rushing along the gameplay. Interacting with History-Puzzle represents a reflection phase, distinct from the true excursion-game, which allows pupils to reflect on their game experience.

An improved version of History-Puzzle is shown in this paper, together with EuroFlags, another educational game about the political geography of the European Union. We are going to investigate the impact of these games (another game is described in [5]) once used in the field, by planning some empirical studies in primary schools in Bari, Italy. We will study educational as well as social aspects of the interaction with such systems, by observing pupils' behavior and by interviewing teachers. The aim is to investigate if and how the presence of the multitouch screen influences their teaching and/or pupils' learning experience.

The paper has the following organization. Section II motivates the use of educational games. Section III describes two educational games on multitouch screen. Section IV illustrates field studies we are organizing and their goals. Section V concludes the paper.

II. Related Work

Game is amusing and fun. Enjoyment is important when endeavoring to achieve teaching goals, because what is enjoyably learned is less likely to be forgotten. A second important aspect of gameplay is that it requires different skills to be deployed simultaneously, and each player can practice those skills felt to be most congenial. Another important point is that gameplay is a relational activity: it encourages group activities, stimulates collaboration, helps with conflict management and is an excellent tool for individuating relational problems [6-9].

The motivating nature of gameplay pushed towards its use for educational purposes, rather than just for pure entertainment. The term *computer-based edutainment*, i.e. education in the form of entertainment, to let people to reach their learning goals by having fun [10], well expresses this concept. Empirical studies indicate the educational potential of interactive games (see for example [1]), primarily to teach mathematics [11-13], physics [14], logic [15], music [16], science [17], shape writing [18], art [19], and history [2, 20].

To the best of our knowledge, proposals of educational games on large multitouch screens are not reported in literature, while several learning applications have been proposed on tabletop displays. Tangible Interfaces for Collaborative Learning Environments (TICLE) strives to find new ways of helping children to learn math and science concepts [21]. Cmate is a tabletop collaborative concept mapping system, which aims at providing a means for a learner to externalize knowledge of a particular domain; Cmate fosters the development of strategies for organizing knowledge and facilitating communication of understanding [22]. The Augmented Knights Castle (AKC) offers children with autism the possibility to configure programmable elements allowing greater individual control and more socially oriented behavior [23].

III. EDUCATIONAL GAMES ON MULTITOUCH SCREEN

This section presents two games to be played by young students interacting with a multitouch screen to reinforce knowledge learned during class lessons. The scenario is that the multitouch system is in a hall of the school and students have free access to it during breaks (see Figure 1).

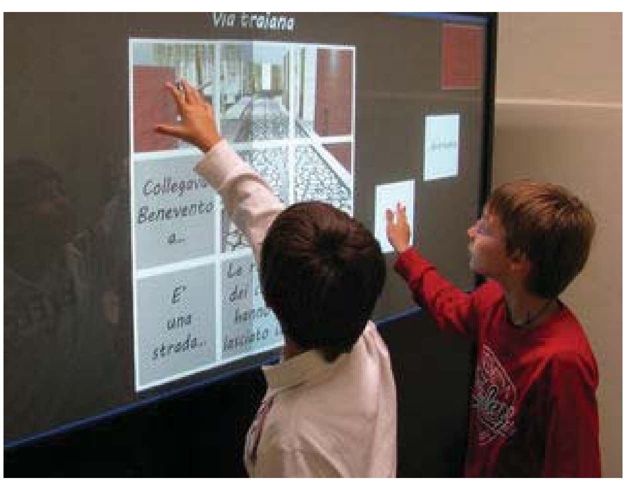


Figure 1. Two pupils interacting with History-Puzzle on the multitouch screen in our laboratory.

A. History-Puzzle

The name of this game comes from the fact that participants have to complete different types of puzzle. For example, the application proposes the map of an ancient city, referred to a specific age, and players have to complete it by placing the tiles representing the monuments/buildings that were present at that time. Another type of puzzle can be completed by putting together the two parts of a phrase reporting an historical notion related to an ancient site. For example, by referring to the game designed for the archaeological park of Egnathia, in Southern Italy, when a player touches one of the images representing a puzzle, such as the Via Traiana, a screen like the one shown in Figure 2 appears. The figure to be discovered by solving the puzzle is in the center of the screen, covered by nine incomplete messages about the selected place. The player chooses the rest of the sentence from the tiles displayed outside and drags it into one of the nine boxes in the central zone. If the selected association is correct, the box will reveal one ninth of the image of the 3D reconstruction of the original place. Figure 2 shows what it looks like when the player has discovered 5/9 of the image.

Figure 2. 5/9 of the "Via Traiana" puzzle have been completed.

In order to increment the difficulty of the game, many tiles are shown; these additional tiles report false answers or answers that do not match any of the nine incomplete sentences currently displayed. When the nine descriptions have been completed and the whole image is displayed, a 3D animated reconstruction of the place will appear. The system also reproduces contextual sounds, e.g. noises of the typical activities carried out in that place when the civilization of Egnathia was alive. In the example in Figure 2, showing the Via Traiana where carts run constantly, the noise of the wheels on the pavement is heard. Finally, the system returns to the map of the park to allow the participants to complete the puzzles of the other places.

For stimulating students to come back to the multitouch screen again, the position of incomplete messages and completing tiles is randomly reassigned every time the puzzles are run. Also the content displayed in such elements changes, thus proposing new challenges to the players.

B. EuroFlags

EuroFlags is primarily targeted at elementary and middle school students, who can engage in the game to test their knowledge on the capital, the geographical position and the flags of each of the European states. Players can choose among four game modalities:

- **Play with capital cities**: player has to put together the capital city proposed by the application and the corresponding country on the map.
- **Play with flags**: player has to put together the flag proposed by the application and the corresponding country on the map.
- **Play with countries**: player has to put together the country proposed by the application and its corresponding position on the map.
- **Mixed**: the application randomly switches among the previous modalities.

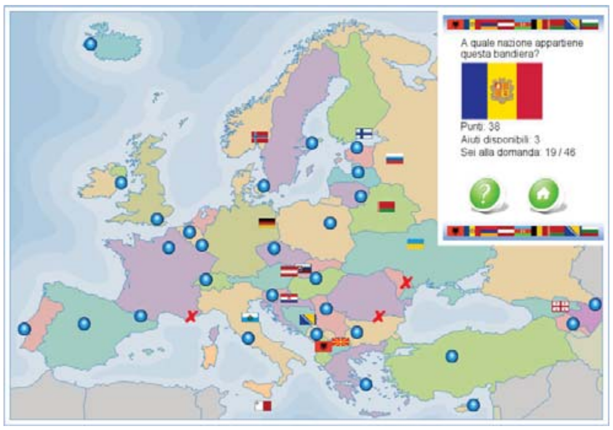


Figure 3. EuroFlags: "Play with flags" modality.

Let us suppose the player has chosen "Play with flags" modality. In the box on the right of Figure 3, EuroFlags is asking "Which country has this flag?", also showing the Andorra's flag. To answer the question, the player has to touch the blue point between Spain and France, which represents the Andorra country. S/he scores 3 points at the first attempt, 1 point at the second; after that, s/he fails and a red cross substitutes the blue point representing Andorra, while EuroFlags will not pose the question about Andorra again. Besides the question and the flag, in the box further information about the current match is displayed: the score (38 points), the hints available (3) and the stage (19 answers given *vs.* 46 questions available). If the player is stuck, s/he can ask for hints by touching on the green button showing a question point. A couple of examples of hints are: "It borders France and Spain"; "It is on the Pyrenees Mountains". The thumbnail flags on the map indicate correct answers the player has already given (e.g. Norway, Finland, Germany, etc.).

IV. The On-Going Evaluations

Preliminary studies on History-Puzzle and EuroFlags with pupils in our laboratory (formative evaluation [24]) have confirmed that they find this novel technology highly engaging. We are now planning more extensive evaluation studies to understand their impact once used in the field. A 47 inches LCD multitouch screen proposing the two games will be placed in the hall of a primary school in Bari, Italy. Through this public display, pupils can interact both with History-Puzzle and EuroFlags during the breaks. The field evaluations we are going to carry out will address both educational and social aspects.

A. Educational aspects

Both History-Puzzle and EuroFlags are educational games to be played by young students interacting with a multitouch screen. They are designed to reinforce knowledge learned during traditional class lessons; students could also interact with History-Puzzle after a possible visit to an archaeological park. The aim of the proposed games is to allow students to exercise the historical/geographical knowledge previously acquired, also discovering their gaps.

Understanding the "value" of an educational experience and analyzing how this value was achieved can be difficult, much beyond what it can be expected at first sight. This difficulty is especially relevant when some degree of innovativeness (in technology or pedagogy) is introduced, since unknown paradigms are at stake, and "what determines what" is less obvious [25]. There is a list of questions, apparently easy but very difficult to answer in practice: What did the pupils learn? Did they acquire specific knowledge or skills? Did they change their attitude and motivations? Did something not trivial begin? How was the technology used and when? How were the pupils instructed, how was technology made available?

Because we want to understand in detail what goes on in the class, questionnaires, with predetermined questions and predetermined answers, seemed too rigid and confined. Thus, interviews to teachers seemed the most suited method of research. The L4A educational partners set up a procedure aiming at making up a "dossier" of the educational experience: teachers are interviewed on expectations before the beginning of the experience; then, they carry out the educational experience and report, on a 'day-by-day' diary of observations, few notes, taken every day, about what went on and the most relevant facts and anecdotes. After the experience's completion, teachers are interviewed again for understanding what actually happened, how the experience was conducted and which were the outcomes in terms of learning, inclusion, etc. Relevant features are then "extracted" from each interview, by means of a schema, which helps in synthesizing all the interesting aspects of the educational experience.

B. Social aspects

Previous studies (see for example [26]) report that it is hard to entice people to interact with public displays; social embarrassment is the key factor which determines whether people will interact in front of an audience. Brignull and Rogers performed two field studies examining several variables: the flow of people around public displays; the level and types of interaction around displays; the transitions between types of interaction; factors that cause social awkwardness and embarrassment around public displays. They observed that interaction with public displays is accepted if the displays have strong physical and social affordances, which allow people to rapidly understand the purpose of the social activity to be performed at the display [26]. Our users are represented by pupils and we want to observe if they will be embarrassed as adults usually are, or if they will use the multitouch display as a stage for showing their skillfulness.

Other studies also investigated social aspect of adults interacting with multitouch public displays; for example, Jacucci et al. considered aspects such as the dynamics of approach (how people notice that there is an interactive display) [27]; the interaction at the display with others (parallel use, teamwork, and playful activities); conflict management; transitions between activities and participants (floor and turn-taking); roles and social configurations. It will be interesting to observe how pupils approach, participate, and interact with the multitouch display installed in the hall of the primary school in Bari, and to compare the results of our field studies with those reported in [27].

V. CONCLUSION

This paper has presented the work we are carrying out about educational games on large multitouch screens. History-Puzzle and EuroFlags have been described, which are two games about history and political geography, respectively. Formative evaluation has confirmed that pupils find games on multitouch screens highly engaging. In order to further investigate educational and social aspects related to the experience of interacting with the proposed system we are going to install a multitouch screen in a primary school in Bari.

ACKNOWLEDGMENT

Partial support for this research is provided by Italian MIUR through grants "L4A" and "CHAT".

REFERENCES

[1] F. Garzotto, "Investigating the educational effectiveness of multiplayer online games for children", *Proc. Int. Conference on Interaction design and children* (IDC'07), ACM, pp. 29-36.

[2] M.F. Costabile, et al., "Explore! possibilities and challenges of mobile learning", *Proc. 26th annual SIGCHI conference on Human factors in computing systems* (CHI'08), ACM, pp. 145-154.

[3] L4A, "Learning for All (L4A): a multi-paradigm, multi-channel and multi-tecnology approach to innovative pedagogy," http://www.learningforall.it/.

[4] C. Ardito, et al., "Experiencing the Past through the Senses: An M-Learning Game at Archaeological Parks", *IEEE Multimedia*, vol. 15, no. 4, 2008, pp. 76-81; DOI 10.1109/MMUL.2008.87.

[5] C. Ardito, M.F. Costabile and R. Lanzilotti, "Gameplay on a Multitouch Screen to Foster Learning about Historical Sites", *Proc. Advanced Visual Interfaces* (AVI'10), ACM, pp. 75-78.

[6] I. Blecic, A. Cecchini, P. Rizzi and G.A. Trunfio, "Playing with Automata. An Innovative Perspective for Gaming Simulation", *LNCS 2493*, Springer-Verlag, 2002, pp. 337-348.

[7] J.P. Gee, "What video games have to teach us about learning and literacy", *Comput. Entertain.*, vol. 1, no. 1, 2003, p. 20.

[8] M. Prensky, "Digital game-based learning", *Comput. Entertain.*, vol. 1, no. 1, 2003, p. 21.

[9] D. Shaffer, *How Computer Games Help Children Learn*, Palgrave Macmillan, 2006.

[10] Z. Pan, "E-learning and game", *Computers & Graphics*, vol. 30, 2006, pp. 1-2.

[11] C. Conati and X. Zhao, "Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game", *Proc. 9th International Conference on Intelligent User Interfaces* (IUI'04), ACM, pp. 6-13.

[12] F. Ke, "Classroom goal structures for educational math game application", *Proc. International conference on Learning sciences* (ICLS 2006), Intern. Society of the Learning Sciences, pp. 314-320.

[13] N. Shin, C. Norris and E. Soloway, "Effects of handheld games on students learning in mathematics", *Proc. 7th International Conference on Learning Sciences* (ICLS 2006), International Society of the Learning Sciences, pp. 702-708.

[14] K. Squire, M. Barnett, J.M. Grant and T. Higginbotham, "Electromagnetism supercharged!: learning physics with digital simulation games", *Proc. 6th Int. Conference on Learning Sciences* (ICLS 2004), Int. Society of the Learning Sciences, pp. 513-520.

[15] R. Lanzilotti and T. Roselli, "An Experimental Evaluation of Logiocando, an Intelligent Tutoring Hypermedia System", *Int. J. Artif. Intell. Ed.*, vol. 17, no. 1, 2007, pp. 41-56.

[16] G. Denis and P. Jouvelot, "Motivation-driven educational game design: applying best practices to music education", *Proc. International Conference on Advances in Computer Entertainment Technology* (ACE 2005), ACM, pp. 462-465.

[17] Y. Rogers, et al., "Ubi-learning integrates indoor and outdoor experiences", *Commun. ACM*, vol. 48, no. 1, 2005, pp. 55-59.

[18] P.O. Kristensson and S. Zhai, "Learning shape writing by game playing", *Proc. 25th annual SIGCHI conference on Human factors in computing systems* (CHI'07), ACM, pp. 1971-1976.

[19] J.S. Cabrera, et al., "Mystery in the museum: collaborative learning activities using handheld devices", *Proc. 7th International Conference on Human Computer interaction with Mobile Devices* (MobileHCI'05), ACM, pp. 315-318.

[20] K. Kardan, "Computer role-playing games as a vehicle for teaching history, culture, and language", *Proc. Sandbox - ACM SIGGRAPH symposium on Videogames* (Sandbox 2006), ACM, pp. 91-93.

[21] L.L. Scarlatos, "TICLE: using multimedia multimodal guidance to enhance learning", *Inf. Comput. Sci.*, vol. 140, no. 1, 2002, pp. 85-103.

[22] R.M. Maldonado, J. Kay and K. Yacef, "Collaborative concept mapping at the tabletop", *Proc. Interactive Tabletop displays* (ITS 2010), ACM, pp. 207-210.

[23] W. Farr, N. Yuill, E. Harris and S. Hinske, "In my own words: configuration of tangibles, object interaction and children with autism", *Proc. International Conference on Interaction Design and Children* (IDC'10), ACM, pp. 30-38.

[24] J. Preece, Y. Rogers and H. Sharp, *Interaction Design*, John Wiley & Sons, Inc., 2002.

[25] P. Paolini, et al., "Assessing and Sharing (technology-based) Educational Experiences", *Proc. World Conference on Educational Multimedia, Hypermedia & Telecommunications* (Ed-Media'11). In print.

[26] H. Brignull and Y. Rogers, "Enticing people to interact with large public displays in public spaces", *Proc. INTERACT 2003* (Enticing people to interact with large public displays in public spaces), IOS Press, IFIP pp. 17-24.

[27] G. Jacucci, et al., "Worlds of information", *Proc. 28th International conference on Human factors in computing systems* (CHI'10), pp. 2267-2267.

A Sketch-Based System for Teaching Geometry

Gennaro Costagliola⋆, Salvatore Cuomo⋄, Vittorio Fuccella⋆, Aniello Murano⋄
gencos@unisa.it, salcuomo@unina.it, vfuccella@unisa.it, murano@na.infn.it

⋆Università di Salerno
Via Ponte Don Melillo
84084 Fisciano (SA), Italy

⋄Università di Napoli Federico II
Via Cinthia, MSA
80126 Napoli, Italy

Abstract

Interactive Whiteboards (IWs) have been massively introduced in schools. While the benefits of these devices in the learning process are well known, dedicated software lack of functionality. In particular, most of the existing tools for IWs are an adaptation of classical software used on Personal Computers, mainly based on buttons and menus. The objective of the research presented in this paper is to completely re-think the interaction paradigm of the software for the IW, in such a way that the input is only composed of pen draws and hand-touches. In order to test this idea in practice, we apply it to the case of the teaching of geometry. Hence, more specifically, in this paper we propose a smart sketch-based tutoring system for IW that combines the recognition of hand-drawn geometrical shapes to the use of multi-touch gestures to support the editing of the drawn shapes.

1 Introduction

Interactive Whiteboards (IWs) are electronic devices having the dimension of a traditional whiteboard, on which it is possible to draw by using virtual pens or the touch screen system. The most advanced of them are enhanced with further capabilities, such as the possibility of supporting both interaction types and distinguishing among the media used for the interaction, making it possible to write with a pen, move objects with a finger and erase with the palm of the hand.

On one hand, a strong increase in the spread of IWs is expected in the future years. In 2006, the *Italian Ministry of Education* launched a plan for introducing 10'000 IWs in schools. With the present trends, a number of 5'300'000 IWs installed will be reached on a world-wide scale within 2015. In proportion, one IW out of 7 classes [8]. On the other hand, one of the drawbacks of these devices is the lack of software tools which can fully operate on them in order to support teaching. Most of the available software tools, in fact, are still built upon the traditional WIMP (Window, Icon, Menu, Pointing Device) interaction style [4]. Even the most recent of them cannot take advantage from all of the advanced capabilities cited above.

In this paper we present a work in progress aimed at enhancing the interaction with IWs. Our work is focused on the support of teaching geometry in middle school. The teaching of geometry requires a massive use of annotated diagrams. In particular, angles, circles, and closed polygons have to be drawn and they can be annotated with labels representing the identifier or the value of a segment or an angle. Furthermore, the syllabi for middle schools include the explanation on how to calculate areas and perimeters of polygons.

Traditional geometrical software tools, such as *CABRI* [1], allow to draw geometrical shapes and to edit them. Nevertheless, they use the traditional WIMP paradigm, which, as remarked in [6], requires a significant amount of mode switching and loss of fluidity within the interface. Besides the time loss in the use of the traditional paradigm, it also requires a good acquaintance of the teacher with its menus and tools in order to perform a live lesson. As reported in [5], in many fields, the use of a sketch-based interaction, besides being more intuitive, can significantly improve the efficiency with which the same tools are used.

The prototypical system we propose can effectively take advantage from the capabilities of the more recent IWs, included the use of pen-based and multi-touch interactions. In particular, it is an intelligent sketch-based tutoring system for IW that combines the recognition of hand-drawn geometrical shapes to the use of multi-touch gestures to support the editing of the drawn shapes. It supports the recognition of angles, circles and regular polygons. The recognized shapes can be annotated and the annotated elements can be referred to in simple hand written expressions. These expressions can be used to set the value of the elements and to calculate measures of interest such as the area and the perimeter of a drawn shape.

The rest of the paper is organized as follows: next section contains a brief literature survey on the most recent tools described in the scientific literature for aiding teaching through sketch-based techniques; Sections 3 and 4 are devoted to describe the system's features design, respectively; in Section 5, a prototypical implementation of the system is described; lastly, some final remarks and a brief discussion on future work conclude the paper.

2 Related Work

Some prototypical software tools based on sketch technology for enhancing the teaching of disciplines in schools have been recently introduced. This is the case of math [6], geometry [4] and physics [3].

One of the first attempt for aiding mathematical problem solving is the $MathPad^2$ tool [6]. The tool supports a set of operations, including the recognition and the solution of hand written mathematical expressions, the recognition and the animation of sketched diagrams, as well as the innovative capability of letting the user associate expressions to diagrams. After an association is made, changes in mathematical expressions can be reflected as changes in the diagram and vice versa.

In [4], authors present a system which recognizes hand-drawn figures and hand-written proof scripts. Figures and proofs are written by the teacher in two separate views of the system interface. As for $MathPad^2$, one of the most innovative capability is that of accurately establishing the correspondence between geometric components and proof steps: parts of the figures are highlighted as soon as they are used in the proof.

Similar to $MathPad^2$, but enhanced with a custom physics engine, is the work described in [3]. The tool provides animation of student-drawn sketches based on the associated mathematics. Such an animation allows students to have a better understanding of physics concepts and can also help students in intuitive verification of their answers to problems.

The surveyed tools have scopes different from the one which has inspired our work. Among those cited, the most related to ours is the work done by Jiang et al. [4]. However, this tool is more oriented to theorem proving and does not fit the needs of middle school programs. Furthermore, it is not specifically designed to support IWs and does not take advantage from multi-touch and multi-modal interaction.

3 System Features

The tool we propose is an intelligent sketch-based tutoring system for IW that combines the recognition of hand-drawn geometrical shapes to the use of multi-touch gestures, in order to support the editing of drawn shapes. The recognition of annotations and short expressions is also supported.

The features of the system have been calibrated upon the middle school syllabus of Italian School. The syllabus published by the *Italian Ministry of Education* [2] foresees the use of audiovisual equipment to support the teaching of science. In particular, for the teaching of geometry, its use is recommended, together with the recourse to drawings and diagrams. The use of experimental activities is foreseen in geometry as well as in other disciplines. According to the syllabus, the study of geometry must profit from a non-static presentation of the geometric figures, emphasizing their properties by showing their transformation.

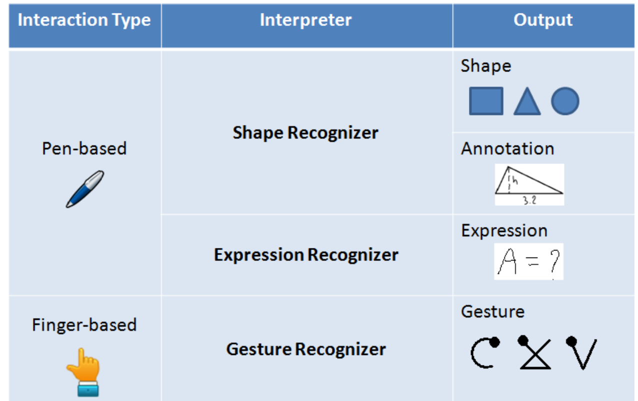


Figure 1. System architecture.

More in detail, the syllabus for geometry includes some basic concepts, such as the study of plane and solid figures, the calculation of areas, perimeters, length of segments and magnitude of angles, the Pythagorean theorem. The processes of building figures is also included and foresee the use of tools such as ruler, set square, and compasses. More advanced topics include the Cartesian coordinate system and the Euclidean transformations.

The proposed system supports a set of features including the recognition and manipulation of angles and plane geometrical figures. The support of solid shapes and Cartesian coordinate systems is foreseen for future enhancements.

4 System Design

The system is composed of three different modules, as graphically summarized in Figure 1: (i) a *Shape Recognizer* which recognizes the sketched geometrical shape and annotations performed on it; (ii) an *Expression Recognizer*, which recognizes simple expressions referred to elements of the drawn shapes; (iii) a *Gesture Recognizer* which recognizes gestures performed on the shapes in order to edit them. The input is dispatched to the right recognizer, according to the type of interaction (pen-based or finger-based). Expressions can be distinguished from shapes and annotations, since they are performed on separated views.

4.1 Shape Recognizer

The *Shape Recognizer* supports the recognition of angles, circles and regular polygons. In our initial prototype, a shape must be completed in a single pen stroke. As soon as a stroke is entered, a *cusp detection* procedure is run, in order to obtain the count and the positions of cusps in the stroke. The stroke is resampled in a sequence of equally spaced points. A stroke can be firstly classified as a *segment*, an *angle* or a *shape* on the basis of the number of cusps it contains (0, 1, > 1, respectively). A further check for *closure* is applied on strokes classified as *shapes*. This is performed by checking that the distance between the last and first points of the stroke is smaller than a given threshold. If this check fails, the stroke is left unrecognized and it is automatically erased from the screen. A stroke classified as a *shape* is further analyzed in order to correctly classify it as a *circle* or a *polygon*. This step is performed by evaluating a measure of circularity as reported in [3]: the standard deviation of the angle subtended at the stroke's centroid by each line segment in the stroke is computed and for values smaller than a threshold, the stroke is classified as a *circle*, otherwise it is classified as a *polygon*. A *polygon* is further classified in a *triangle* (*equilateral*, *isosceles* or *scalene*), a four-sided shape (whose type can be *square*, *rectangle*, *diamond* or *rhomboid*), or an n-sided regular polygon (*pentagon*, *hexagon* and *octagon*).

As soon as a stroke is classified, it is immediately beautified. In past research on sketch recognition, beautification has not always been regarded as a desirable feature. In particular, Plimmer and Apperley [9] found that in the early stages of a design process, the users appreciated that the look and feel of a rough sketch was maintained. Nevertheless, the above findings were recorded on the recognition of user interface diagrams. These diagrams can contain many elements and the precision with which single elements are drawn is not essential for their understanding. In our case, instead, a single shape is entered and a higher precision is required, so we regarded beautification as desirable. The procedures for recognition and beautification of the entered strokes are based on the algorithms described in [7].

After they have been recognized, the shapes can be annotated. An annotation can be a *number*, a *letter*, or a *dashed line*. The annotation is assigned to the annotated elements by proximity: it must be drawn within a given distance threshold and is assigned to the closest element to which it can be referred. Given the sets of points of both a graphical element and the annotation, the distance is calculated as the minimum of the distances between any two of their respective points. A *number* can be used to set the length of the side of a polygon, the circumference of a circle, or the magnitude of an angle. A special convention is used for *letters*: capital letters must be associated to vertexes and lower-case letters to sides. A *dashed line* can be used to highlight parts of interest of a shape: the bisector of an angle, the height of a triangle, the radius or the diameter of a circle, the apothem of a regular polygon and so on.

4.2 Gesture Recognizer

The entered shapes can be edited through the use of multi-touch gestures. The semantics of the gesture set has been thought to enable the most desirable editing functionalities for each shape. First of all, it is possible to enlarge and shrink the size of a drawn shape. This is performed by *pinching inwards* or *outwards* of the shape.

If an angle has been drawn, its magnitude can be changed through a two-finger *rotate* gesture: a finger must be pointed on the vertex and the other on one of its rays. With the first finger standing on the vertex, the second finger is rotated and the underneath ray is moved jointly with it. Another set of gestures is available for shapes. By keeping the finger pressed on an angle or on a side of a shape and then *swiping* the finger, it is possible to move the element and change the aspect ratio of the shape: with this gesture it is possible to change the type of the entered shape. For instance, a *rectangle* can be changed to a *square* by pressing on one of its short sides and then swiping it to narrow the long sides. A *scalene triangle* can be changed to *isosceles* by moving one of its vertexes.

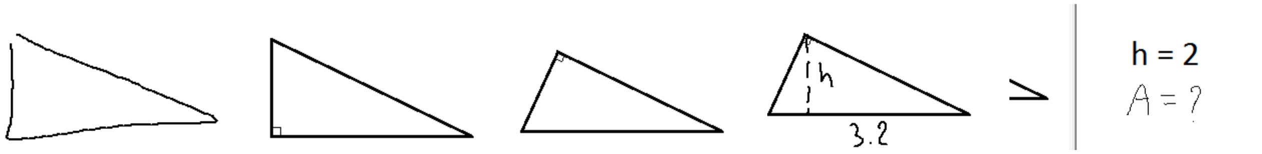
When editing the shape, visual feedbacks are given to the user to point out that the shape is going to change its type. When an editing operation entails a change in the magnitude of an angle, a small circular arc centered at the vertex of the angle is shown. As soon as its magnitude is close to 90 degrees, a small square is shown in place of it. This is useful to let the user easily set right angles. The contour of a shape is also colored in green as soon as a change in shape type is detected. Numeric labels, with which sides and angles are annotated, are also changed when the shapes are modified. More precisely, their values change proportionally to the extent of the modification. The precision with which the new value is reported is the same of the original value. This holds for the number of decimal digits used to report the length of the side of a shape and the fractions of a degree (e.g., minutes and seconds) used to report the magnitude of an angle.

4.3 Expression Recognizer

The *Expression Recognizer* recognizes simple handwritten scripts. The scripts can be composed of different *steps*. In each step, geometrical elements of the drawn shapes can be referred with the letters used to annotate them. An *assignment* statement (with the = operator) can be used to set angle magnitudes and side lengths. Angles can be referred to using both the classical notation (three letters identifying its rays) and the simplified notation using only its vertex. It is worth noting that an assignment can cause a change in the shape proportions or type. In this case, a refresh of the view is triggered.

When the '?' special character is used on right-hand side of the assignment statements, it is automatically substituted by the system, provided that its value can be calculated using the data annotated on the shape or entered in the previous *steps*. P and A (capital) characters are used to refer to the perimeter and the area of the currently drawn shape, respectively. As soon as an expression is evaluated and recognized as valid, it is converted to printed characters.

The *Expression Recognizer* is a simplification of the one presented in [4]. Firstly, a partition of the hand writing is performed in order to distinguish different symbols in each expression. Then, the symbol are recognized. The set of symbols only includes number, letters and the '=' and '?' special characters. The semantic interpretation is easier than that used in [4], due to the smaller number of avail-



(a) A stroke is entered (b) The shape is beautified (c) The shape is modified through a finger-based interaction (d) The shape is annotated (e) Expressions are used to calculate area

Figure 2. Some screenshots of the system interface

able symbols and operations. In our design, an expression is evaluated as soon as a timer expires from the time of the last input event. If the expression is syntactically correct, the step is executed and a beautified version of the expression is shown in place of the sketched one, otherwise, the system waits for further input.

5 Implementation

A prototype implementing most of the features of the above described system has been developed in Java language. The interface of the application is vertically divided into two parts. On the left side, figures can be drawn and annotated. The right side is for expressions.

Figure 2 shows some screenshots of the system interface as soon as an action is performed: in Figure 2(a), a stroke representing a right-angled triangle has been drawn; the system substitutes the sketched version with a beautified one (the lower-left vertex is recognized as a right angle, as shown in Figure 2(b)); by pressing and then swiping the finger on the upper vertex the shape is changed and a new right-angled triangle is produced (Figure 2(c)); the triangle is then annotated with a number (3.2) to set the length of its base. Furthermore, the height from the upper vertex is traced out and annotated with the h character (Figure 2(d)). In the right view, the expression $h = 2$ is entered. As soon as the expression is recognized, the sketched version is beautified. The user can enter a new expression ($A =?$, as shown in Figure 2(e)), in order to calculate the area of the triangle. As soon as the last expression is entered, the system substitutes it with a beautified expression reporting the value of the area.

6 Conclusion

In this paper, we have introduced a new software tool for the IW. The underlying idea is the replacement of the classical WIMP paradigm in favor of a more natural interaction, based only on pen draws and hand-touches. We have applied this idea to the case of the teaching of geometry. The proposed system is a smart sketch-based tutoring system prototype for IW that combines the recognition of hand-drawn geometrical shapes to the use of multi-touch gestures to support the editing of the drawn shapes. In a first informal test with pilot users, we have noted several positive aspects. First, we report an appreciation of the new interaction paradigm, particularly in relation to the ease of use of the tool: it is not required to remember menu layouts to use specific functionalities. Second, we report an increased fluidity in teaching: indeed, the new interaction paradigm does not compel the teacher to introduce annoying pauses in the lesson, as with classical menu-based applications. Finally, due to the absence of menu and toolboxes, the whole space of the whiteboard is fully available for the lesson.

More formal experiments are necessary to fully evaluate the effectiveness and the efficiency of the system. In particular, as future work, we are planning to carry out an experiment in which specific tasks are executed. These tasks must simulate the execution of traditional and online short training sessions. As a result, we aim at comparing the efficiency of interacting with the IW with both the classical WIMP and the proposed paradigms. We will also evaluate the satisfaction of tutors and learners.

References

[1] Cabri ii plus. http://www.cabri.com/cabri-2-plus.html, 2011.

[2] I programmi della scuola media - d. m. 9 febbraio 1979 (in italian). http:// www.edscuola.it/ archivio/ norme/ programmi/ media.html, 2011.

[3] S. Cheema and J. J. LaViola, Jr. Applying mathematical sketching to sketch-based physics tutoring software. In *Proceedings of the 10th international conference on Smart graphics*, SG'10, pages 13–24, Berlin, Heidelberg, 2010. Springer-Verlag.

[4] Y. Jiang, F. Tian, H. Wang, X. Zhang, X. Wang, and G. Dai. Intelligent understanding of handwritten geometry theorem proving. In *Proceedings of the 15th international conference on Intelligent user interfaces*, IUI '10, pages 119–128, New York, NY, USA, 2010. ACM.

[5] L. B. Kara and T. F. Stahovich. An image-based, trainable symbol recognizer for hand-drawn sketches. *Comput. Graph.*, 29:501–517, August 2005.

[6] J. J. LaViola, Jr. and R. C. Zeleznik. Mathpad2: a system for the creation and exploration of mathematical sketches. *ACM Trans. Graph.*, 23:432–440, August 2004.

[7] 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.

[8] A. N. per lo Sviluppo dell'Autonomia Scolastica. La lavagna interattiva multimediale (in italian). http://www.indire.it/, 2011.

[9] B. Plimmer and M. Apperley. Interacting with sketched interface designs: an evaluation study. In *CHI '04 extended abstracts on Human factors in computing systems*, CHI EA '04, pages 1337–1340, New York, NY, USA, 2004. ACM.

Introducing ePortfolio to Architectural Course : The Integrated Archiving Environment

Yuki TERAWAKI[1], Taketo Shimohigoshi[2], Makoto Watanabe[2], Yasushi KODAMA[1], Kazuo YANA[1]
[1]Research Center for Computing and Multimedia Studies, Hosei University, Tokyo, Japan
[2]Hosei University Faculty of Engineering & Design, Department of Architecture, Tokyo, Japan

Abstract—**We present results from an architectural course to design an ePortfolio to improve learning environment of architectural course and to support the assessment of Japan Accreditation Board for Engineering Education (JABEE). This paper describes the Integrated Archiving Environment (IAE). IAE achieves ePortfolio for architectural course. IAE not only achieves ePortfolio but also provides a tool of digitalization for the learning for which it was not originally digitalized, and the whole results of learning are stored in IAE server. IAE can enhance the learning achieves without making any changes to them.**

Keywords— **requirements engineering, reflective learning, ePortfolio, QR code, Eye-Fi Card**

I. INTRODUCTION

Department of architecture Hosei University plans to receive an assessment for Japan Accreditation Board for Engineering Education (JABEE). JABEE is assurance that a college or university engineering education programs meet the quality standards established by the profession for which it prepares its students. The evaluation team of JABEE reviews evidence such as course materials or outcomes such as student projects. In the accreditation of JABEE, the result of study (outcomes) and education (evidence) are evaluated. All of the outcomes and the evidences need to be archived. They need to be searched responding to the diverse education history of each student. Therefore, we plan to introduce a means of digitalization for all of outcomes and evidences. They are archived and searched according to educational practice of department of architecture. We developed web-based application that we call 'integrated archiving environment'. We consider the IAE a basis for ePortfolio.

ePortfolios are electronic versions of portfolios. Portfolios are normally used to showcase and assess a person's achievements and learning goals [1][3]. Portfolios have been used in arts and architectural course for years. However, when all outcomes that are created on an architectural course would plan to be digitalized, the digitalization process may encounter many difficulties.

- Means of Digitalization
  The students of department of architecture produce a lot of architectural works. Most of them are not originally digitalize. For example, there are architectural model, handwritten and drawn sheets and presentation board. These architectural works are hard to digitalize. The paper size of handwritten and drawn sheet is more than A3. And architectural model is not flat. Therefore, in order to digitalize, some special instruments are required.

- Promptness of Digitalization
  Moreover, promptness of digitalization should be considered. For instance the laboratory teaching situation and the works of student are recorded by using camera during the class. Usually, the shot picture save into the media of camera such as memory-card according to specific rule of file naming. The shot picture and its students are specified and then the file name generated by the camera is changed accordingly. If the shot picture and its student were specified immediately, the shot pictures will become available as material of discussion in classrooms additionally if would make academic staff's lives easier.

- Storage for Large Files
  In HOSEI University, Course Management System (CMS) has been used since 2007 by users campus-wide. The CMS has already been available to distribute the handout or to submit the report. Part of outcomes and evidences for department of architecture has been already stored in CMS. However, all outcomes and evidences cannot archive in the CMS because limit of file capacity becomes a problem. Generally, the report file for classroom lecture requires 1 to 2 megabytes. However the report file for laboratory teaching requires anywhere 30 to 100 megabytes per student in department of architecture. Therefore, CMS that are available throughout the Hosei University cannot save the whole outcomes for department of architecture.

Therefore, toward the realization of so-called ePortfolio, we have to focus on the means of digitalization at the first setout. Moreover, current learning activities are never modified. The underpinning initiative of this research is to develop a means of computerization and to manage the computerized outcomes and evidences.

This paper is organized as follows. First we will list the requirements of IAE. In the following section, we describe IAE. Finally, we explain the future work and the conclusion.

II. Requirements for IAE

Many software problems arise from shortcoming in the ways that people gather, agree on and modify the product's requirements. In fact errors made during the requirements stage account for 40 to 60 percent of all defects found in a software project [8]. So, we had a great deal of time discussing for requirements of IAE.

The Requirements Engineering (RE) Process has 4 processes. There is requirements elicitation, evaluation, specification (documentation) and quality assurance. This process is iteration on successive increments according to a spiral model [6][7]. We obtained the requirements of IAE through a few requirements process.

IAE should meet the following requirements. These requirements will offer numerous result suggestions to other area such as art courses.

A. Top Goal

We are aiming to enrich the educational environment of the Department of Architecture by enabling both teachers and students to share fully the learning experiences. To this end, all the educational outcomes and evidence of the department must be stored and maintained properly. That material includes the educational guidance provided by the teachers and the work produced by the students such as lecture reports, and handwritten design drawings and architectural models created in laboratory teaching classes. Some of them are inherently digital, and the others need to be converted to digitalized.

Therefore, we have to consider educational practice of individual classes and then we prepare the appropriate data-saving environment accordingly. Moreover, we will provide the means of digitalization for non-electronic material.

B. Data Capacity

IAE data should be maintained at least for four years, these have to be browsed and retrieved at the appropriate rate.

The students' learning history is used for reflection [2]by them. Moreover teachers can refer to the previous works of students for information about level of attainment. We have to calculate precisely what capacity the IAE will need. For instance, in the laboratory teaching classes of our department of architecture, students submit works over three times per semester. If students compress a file of their work, on average, the volume of data of each student's work is 100 MB. Department of architecture has over 150 people per in every grade.

C. Functionality

In laboratory teaching classes, photographs of learning activities and students' work are taken by teaching assistant. The creation process of each student are recorded and can be used for future guidance.

Usually, the shot picture save into the media of camera such as memory-card according to specific rule of file naming. Therefore, after class, teaching assistants have to change the file names of those photographs from automatically assigned ones to student numbers to identify who the photographs belong to. Because that operation depends entirely on the assistants' memory, it is subject to human error.

Therefore, we need to prepare preventive measures against such mistakes. If real-time photograph identification becomes possible and image files linked to each student can be stored directly onto the IAE server, it will not only prevent errors but also expand the possibilities of taught in a laboratory. For example, teachers will be able to hold discussions or give guidance, retrieving and presenting the image files of a student to other students in the class.

C. Interoperability

It is important to consider that interoperability between information system provided by Hosei University and IAE.

1) The Authentication Check

In the light of password management and user-friendliness, the authentication check of IAE needs to work together with the authentication system constructed by our university.

2) Course Registration to IAE

Class information, such as class names or teachers in charge, is managed by the academic affairs system. Therefore, the data required to set up the IAE has to be imported from academic affairs system.

3) Retrieve data of CMS

In some didactic manner classes, the receiving of report is performed via CMS provided by the university. We have to consider how the data of CMS are imported and subsequently how imported data are saved in IAE server.

III. IAE

Figure 1 shows conceptual image of IAE [5]. IAE provides function to store the outcomes and evidences, to accept the outcomes and evidences, and to digitalize the handwritten materials. We'd like to say, IAE providing the means of digitalization and data management without changing educational practice of department of architecture.

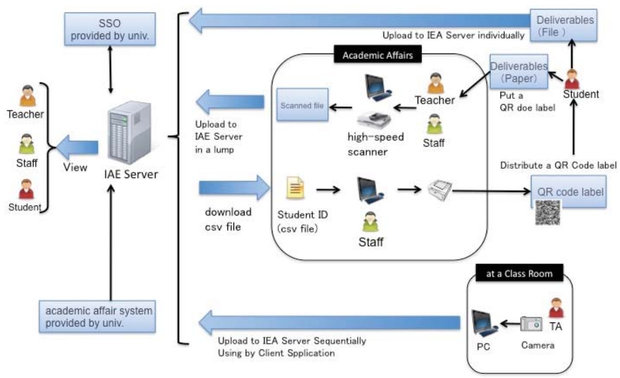


Fig. 1. The conceptual diagram of IAE

IAE consists of three tools: First is management of outcomes and evidences (store and accept). Next is the means of digitalization using by Quick Response (QR) code and thirdly, there is the sequential uploading system for the management of photograph.

A. Management of outcomes and evidences

Year is basic to manage for outcomes and evidences. Web application of IAE also is realized on a year basis. Figure 2 shows sitemap of IAE. Each site is arranged in a hierarchy basis on year. Moreover, user is able to operate IAE in the same operational feeling as PC such as Finder of Mac or Explorer of Windows.

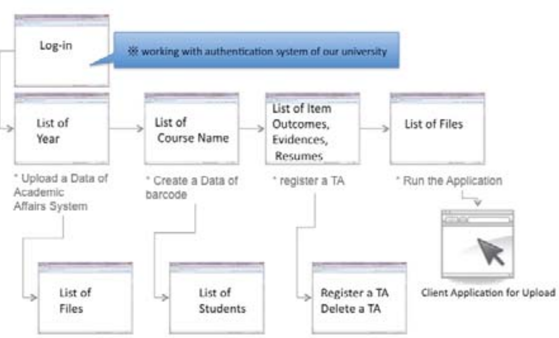


Fig. 2. Sitemap of IAE

B. Means of Digitalization using by QR Code

In means of digitalization, barcode is used for identification of student ID and his or her hand drawn report.

The type of barcode is found in several types, such as code39, JAN-13/ EAN13 and QR code. In the light of error correcting level for code reading, we decided to employ the QR code.

QR code is a specific matrix barcode. The code consists of black modules arranged in a square pattern on a background. The information encoded can be text, URL or other data. QR codes are used in a much broader context, including marketing campaigns, ticketing and email marketing and couponing. Specification of QR code is standardized as ISO/IEC 18004 (JIS X 0510) [4].

Figure 3 shows QR code label of IAE. In IAE, student IDs are encoded as QR code label, and then the QR code label are distributed to student. IAE provides the means of digitalization by QR code label is put on every hand drawn report. Initially, QR code label is created by using provided data from the academic affairs data. Staff downloads the student ID data from IAE server to staff's computer, and then QR coded students ID are printed by application for generating QR code. Ten QR code labels per student are distributed at department guidance of the beginning of a semester. If student finished up the QR code label, staff can always print out it because QR code label can be easily printed. Next, The report that put the QR code seal is read with a high-speed scanner, and is made a digitalized file that the student number is identified. These digitalized files are uploaded all at once. Finally, these files are automatically saved on the appropriate folder of class name to student ID basis.

Fig. 3. QR code label of IAE

B. The sequential uploading system for the management of photograph

We've combined various technologies in order to make it possible to promptly upload photograph after shooting works of students. The procedure is described followings. First, photographs accumulated in the camera are automatically transferred to a dedicated PC. Then, the accumulated photographs are uploaded to IAE server by using software which works to identify photographs and student numbers and to upload them to IAE server.

The efforts performed by the camera and PC are as follows. We use the camera that accept Eye-Fi memory card. Eye-Fi memory card is a kind of SD memory card with built-in wireless LAN client function. Therefore, it is possible to transfer recorded photographs to PC or server without ejecting memory cards from digital camera or connecting digital camera to PC. In order to smoothly move photos from camera to PC, we set up the extremely small wireless adopter to PC. So, PC can operate as an access point.

To up-load to the IAE server, we developed software. Figure 4 shows conceptual image of sequential uploading system for the management of photograph and figure 5 shows screen shot of this software. This software will support to specify a photographed picture and student number easily and then execute to upload the picture to IAE server. Moreover, we take a careful note of the maintenance of software. To be more precise, if there is a change for software, the program update is executed without reinstalling the software to individual PCs.

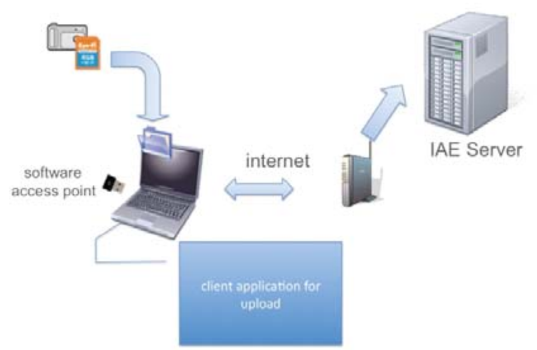


Fig. 4. Conceptual imag of the sequential uploading system for the management of photograph

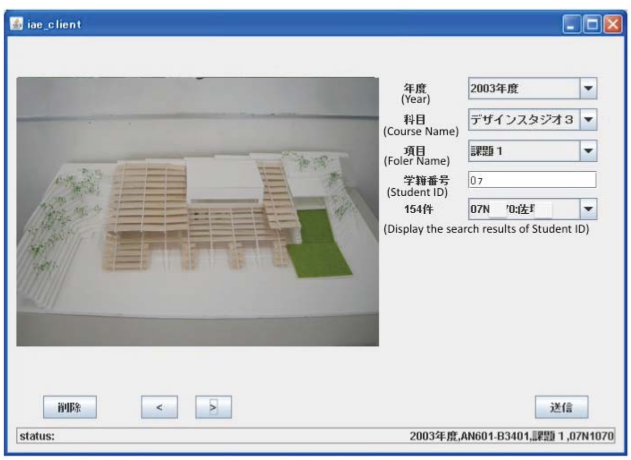


Fig. 5. Screen shot of client application for upload

IV. CONCLUSION

In this work, it aimed for sharing the environment of "Learning" between teachers and students and enriching the educational environment of the architectural course, and the integrated preservation environment for outcomes and evidences was constructed. We elicited the requirements for IAE, and then specified and completed it as a system.

IAE started the operation from April 2011. Because it has been only for several weeks since its utilization, the evaluation with questionnaire and system log analysis would become the future issue. However, the evaluation seems to be generally favorable for the several courses that have already started the utilization. We would like to proceed the evaluation while accumulating the usage results from now on.

In e-portfolio, it will be desired to coordinate the all processes such as how an active learning is executed in class, what kind of change is generated during the learning process, and what kind of learning outcome is accomplished. Learning results by printing media and electronic media are obviously mixed in college unless originally it is required to create the learning results electronically like a coursework in college for correspondence courses or Internet college. Like the examples in this article, it will be difficult to computerize or continuously preserve all learning results for the courses that put more importance on creations such as architecture and art course. IAE accomplished a great contribution by providing the support and eventually conquering the difficulty.

E-portfolio started getting attention even in Japan from around the last year. In Japan, e-portfolio has been utilized for two purposes. Firstly it aims for personal progress centered on self-evaluation and secondly it aims for how a learning procedure can be facilitated by the learning process in class. Our approach is to support the latter. For approaching with e-portfolio as how the learning process can facilitate a learning procedure, it will be indispensable to approach for computerizing the learning results while constructing the e-portfolio system at the same time.

REFERENCES

[1] E.A. Northwest, "Portfolios", Portfolio News, vol.2, 1991.

[2] Donald A. Schon, "The Reflective Practitionaer: How Professionals Think in Action", Basic Books, Inc. 1983.

[3] P. Bulter, "A Review of the Literature on Portflios and Electronic Portfolios", Massey University, Palmerston North 2006.

[4] ISO/IEC, "ISO/IEC 18004:2006 Information technology -- Automatic identification and data capture techniques -- QR Code 2005 bar code symbology specification", 2006.

[5] IAE 2011 Server, Retrieved April 8, 2011, from https://iae.hosei.ac.jp, 2010

[6] Boehm, B. W., "A Spiral Model of Software Development and Enhancement", IEEE Computer, Vol.21, No. 5, 61-72, 1988.

[7] Kotonya, G. and Sommerville, I., "Requirements Engineering: Processes and Techniques", John Wiley & Sons Ltd, 1997.

[8] Davis, A. M., "Software Requirements: Objects, Functions, and States" Englewood Cliffs, NJ, Prentice Hall PTR, 1993.

Address of the corresponding author:

Yuki TERAWAKI
Research Center for Computing and Multimedia Studies,
Hosei University
3-7-2 Kajino, Koganei City Tokyo 184-8584
Email: yuki.terawaki.dc@k.hosei.ac.jp

Proceedings
International Workshop on Visual Languages & Computing (VLC 2011)

Editor

Levent Burak Kara, *Carnegie Mellon University, USA*

Drawing on the World: Sketch in Context

Andrew Correa
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA, USA
acorrea@csail.mit.edu

***Abstract*—In this paper we introduce our approach to implementing context-rich sketch-based interfaces. By "context-rich" we mean interfaces for systems that refer to real-world objects. For example, a system that allowed the user to draw on an annotated video feed instead of a blank canvas would yield this kind of context-richness. We describe "Drawing on the World", the concept that taking what is being drawn on into consideration results in increased ease of development and a better user experience in sketch recognition.**

We created a description language, called StepStool, and an engine to interpret StepStool descriptions. StepStool is used to describe the relationship between context (objects on the canvas) and drawn shapes to determine the shapes' meanings. We used StepStool to implement the context-rich control interface to a robotic forklift (see [2], [11]). We use that interface in this paper to describe StepStool's use.

In our future work, we propose two extensions to StepStool that could make it more broadly applicable. The first extension allows StepStool to be used with non-robotic systems. The second proposed extension allows StepStool to be used with other modalities—e.g., hand gesture recognition.

***Index Terms*—Gesture Computing, Human-Machine Interface Design, Sketch and Gesture Based Design, Sketch Recognition**

I. Introduction

Sketch is everywhere. From an early age we are taught to communicate ideas to one another using pen and paper and to interpret drawn diagrams. In the Multimodal Understanding Group we are interested in simplifying the way people interact with systems, by giving those systems the ability to understand human modes of communication, i.e., speech, sketch, and hand gesture. The work presented here focuses on sketch.

We present work that simplifies the task of creating sketch-based systems. We do this by separating sketch recognition into its own module. Within this module we separate the lower-level sketch recognition from the higher-level domain specification. We call the low-level task "shape recognition", because it recognizes a stroke as one of a set of shapes. We call the higher-level task "context recognition", because it involves comparing the shape to context. The work presented here focuses on context recognition. We introduce a description language, called StepStool[1], that facilitates describing shape-to-object relationships.

A. Related Work

Much work has been done on the task of low-level primitive shape recognition. Earlier systems [8], [12] use simple corner-detection-based recognizers to determine when a stroke should be broken into line segments, and classify strokes accordingly. Paulson and Hammond [6] created a system, called "PaleoSketch", that uses heuristics to classify a stroke as one of eight primitive shapes. We partially reimplemented PaleoSketch, and used it with our StepStool engine. Our work treats the shape recognizer as a black box and works with its output.

Much work has been aimed at developing domain-independent sketch recognizers. Hammond and Davis' approach to this task [4] involved creating a new language to describe a sketch domain—a lexicon of shapes, shape beautifications, and editing gestures for shapes—and a compiler for that language. Their system could make low-level shape recognitions (e.g., lines, circles, and curves). It involved describing geometric relations between them using a description language, called LADDER, to define higher-level shapes. Our language, StepStool, is inspired greatly by this work in both spirit and name.

Context in sketch recognition has been widely used. Alvarado et al. [1] incorporated context to distinguish various hand-drawn shapes from one another in the domain of free body diagrams. Similarly, Ouyang and Davis [5] present a machine learning approach to achieving domain independence in sketch recognition. They do this by incorporating what they call "local context" in their sketch recognition system. Local context consists of the area around the strokes being classified. A major drawback of Ouyang's approach (common to all machine learning approaches) is that it requires training data. We decided to use a description-based approach to avoid the need for a training step.

Several systems have been developed that adopt a remote-command-and-control architecture [3], [7], [9], i.e., an architecture in which a human controls a robot with a sketch-enabled mobile device. These systems have the ability to detect objects in the world. We built off the idea that we could annotate camera images with these detected objects and use the context of the camera image to control the robot's actions in the world. We describe our approach to this command-and-control architecture in [2].

II. Terminology

Before describing our work, we begin by defining the sketch terminology we use in this paper.

[1]StepStool stands for, "A Stroke to Environment, Peripherals, Scene, Terrain, and Objects Orientation Language".

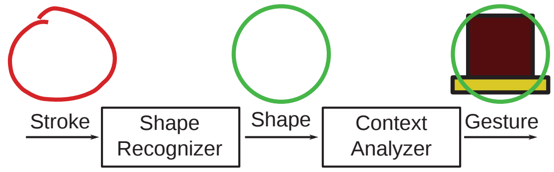


Fig. 1. The progression of a stroke through a shape into a gesture. In this example, a stroke is recognized as a circle, and then classified as a "Pickup" gesture because it was drawn on an item of cargo.

A. Strokes, Shapes, and Gestures

Fig. 1 shows the relationship between strokes, shapes, and gestures as we define them in our system. Strokes are a sequence of timestamped (x, y) coordinates, and shapes are geometric primitives. Strokes are interpreted to be one of a fixed set of shapes by a shape recognizer (like Paulson's PaleoSketch [6]). Gestures are shapes with meaning. They are given that meaning based on what the shape was drawn on and around, i.e., based on context. We call the part of the system that classifies shapes as gestures the "Context Analyzer". The focus of our work is on implementing this context analyzer.

B. Domain Knowledge vs. World Context

Our system assumes some kind of "world" exists. That is, our system assumes that the user is not only drawing sketches on a blank canvas, but drawing sketches on a world populated with objects. This involves making a conceptual switch. Fig. 2 demonstrates this idea. On the left are several strokes. Looked at alone, their meaning is unclear. When world context is added, as it is on the right, suddenly some meaning can be attributed to the strokes.

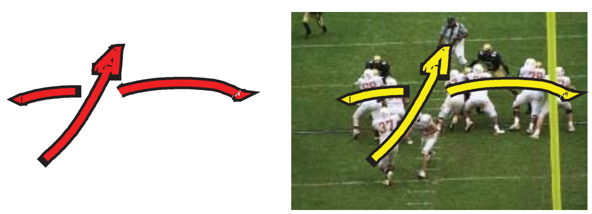


Fig. 2. Strokes' meanings may not be apparent without context.

Our system distinguishes between "domain knowledge" and "world context". We define "domain knowledge" as "the lexicon of shapes we need to recognize". For example, in circuit diagrams, the lexicon includes diodes, wires, and resistors. We define "world context" to mean "what is being drawn on and around". For example, in circuit diagrams a resistor's context is "between two wires". Our system uses both world context and domain knowledge to recognize gestures.

C. The World, the World Model, and the Scene

We distinguish between the "world", "world model", and "scene" in our system. The world is the place that you and I live in. In robotic systems, we want the system to interact with the world. The world model is where a system stores select information about the world. In the case of a robot, this could be the position and orientation of a soda can it is trying to pick up. The scene is a StepStool-specific construct that stores a subset of the information in the world model. It is separate from the world model to decrease the complexity of the system. This simplification facilitates writing correct StepStool descriptions.

D. Grounding Gestures to Referents

Lastly, we define the meaning of "grounding gestures to referents" in sketch recognition. Referents are defined by their association to a gesture. Each gesture must reference some object that the gesture is meant to act on. This object must exist in the world. For example, without a reference to the soda can, the system could not understand a "pickup" command. "Grounding" is the process of making that association.

III. StepStool

Inspired by LADDER [4], StepStool is a description language meant to simplify how context-rich sketch-based systems are implemented. StepStool allows the programmer to focus completely on the rules that describe gestures instead of having to implement shape and context recognizers. StepStool's purpose is to distill domain knowledge and world context, thus facilitating the implementation of context-rich sketch interfaces.

Our main motivation when designing StepStool, was to create a language that is straightforward and human-readable. StepStool is straightforward in the sense that small conceptual changes to the software specification result in small changes to the StepStool description files. As a result of being straightforward, shapes can be "overloaded", or assigned more than one meaning depending on context. For example, a circle could either be a command to pick something up, or a command to put something down. The difference would be what is being circled—an object, or the ground. Stepstool is human-readable relative to a programming language.

We implemented a StepStool engine that interprets StepStool descriptions. The following sections go into how the StepStool engine is intended to be used as a library (§III-A), describe what the syntax of the language looks like and why (§III-B), and present an example of the system in use in a robotic forklift (§III-C & §III-D).

A. Use in a General System

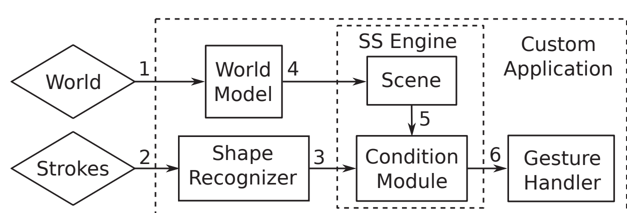


Fig. 3. The control flow of a system using the StepStool interpretation engine.

We implemented a StepStool interpretation engine that allows an application to use StepStool to implement a sketch-based interface. Fig. 3 shows a high-level view of how the engine is meant to fit into a sketch-based system.

1) An object is detected in the world and the world model is populated with a representation of that object.
2) The user draws a stroke, which is passed to the shape recognizer.
3) The shape recognizer classifies the stroke as a shape and passes the shape into the StepStool engine.
4) The scene is populated with relevant 2D data from the world model.
5) The Stepstool engine's condition module compares the shape with the objects in the scene.
6) The StepStool engine returns a gesture classification, that the system can handle in a system-specific way.

Note that despite not having access to the real world via sensors as robots do, we anticipate that StepStool could work well with non-robotic systems. StepStool's only interaction with the world is through the world model. The resultant gesture classification is sent to a system-specific handler. Thus, the system can use StepStool to classify new strokes based on the results of previous strokes' classifications. We describe this idea further in our future work section.

B. Syntax & Details

$\langle S\rangle$ ::= $\langle$*object*$\rangle$ | $\langle$*shape*$\rangle$ | $\langle$*gesture*$\rangle$

$\langle$*object*$\rangle$ ::= "object" $\langle$*ident*$\rangle$ { $\langle$*attrib-list*$\rangle$ } "end"

$\langle$*shape*$\rangle$ ::= "shape" $\langle$*ident*$\rangle$ { $\langle$*attrib-list*$\rangle$ } "end"

$\langle$*attrib-list*$\rangle$ ::= "has" $\langle$*ident*$\rangle$ { "," $\langle$*ident*$\rangle$ }

$\langle$*gesture*$\rangle$ ::= "gesture" $\langle$*ident*$\rangle$ "shapeof" $\langle$*ident*$\rangle$ "referent" ($\langle$*ident*$\rangle$ | "projected") { "given" $\langle$*ident*$\rangle$ $\langle$*ident*$\rangle$ } { $\langle$*cond*$\rangle$ } "end"

$\langle$*cond*$\rangle$ ::= ("shape" | $\langle$*full-ident*$\rangle$) ["not" | "!" | "approx" | "~"] $\langle$*cond-kw*$\rangle$ ($\langle$*full-ident*$\rangle$ | "true" | "false" | $\langle$*num*$\rangle$)

$\langle$*full-ident*$\rangle$ ::= $\langle$*ident*$\rangle$ ["." $\langle$*ident*$\rangle$ | "[*]" | "[%]"]

Fig. 4. The extended Backus-Naur Form for StepStool descriptions. $\langle$*ident*$\rangle$ (short for "identifier") is a placeholder for any string. $\langle$*num*$\rangle$ denotes a number. $\langle$*cond-kw*$\rangle$ is a condition keyword from fig. 6. Parenthesis with bars ("(" and ")" with "|") denote one of a set of elements must be inserted. Square brackets ("[" and "]") denote optional elements. Curly braces ("{" and "}") denote optional repetition (i.e., 0 or more).

StepStool defines three kinds of descriptions: shape descriptions, object descriptions, and gesture descriptions. Shape and object descriptions are lists of geometric attributes[2], while a gesture description lists the conditions that classify a shape as a gesture. Fig. 4 presents the syntax of all three StepStool description types in extended Backus-Naur Form.

Fig. 5 shows one example each of an object description, a shape description, and a gesture description. Object and shape descriptions are meant to describe the objects found in the scene and the shapes to be recognized. Object and

[2]Currently these attributes are all stored as integers.

```
object Person        gesture Follow
  has vx, vy           shapeof Circle
end                    referent person

shape Circle           given Person person
  has cx, cy           shape on person
  has r                shape sizeof person
end                  end
```

Fig. 5. Examples of an object, shape, and gesture description. The "Follow" gesture is a command to follow a circled person. The person and circle descriptions exist to be referenced in the gesture. The person is the gesture's referent. The circle is the gesture's shape.

shape attributes—denoted by the `has` directive—are compared together by the StepStool engine according to the conditions listed in gesture descriptions. By default, each shape and each object is given position and size attributes. These attributes are used to evaluate the conditions listed in fig. 6.

Condition Keywords						
inside	on	sizeof	below	leftof	larger	is

Fig. 6. StepStool condition keywords.

A gesture description lists the conditions that classify a gesture. These conditions are of the form "`x [cond] y`", where `[cond]` is one of the conditions from fig. 6. For example, the gesture in fig. 5 has two conditions: "`shape on person`" and "`shape sizeof person`". "`shape on person`" means the shape must be drawn on the person. "`shape sizeof person`" means the shape must be the same size as the person.

The three elements of a gesture description, that are not in fig. 6 are the `shapeof`, the `referent`, and the `given` keywords. The `shapeof` condition limits what kinds of shapes can be classified as the gesture. The `referent` directive assigns a referent to the gesture, i.e., it denotes what the gesture is meant to be grounded to. This referent must appear in a `given` directive. The `given` directive allows the description to reference objects. This is useful when the referent is an object in the scene or when the drawn shape needs to have a specific geometric relationship to specific objects in the world.

```
object Truck         object Box end
  has vx, vy
end                  gesture Path
                       shapeof Line
shape Line             referent projected
  has x1, y1           shape not on Box[*]
  has h2, y2           shape not on Truck[*]
end                  end
```

Fig. 7. Additional examples of object, shape, and gesture descriptions. The gesture demonstrates the use of the "`projected`" keyword and the "`[*]`" (all) modifier. The "`Path`" gesture is a command to move. According to the gesture description, a line will be classified as a `Path` only if the line is not drawn on any boxes or trucks.

Some gestures do not act upon an object in the world.

Instead, those gestures have an effect upon the system. For example, in fig. 7 the `Path` gesture is a command for the forklift to follow a path. This gesture does not directly effect any object. Instead, it effects the system. This interaction between the `Path` gesture and the system can be modeled with projection. The gesture is projected into the world. The `Path` gesture's referent is its projection. Being grounded to "`projected`" means that a new referent must be created for this gesture.

```
gesture Select
  shapeof Circle
  referent body
  given Body body
  shape on body
  shape approx larger body
end
```

```
gesture AddBody
  shapeof Circle
  referent projected
end
```

Fig. 8. The two gestures here demonstrate the need for an evaluation priority. When the conditions of the one on the left are satisfied, the conditions for the one on the right will also be satisfied. To determine which classification to make, StepStool implements a priority queue of gestures. The programmer must assign the `Select` gesture higher priority so that when a body is circled, it is selected. When no body is circled, the `AddBody` gesture is classified and the system will project a circular body.

Fig. 8 shows a situation where a shape could be classified as more than one gesture, namely a `Select` gesture or an `AddBody` gesture. To break these sorts of ties, the gestures are stored in a prioritized list. Gestures are evaluated sequentially by priority. StepStool classifies a shape as the first gesture found to be valid.

StepStool also has several constants, modifiers, and operators. Constants include all the numbers, `true`, `false`, and `shape`. `shape` is used as a constant to allow a gesture to compare the drawn shape (the one being classified) to the objects in the scene. The `not` (or `!`) and `approx` (or $\sim$) modifiers negate and loosen conditions, respectively. The statement "`shape larger body`" evaluates to false if the shape is the same size as `body`. The statement "`shape approx larger body`", however, evaluates to true when the shape is the same size as the body. Thus, we do not need to implement an "`outside`" condition; we write "`not inside`" instead.

Lastly, there are situations where each or all of a certain class of object need to be referenced. For example, the path in fig. 7 cannot intersect any `Box` or `Truck` objects ("`shape not on Box[*]`" and "`shape not on Truck[*]`"). Similarly, when commanding the forklift to pick up some cargo, the drawn shape must be the same size as at least one item of cargo ("`shape approx sizeof Cargo[%]`"). For this purpose, we included the `[*]` and `[%]` modifiers to mean "ensure all of this type of object satisfy this condition" and "ensure at least one of this type of object satisfies this condition".

C. Implementation: The Forklift Project

We implemented a robotic control interface [2] to an autonomous forklift [11] using our StepStool interpretation engine. In this section, we describe a subset of the descriptions used to implement [2]. With this subset, we demonstrate StepStool's intended use in a system, and we specify the purpose of each description type.

```
object Forklift
  has vx, vy
  has loaded
end
```

```
object Pallet
  has pickup
  has id
end
```

```
object Truck
  has vx, vy
  has pallets
end
```

Fig. 9. These three object descriptions are lists of attributes describing the objects that appear in the world. In our example there are only forklifts, pallets, and trucks in the world.

Fig. 9 shows example object descriptions. The `Forklift` description is meant to make the state of the robot visible to the StepStool engine. The `Pallet` and `Truck` descriptions are meant to describe common warehouse objects[3]. The `vx` and `vy` attributes keep track of velocity, and the `loaded`, `pickup`, and `pallets` attributes keep track of elements of the objects' states; namely whether the forklift is loaded (i.e., already carrying a pallet), whether a pallet is already scheduled for pickup, and how many pallets are on the truck.

```
shape Dot
end
```

```
shape Line
  has x1, y1
  has x2, y2
end
```

```
shape Circle
  has x, y, r
end
```

Fig. 10. Each of these three shape descriptions is a representation of what the shape recognizer recognizes. Note that the `Dot` description has no attributes, because position and size attributes are enough to describe it.

Fig. 10 shows example shape descriptions. These descriptions are StepStool's interface to the shape recognizer. They describe useful shape attributes. The shape recognizer we use in this example recognizes only three shapes: a dot (or "click")[4], a line (or curve—any open region), and a circle (or polygon—any closed region).

Fig. 11 shows example gesture descriptions. Notice the use of gesture overloading in these examples. There are contexts in which a circle denotes a "pick up that pallet" command and there are contexts in which it denotes "place that pallet down". The dot has a similar dual nature. The priority given to these gestures is left-to-right then top-to-bottom. Thus, if the forklift is not "loaded" (i.e., not holding any cargo) when an appropriately-sized circle is drawn and no pallet is found in the scene, then the `PickUpP` gesture will be recognized. If the user draws a circle around a detected pallet, however, the StepStool engine would recognize the `PickUpR` gesture.

[3]Pallets are the smallest unit of transported goods handled by a forklift. Pallets provide a way to package cargo so that forklifts can manipulate that cargo—by inserting the forklift's tines in two slots in the pallet.

[4]Note that the `Dot` description has an empty body, because the default position and size attributes suffice to describe dots. The purpose of adding this "empty" description is to distinguish it from other shapes.

```
gesture DropOffR
  shapeof Dot, Circle
  referent pal

  given Pallet pal
  given Forklift fork

  shape on pal
  fork.loaded is true
end
```

```
gesture PickUpR
  shapeof Dot, Circle
  referent pal

  given Pallet pal
  given Forklift fork

  shape on pal
  fork.loaded is false
end
```

```
gesture DropOffP
  shapeof Circle
  referent projected
  given Forklift fork

  shape ~sizeof Pallet[%]
  fork.loaded is true
end
```

```
gesture PickUpP
  shapeof Circle
  referent projected
  given Forklift fork

  shape ~sizeof Pallet[%]
  fork.loaded is false
end
```

```
gesture Path
  shapeof Line
  referent projected
  shape not on Forklift[*]
  shape not on Truck[*]
  shape not on Pallet[*]
end
```

Fig. 11. These gesture descriptions specify the conditions under which a shape is to be classified as a gesture. They are sorted by priority: left-to-right then top-to-bottom. When multiple gestures describe the scene accurately, the highest priority gesture triggers a classification. The remaining gestures are not evaluated.

D. Forklift Project Run-Through

This section follows the process the StepStool engine takes when interpreting a gesture. For this example, we will assume the descriptions from fig. 9, fig. 10, and fig. 11 have been loaded into the system in order.

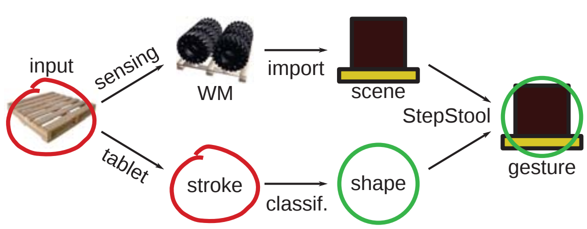


Fig. 12. The process of recognizing a pickup gesture. The top path shows the steps through the forklift's object detection system. The bottom path shows the steps through the interface's recognition system. At the far right, we see that a pickup gesture was recognized.

The process starts with the user drawing a stroke (shown in red in fig. 12) on top of a pallet. The shape recognizer recognizes it as a circle (shown in green along the bottom path in fig. 12) and passes the shape to StepStool. The forklift's sensors have already detected the pallet and made some representation of it in the world model. When StepStool sees that a stroke has been drawn, it imports all the objects displayed on the canvas into the scene. Note, that one of these objects is a `Forklift` object. This is a representation of the forklift's current state.

The StepStool engine operates as follows. First, since the shape is a circle, all gestures that include "`shapeof Circle`" (the top four in fig. 11) are evaluated in order of priority. First, StepStool evaluates the `DropOffR` gesture. According to the `given` statements there must be at least one pallet and one forklift object in the scene. There are—as mentioned above, they were just loaded into the scene—so the process continues. First, "`shape on pal`" is evaluated. This evaluates to true, so the next condition, "`fork.loaded is true`" is evaluated. This second statement evaluates to false, because the forklift is not loaded. The StepStool engine discards this gesture and moves on to the next one: `PickUpR`. The StepStool engine goes through the same steps as with the previous gesture description. This time, the last condition, "`fork.loaded is false`" evaluates to true. All conditions in the `PickUpR` gesture have evaluated to true, so the StepStool engine returns it. The engine does not look at either of the remaining 2 descriptions.

E. An Aside about "Correcting" Shape Classifications

Early versions of the forklift project's shape classifier often returned inaccurate shape classifications. If a user's stroke was not a perfect circle or overlapped itself to too great a degree, the shape recognizer would classify the stroke as a polyline instead of a circle. We found that the inaccuracy of our primitive shape recognizer was acceptable, because there were distinctive elements of the scene in our domain, that distinguished mis-classified polylines (strokes the user intended to draw as circles) from actual polylines. Thus we think that it is worth noting that adding context (with StepStool or otherwise) has the ability to improve stroke-to-gesture recognition in certain cases.

IV. Future Work

Development of our StepStool system was directed heavily by the forklift project (described in [11]). We expect that the language in its current form will not suffice in implementing many sketch-based systems. However, we see a lot of potential in our approach. The following sections list what we consider promising extensions to StepStool. They have the potential to make StepStool more broadly applicable.

A. Non-Robotic "Real Worlds"

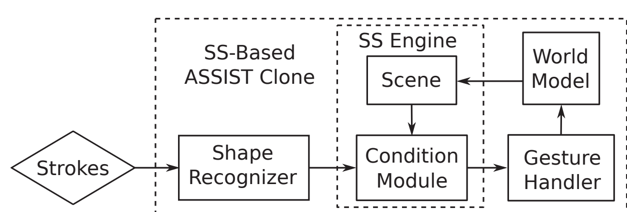


Fig. 13. The architecture used when implementing a system without the ability to sense the real world. The world model stores the state of the canvas. Gestures act directly upon the world model by adding things to the world model, or editing things in the world model.

We propose making an ASSIST [1] clone that uses drawing on the world, to test the efficacy and flexibility of using StepStool on systems with no connection to the real world (i.e., non-robotic systems). Fig. 13 shows the architecture of such a system. The "world" that the user is drawing on is the world of the canvas. When the program starts, the world model is empty. The user draws an outline that is recognized as a circle. The circle is interpreted to be a body and added to the world model. Later, when a circle is drawn around that body, the system classifies it as a selection gesture.

```
gesture StartBody
  shapeof Line
  referent projected
  shape not on Body[*]
end
```

```
gesture AddBody
  shapeof Circle
  referent projected
  shape not on Body[*]
end
```

```
gesture AddWheel
  shapeof Circle
  referent projected
  given Body body
  shape on body
end
```

```
gesture ContBody
  shapeof Circle
  referent bp
  given BodyPart bp
  shape approx on bp
end
```

```
gesture EndBody
  shapeof Line
  referent bp
  given BodyPart bp
  shape approx on bp
end
```

Fig. 14. Possible gesture descriptions in the ASSIST implementation.

Fig. 14 gives examples of how these gesture descriptions could look. The `StartBody`, `AddBody`, and `AddWheel` gestures' referents are "`projected`" because their purpose is creating bodies in the world model. In the case of `StartBody`, a `BodyPart` is made, because the shape that was drawn did not fully close. Later, when a continuing or completing stroke finishes the body, it can be added to the world model.

B. Other Input Modes

We would like to extend StepStool to include other input modalities. For example, we could write a hand gesture recognizer that compares hand gestures against context to yield context-rich hand gestures. Examples of this would include work from [10], in which machine learning approaches are used to distinguish between a subset of flight deck hand signals. In this system, we would have to switch our vocabulary from "strokes" to "hand movements" and from "shapes" to "hand signals". "Hand gestures" would then be specific commands. The context of these commands would be the state of the flight deck.

V. Conclusion

In this paper, we presented our approach to implementing context-rich sketch-based interfaces. Our approach was based on distilling the details of context into a description language we call StepStool. We implemented and described a StepStool engine, that followed StepStool descriptions to make gesture interpretations. We described how the engine was meant to be used in an arbitrary system. Finally, we gave an example of StepStool's use in the mobile command interface to a robotic forklift and followed the recognition of a stroke in that system through a shape into a gesture.

Acknowledgments

We gratefully acknowledge the support of the U.S. Army Logistics Innovation Agency (LIA), the U.S. Army Combined Arms Support Command (CASCOM), the Agile Robotics group at CSAIL, MIT, and Randall Davis for his guidance in this work.

This work was sponsored by the Department of the Air Force under Air Force Contract FA8721-05-C-0002. Any opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

References

[1] Christine Alvarado and Randall Davis. Resolving ambiguities to create a natural computer-based sketching environment. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, pages 1365–1371, Seattle, Washington, USA, August 2001. Morgan Kaufmann Publishers.

[2] A. Correa, M. R. Walter, L. Fletcher, J. Glass, S. Teller, and R. Davis. Multimodal interaction with an autonomous forklift. Osaka, Japan, March 2010.

[3] Terrence Fong, Charles Thorpe, and Betty Glass. PdaDriver: A handheld system for remote driving. In *In Proceedings of the IEEE Internationall Conference on Advanced Robotics*, July 2003.

[4] Tracy Hammond and Randall Davis. Ladder, a sketching language for user interface developers. *Elsevier, Computers and Graphics*, 28:518–532, 2005.

[5] T. Y. Ouyang and R. Davis. Chemink: A natural real-time recognition system for chemical drawings. In *International Conference on Intelligent User Interfaces (IUI '11)*, February 2011.

[6] B. Paulson and T. Hammond. Paleosketch: Accurate primitive sketch recognition and beautification. In *Intelligent User Interface Conference (IUI '08)*, January 2010.

[7] D. Sakamoto, K. Honda, M. Inami, and T. Igarashi. Sketch and run: A stroke-based interface for home robots. In *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, pages 197–200, Boston, MA, April 2009.

[8] Tevfik Metin Sezgin, Thomas Stahovich, and Randall Davis. Sketch based interfaces: Early processing for sketch understanding. In *Workshop on Perceptive User Interfaces, Orlando FL*, pages 1–8, 2001.

[9] M. Skubic, D. Anderson, S. Blisard, D. Perzanowski, and A. Schultz. Using a hand-drawn sketch to control a team of robots. *Autonomous Robots*, 22(4):399–410, May 2007.

[10] Yale Song, David Demirdjian, and Randall Davis. Continuous body and hand gesture recognition for natural human-computer interaction. In *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2011.

[11] S. Teller et al. A voice-commandable robotic forklift working alongside humans in minimally-prepared outdoor environments. In *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*, May 2010.

[12] Bo Yu and Shijie Cai. A domain-independent system for sketch recognition. In *GRAPHITE '03: Proceedings of the 1st International Conference on computer Graphics and Interactive Techniques*, pages 141–146, February 2003.

Sketch Input of Engineering Euclidean Solid Models

Pedro Company
Institute of New Imaging Technologies
Universitat Jaume I
Castellón, Spain
pcompany@emc.uji.es

Peter Varley
Institute of New Imaging Technologies
Universitat Jaume I
Castellón, Spain
varley@emc.uji.es

Raquel Plumed
Dept. Mechanical Eng. and Construction
Universitat Jaume I
Castellón - Spain
plumed@emc.uji.es

***Abstract*—This position paper describes some open problems of sketch input of engineering Euclidean solid models. After a brief historical introduction, we discuss traditional design using pencil and paper, and how this paradigm has been adapted (or forgotten) by various current approaches to computer interpretation of sketches. We discuss three particular open problems, those of modalities, annotations, and assemblies. We also analyse current perception-based approaches in order to identify the most important areas in which further work is needed: detection of geometrical, perceptual and manufacturing cues, and a taxonomy of cue interdependencies which would prevent incompleteness and contradictions.**

***Index Terms*—Pen computing, Sketch-based modelling, Sketching interface, Cues.**

I. INTRODUCTION

Pen computing refers to those computer user interfaces where users interact with computers by means of pens which guide pointers moving around a screen. In this paper, such interfaces are considered distinct from classic keyboard/ display/mouse environments only as far as they are used to simulate the behaviour of actual paper and pencil. We are only interested in "virtual" paper and pencil which allows free sketching and annotating, not in the "pointing mode" of WIMP environments. Similarly, we are interested in handwriting as far as it conveys *annotations* which complement drawings, but not in handwriting recognition as a standalone problem.

Johnson et al's [1] recent overview of the evolution and current state of the art of computational support for sketching describes the advantages of sketching as part of design and the open problems which must be solved before computer-based sketching can be integrated into the design process. However, it misses some important developments in the area of Sketch-Based Modelling (SBM), which aims at developing programs capable of producing 3D geometric models from 2D sketches.

SBM aspires to encourage professional designers to discard physical pen and paper in favour of a computer-based interface in the creative phase of the design process. However, it often fails to produce the output 3D geometric model which best matches the design intent embedded in the input sketch, and its current success ratio is not good enough to persuade designers and engineers to abandon actual paper and pencil.

In order to make the problem more tractable, SBM is generally divided into sub-problems.

The first distinction to be made is between freeform surfaces and Euclidean solid geometry (in which the object wireframe can be constructed from straight lines and arcs). Different techniques are required: the input information is different (edges for Euclidean solids vs. contour for freeform surfaces), as is the output (Euclidean solid geometry typically requires much greater geometrical exactness). In this paper, we discuss only Euclidean solid geometry—the sheer quantity of recent work on freeform surfaces (of which we highlight [2], [3], [4], [5] and [6]) makes it impractical to discuss that too.

Another important distinction is that between *perception-based* approaches, using perceptual cues to guide the interpretation process, and *geometry-based* approaches such as those of [7] and [8], which try to produce 3D models solely from the geometric information contained in the 2D input. In this paper, we argue strongly in favour of perception-based approaches: the problem of determining the embedded design intent is inherently cognitive, not purely geometrical.

The prevailing paradigm for sketch input of engineering Euclidean solid models is representative of the perception-based category of approaches [9]: (i) converting 2D sketches of wireframe drawings into 2D line drawings; (ii) deduce the faces implied by the line drawings; and (iii) use optimisation-based inflation to produce 3D geometric models from 2D line drawings. This approach was first proposed by [10], and later developed by other authors including [11] and [12]. Recent advances in face detection [13] and inflation [14] combine well, leaving vectorisation as the main open problem.

Alternative paradigms which correspond better to design engineering practice but for which the algorithms are not so well-developed include use of natural line drawings [15] or linear 3D scaffolds as guidelines for freeform shapes [16].

In choosing between paradigms, it is important to be aware how the tool conditions the task. In Section II, we discuss traditional design using pencil and paper, and how this paradigm has been adapted (or forgotten) by various current approaches to computer interpretation of sketches.

In Section III, we discuss three particular open problems, those of modalities, annotations, and assemblies.

In Section IV, we analyse the current status of optimisation-based inflation of 3D engineering models from 2D design sketches. We highlight the main limitations of this approach, and argue in favour of some strategies which may be helpful in overcoming these limitations. Improving the detection of design intent is the key to solving both problems, so detection

This work was partially funded by financial support from the Ramon y Cajal Scholarship Programme and by the "Pla de Promoció de la Investigació de la Universitat Jaume I", project P1 1B2010-01

of geometrical, perceptual and manufacturing cues embedded in the input is the first step. Then, a full taxonomy of cues and their interdependencies, aimed at preventing incompleteness or contradictions, will help to make explicit this design intent.

II. Historical Background

The origins of SBM can be traced back to the pioneer work of Perkins [17], [18], who analysed how people perceive drawings as representing objects, which geometrical relationships must be obeyed, and the circumstances in which geometrical relationships can be ignored.

The area now known as SBM comes from developments which were earlier known as Geometric Reconstruction. The former goal of Geometric Reconstruction was extracting information from old engineering blueprints, essentially archaeological recovery of lost know-how. However, the task proved too difficult. The main (and still unsolved) problems were those of *vectorisation* and *annotation*.

The problem of *vectorisation* (converting raw data, patterns of adjacent black and white dots, into coherent information) was too complex. The most popular academic contributions are still based on Sparse Pixel Vectorisation (SPV) algorithm [19], but Bartolo et al.'s recent work [20] is interesting as it describes the particular problem of extracting lines from scribbled drawings.

So was the problem of *annotation.* Engineering drawings convey both 3D information represented through diverse views (main orthographic views, particular views, sections, etc) and annotations (dimensions, tolerances, etc) [21].

In the short term, the problem was solved by brute force: several CAD companies offered commercial "paper-to-CAD conversion", translation of old engineering blueprints as an additional service.

It is worth noting that architecture community continues to aim for this original goal [22]. However, in the engineering community, the main goal of reconstruction changed in the 1990s. Nowadays, most systems are oriented towards conceptual design via sketch-based modelling [23].

Thus we can note how the goal has changed over time from *vectorisation* (2D + paper => 2D + computer) via *reconstruction* (2D + paper => 3D + computer) to full *sketch-based modelling* (mental model => 3D + computer). For a more detailed exposition of our view of this historical background of computer interpretation of engineering drawings, see [24] and [25].

Currently, as reported in recent surveys such as [9] and [26], SBM contains a variety of sub-problems. As we shall see in the sections below, there is no general approach which solves them all. Different *critical features* (see [9] for a taxonomy) produce different bottlenecks, and levels of development are different for each critical feature.

Design intent and CAD have been linked for a long time. However, the definition of design intent is ambiguous. In 1989, design intent was associated with design constraints and the methods of manipulating design constraints during product design activities [27]. This definition continues to be used even now. However, the word *design* in the CAD community is a synonym of *model*: modelling is representing the design intent in some way, and design intent is a basic concept of *design for change*. The model being created is flexible through changes, but the intent remains the same. Some work has been done in the SBM sector to cope with this new understanding of design intent as design-for-change (one example would be the idea that sketching a single line and then removing the central segment implicitly conveys the design intent that the remaining two segments are collinear).

However, we prefer to understand design intent as a broader concept, a mix of: *geometry*, as far as it is linked to the shape; *psychology*, as far as it is not always explicit in the sketches; and *engineering*, as far as it is linked to the function.

When geometry dominates, design intent is mainly conveyed through geometrical features. These have already been studied as *regularities* [11], [28], [29].

The psychological component appears because information not explicitly included in the input is nevertheless perceived through *perceptual cues*. Fundamentals of perceptual cues have been studied in the general perception literature, e.g. Palmer [30] and Hoffmann [31]. More specific to SBM, Perkins [18] was probably the first to realise that looking at pictures is different from looking at objects.

When function dominates, design intent is mainly conveyed through *engineering features*, where attention is given to the machining process which creates the given geometry. For example, in Figure 1, the cylindrical holes (geometry) are naturally interpreted as drilled holes (machining features), and the blends (geometry) are naturally interpreted as rounding (machining features). Both are modifications to the geometrical skeleton.

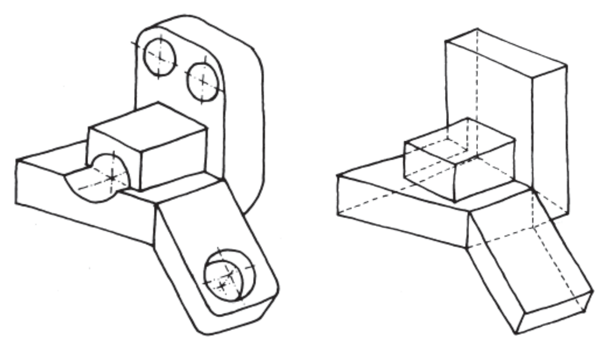


Figure 1. Geometry with features (left) and skeleton (right)

Consequently, we can define design intent as the set of intentions in sketches, conveyed through cues, which, when perceived, reveal regularities or features of the object.

It should be noted that only a few design intent cues have been studied in any detail—examples include face planarity, edge parallelism ([15]) and bilateral symmetry ([32], [33], [34] and [4]). There are clear examples where design intent should be studied further.

III. Open Problems with Virtual Paper and Pencil

This section discusses three particular open problems.

A. Pencil and Paper Modalities

How do engineers actually use pencil and paper in practice? Do current sketching tools offer all the modalities which users of pencil-and-paper expect? What is missing?

User studies such as [35] assert that current SBM tools are still less usable than pencil and paper sketches. The "hardware" of pencil and paper sketching is simple, but its operation is sophisticated as pencil and paper sketching is multimodal. Note, for example, the scaffolding lines, highlighting and hatching in Figure 2, the different inking of axes, hidden lines and autocorrection in Figure 3, the different uses of overtracing (for decoration and for "thinking over the line"), and the annotations in Figure 4.

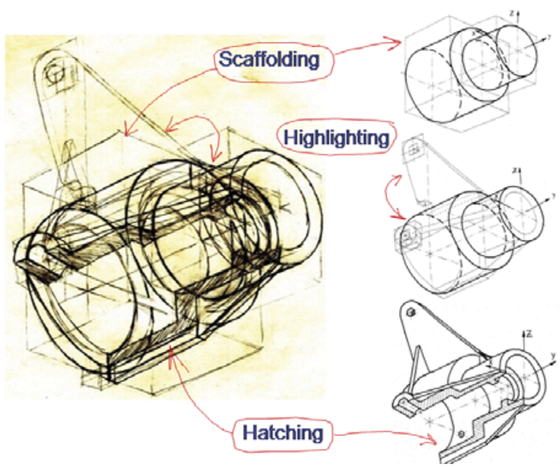


Figure 2. Multiple Modes (I)

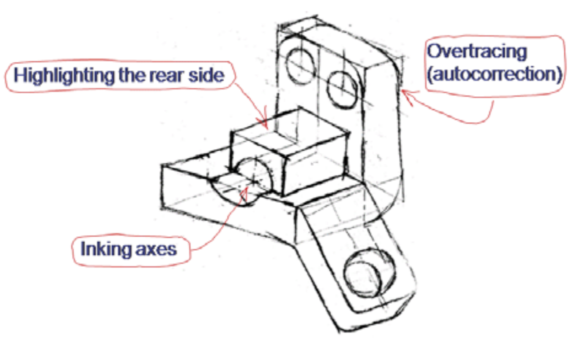


Figure 3. Multiple Modes (II)

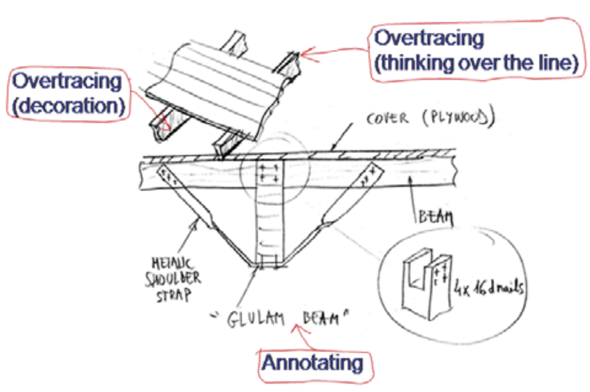


Figure 4. Multiple Modes (III)

The "overloaded" use of thin lines in Figure 3 is interesting: the human eye can easily distinguish scaffolding lines, hidden lines and axes, even though their graphic representation is the same. It is also worth noting the informal mixing of views in Figure 4: the same sketch contains an orthographic view, a detailed view and a pictorial view.

Thus we can conclude that the set of operating modes available in pencil and paper sketching is complex. Using just the pencil, we have: *views of the object* (orthographic, detailed, pictorial); *scaffolding*; *highlighting*; *hidden lines*; *axes*; *hatching*; *overtracing* (autocorrection, thinking, decoration); and *annotating*. What is more, the switching strategy is automatic and non-intrusive.

Some typical computer operations are also possible with pencil and paper sketching, albeit with a change of tool. For example, lines can be *erased* using correction fluid, a literal *cut and paste* can be performed using scissors and glue, and *regularities* can be created by tracing over a displaced copy of the original. However, in these specific cases, the computer already has an advantage.

This list of operating modes is illustrative, but far from exhaustive. We still require a full taxonomy of operating modes, including their mutual relationships and the cues used to discriminate between them.

There are two possible strategies we can take when we aim to reduce the gap between pencil and paper and SBM tools: either we try to replicate pencil and paper as closely as possible, or we try to add extra features to compensate for the loss of some of the power of pencil and paper.

Until recently, SBM tools have aimed to provide as many pencil and paper modes as possible. However, they have been based around the wrong interface paradigms, those of command-driven and menu-driven interfaces—the greatest advantage of the pencil is that it requires no commands or menus to use. But identifying an alternative interface paradigm is a difficult task in itself, and we do not know of one which would suit experienced design engineers.

In terms of adding extra features, there is broad agreement on some of the main advantages which computer-based systems already possess: it is easier to edit work; it is easier to file work; and it is easier to interface work to other applications. This list could easily be extended by adding some current CAD operations which help in reducing editing tasks: *extrusions* and *sweeps*.

But more work is needed. To achieve our final goal, making CAD tools as easy to use as pencil and paper, we need both hardware advances and software advances. Two example problems illustrate this. Firstly, current graphics tablets have proved to be less comfortable to use because of the small gaps in both time and location between the pen tip and the cursor [36]. Secondly, use and maintenance of computers still require technical knowledge which several designers have told us (and we can only agree) should not be part of their job.

User studies also assert that current SBM tools do not possess significantly improved functionality [37]. Before we can implement SBM properly, two questions must be posed: how many functions can be provided without buttons and menus? and how many functions does a design engineer require?

Clearly, if the second answer is larger than the first, a new interface paradigm is required.

It is already clear that full modeless operation is not the goal. If we wish to replicate real pencil and paper scenarios in a virtual environment, we must be aware that real pencil and paper sketching includes a rich variety of modes. Replicating pencil and paper scenarios in virtual environments is still infeasible—although the goals of SBM have supposedly been accomplished, practical implementations have unfriendly interfaces.

So far, three types of input have been considered for SBM: perfect line drawings, line drawings containing minor geometric errors, and freehand sketches. But this too is an oversimplification, and in need of refinement. For more details on this topic, see [38].

B. Annotated Sketches

This section discusses the problem of interpreting annotated sketches. Which strokes are annotation rather than object? What does the annotation mean? How should it be applied?

In general, we can consider three purposes for sketching: *thinking*, *talking*, and *prescribing* [39]. We can combine this with two levels of geometric information: *line drawings* and *sketches*; and with two levels of non-geometric information: *with annotations*, and *without annotations*. The particular open problem we discuss here is that of interpreting *annotated engineering sketches*.

Annotations is a generic term. It includes: *dimensions*; *cut views with hatching*; *icons*; *symbols*; and many other standardised conventions.

The proposed approach for producing 3D models from annotated engineering drawings is: (i) capture and record the data; (ii) separate annotation data from drawing data; (iii) interpret the drawing data (without annotations); (iv) interpret the annotations (separately from the drawing)—see, for example, [38]; (v) apply the interpreted annotations to the interpreted drawing.

The subtask of interpreting engineering symbols is not trivial. Behind even apparently quite simple drawings, there are hundreds of standards defining the exact meaning of many symbols and conventions (ISO, DIN, BSI, ANSI, …). It is obvious that communication of relevant information depends on the meaning of symbols (see, for example, [40] for historical examples where misunderstanding of symbols has caused information loss).

The problem becomes still more challenging when we recognise that new standards (e.g. ASME Y14.41-2003) already allow annotations in 3D models.

Currently, computers are blind to these annotations—the annotations are just labels added to the model. The user can read and modify them, but the geometric engine does not use them either to construct, to edit or to validate the model.

One interesting open problem area is that of interpreting sketched data input for Computer-Aided Engineering (CAE) applications. The particular problem here is that data is input through either of two WIMP interfaces: stand-alone CAE preprocessors which define both geometric data and attributes, and combinations of CAD applications (which export the geometry) and downstream CAE preprocessors (which add attributes).

What we want is a tool which takes as its input the sketches which designers typically draw in order to fix their ideas before interacting with CAD preprocessors, and which creates as its output a file which meets the specification of the desired analysis tool. It would be based on two reasonable assumptions: that the input sketches are drawn directly onto a computer screen which acts as virtual paper and pencil; and that the user is still in the process of conceptual design and is not ready to progress to a detailed design stage.

Hence our goals are: to supply the user with a computer interface similar to pencil and paper; to minimise the amount of information required from the user; and to give the user more freedom in inputting and editing this information.

Examples of such tools include Pre/Adef [41] and FEAsy [42]. Although they represent the current state of the art, they remain unsatisfactory. Their advantages are: they produce valid output (whereas all that pencil and paper can do is fix ideas), no training is required, and the user is not forced by the system to add unnecessary information such as dimensions. Their disadvantage is that explicit mode-switching is required (although only when switching to a very different task) and that practice has shown that users do not always feel comfortable with an on-line parser. Such interfaces are similar to, but not yet as good as, real paper and pencil.

C. Assemblies

This section considers the possibility of sketching assemblies of parts.

Currently, we are limited to reconstructing isolated parts. However, engineering parts typically work in combination with one another. Our colleagues in mechanical design engineering sketch not only single parts but also assemblies of several parts. If they are to be persuaded to use SBM systems, SBM systems must support sketching of assemblies, perhaps by defining and implementing a set of symbols which can help an SBM system to assemble 3D models obtained from 2D sketches.

The basic guidelines of such an approach should be that: the symbols must themselves be sketched, as part of a natural design process; the meanings of the symbols must be robust (in the sense that they must be correctly interpreted by the geometric engine in charge of assembling the parts); and the symbols should not be subject to the faults of current sets of CAD operations.

We must therefore analyse what is wrong with current CAD applications. In current applications, components can be positioned within the product assembly using either reminiscent absolute coordinate placement methods, or *mating conditions*. Mating conditions are definitions of the relative positions of components with respect to one another, for example alignment of axes of two holes, or distance of two faces from one another. The final position of all components based on these relationships is calculated using a geometry constraint engine built into the CAD or visualisation package.

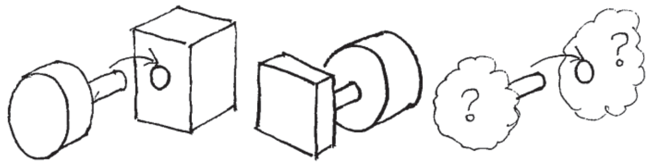


Figure 5. Assembling conceptual models

Commercial tools which include mating conditions are familiar, and assist the 3D CAD user to get an intuitive and friendly set of constraints (for example, SolidEdge's FlashFit option). As users place parts in an assembly, assembly relationships position new parts relative to parts already in the assembly. There are several possible assembly relationships.

However, we find a major drawback. Only complete and consistent parts can be assembled. Current CAD packages can assemble concrete shapes (e.g. Figure 5 left), and even modify them after assembling (e.g. to Figure 5 centre), but cannot assemble blurred or incomplete shapes (e.g. Figure 5 right). We need a complete design before assembling may proceed.

Our vision is creating a sketch-based environment in which design intent can be specified, so that we are able to assemble different parts which are not yet fully defined.

IV. Cues and Design Intent

As noted above, when geometry dominates, design intent is mainly conveyed through geometrical features or *regularities*. Algorithmic solutions such as [43] can process this information (although they could still benefit from improvement).

In other cases, cues are based on perception, and we meet a problem: most researchers in perception only consider looking at physical objects, not looking at pictorial representations. Hence, although perceptual cues have been studied in the literature, applying them to produce explicit geometric constraints which may guide the construction of 3D models has barely been considered. We know of nothing in this area which goes beyond the limited detection of features in [15].

Finally, in *engineering features* attention is given to the machining process which creates the given geometry. Some cues linked to simple manufacturing features have already been studied. Extrusion is one such [44]. Rounds and chamfers are another simple but very frequent manufacturing feature [45]. Some interactions between features, such as the one shown in Figure 6 (left), interaction between two rounds, have also been studied, but there are many which have not.

When we detect several cues, the geometric criteria derived from these cues may well contain contradictions. Furthermore, as we add more cues, the mutual dependencies arising between them will lead to redundancies amongst the resulting geometric constraints, a problem which some authors (e.g. [28]) aim to detect and circumvent. General strategies to cope with these problems have already been proposed. For example, multiagent-based management of cues was first proposed by Juchmes and Leclerc [46], and [47] has recently reported that it is feasible to simultaneously detect multiple cues.

Nevertheless, a simple example is enough to show that much more work is required here. Consider the regularities which appear in the upper face of the parallelepiped represented in Figure 6 (right): 1) a face planarity constraint on (*a, b, c, d*); 2) two parallelism constraints (*a//c* and *b//d*); 3) two perpendicularity constraints ($a \perp b$ and $a \perp d$) detectable by line orthogonality; 4) four skewed symmetries (with axes *e1*, *e2*, *e3* and *e4*); and 5) four orthogonal corners (*ab, bc, cd, da*) detectable by line orthogonality. Thus just one quadrilateral face can generate thirteen (or more) constraints. Considering the parallelipiped wireframe as a whole, there are similar sets of constraints for the remaining five faces, giving us a total of 78 constraints. Yet, if we use the classical optimisation approach to inflate the model, we have just eight variables: the z-coordinates of the vertices. With more constraints than variables, but no inconsistencies, it is clear that some constraints must be redundant.

We can note that, back in 1971, the third paper in [18] recommended finding the "geometrically maximal combination of constraints". This hides a problem which, forty years later, has still not really been solved: how geometric constraints interact in 3D.

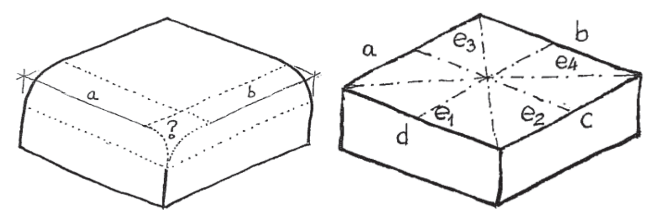


Figure 6. Interdependencies between features (left) and cues (right)

V. Conclusion

WIMP interfaces are not appropriate for conceptual design stages. However, SBM tools are not yet used—they look suitable for the task, but need improvement. We can classify problems into (a) those where a reasonably good solution exists (improvements may still be possible) and (b) open problems.

We have studied three open problems of sketching interfaces which discourage designers from using virtual paper and pencil, those of modalities, annotations, and assemblies.

We have also described the main weaknesses of cue detection and the process of finding the underlying implied design intent. We understand the capture of complex design intent from input sketches as a mix of geometry, psychology and engineering. The detection of cues needs improvement in all three areas, but is most critical in the areas of perceptual cues and engineering features.

Developing a catalogue of cues and their interdependencies should help to reduce current problems of incompleteness and contradictions which cause malfunctions in the strategies (such as optimisation algorithms) used for computer interpretation of sketches.

References

[1] G. Johnson, M. D. Gross, J. Hong, and E. Y. Do, "Computational support for sketching in design: A review". Foundations and Trends in Human–Computer Interaction. Vol. 2, No. 1, 1–93, 2009.

[2] L. B. Kara, K. Shimada, "Sketch-based 3D-shape creation for industrial styling design", IEEE Computer Graphics and Applications, 27 (1), pp. 60-71, 2007.

[3] A. Andre, S. Saito and M. Nakajima,"CrossSketch: Freeform surface modeling with details", 4th Eurographics Workshop on Sketch-based interfaces and modeling, Riverside, California,August 02-03, 2007.

[4] S. Bae, R. Balakrishnan and K. Singh, "ILoveSketch: As-natural-as-possible sketching system for creating 3D curve models", ACM Symposium on User Inteface Software and Technology, Monterey, CA, USA, October 19-22, 2008

[5] A. Rivers, F. Durand, and T. Igarashi, "3D modeling with silhouettes", ACM Transactions on Graphics 29 (4), art. no. 109, 2010.

[6] S. Roth-Koch, and E. Westkaemper, "The implementation of a sketch-based virtual product development", Prod. Eng. Res. Devel. 4:175–183, 2010.

[7] K. Sugihara: "Machine interpretation of line drawing", MIT Press, 1986 (http://home.mims.meiji.ac.jp/~sugihara/openbooks/mild.pdf)

[8] Cooper M.: Line Drawing Interpretation, Springer, 2008.

[9] P. Company, A. Piquer, M. Contero, and F. Naya, "A survey on geometrical reconstruction as a core technology to sketch-based modeling", Computers & Graphics. 29 (6), 892-904, 2005.

[10] T. Marill, "Emulating the human interpretation of line-drawings as three-dimensional objects", Int. J. of Computer Vision; 6(2): 147-161, 1991.

[11] H. Lipson and M. Shpitalni, "Optimization-based reconstruction of a 3D object from a single freehand line drawing", Computer-Aided Design 28 (8), 651-663, 1996.

[12] P. Company, M. Contero, J. Conesa, and A. Piquer, "An optimisation-based reconstruction engine for 3D modelling by sketching", Computers & Graphics. 28 (6) 955-979, 2004.

[13] P. A. C. Varley and P. Company, "A new algorithm for finding faces in wireframes", Computer-Aided Design 42(4), 279-309, April 2010.

[14] Tian C., Masry M., Lipson H. "Physical sketching: Reconstruction and analysis of 3D objects from freehand sketches", Computer ‐ Aided Design, Vol. 41, pp. 147 ‐ 158, 2009.

[15] P. A. C. Varley, Automatic Creation of Boundary-Representation Models from Single Line Drawings, PhD Thesis, Cardiff University, 2003.

[16] R. Schmidt, A. Khan, K. Singh, and G. Kurtenbach, " Analytic drawing of 3D scaffolds" ACM Trans. Graph., Vol. 28, Nr. 5, 2009.

[17] D. N. Perkins, Cubic Corners, Quarterly Progress Report 89, 207-214, MIT Research Laboratory of Electronics, 1968.

[18] D. Perkins, Cubic Corners, Oblique Views of Pictures, the Perception of Line Drawings of Simple Space Forms. Geometry and the Perception of Pictures: Three Studies. Harvard Project Zero. Technical Report No. 5, 1971. http://www.eric.ed.gov/PDFS/ED114328.pdf

[19] D. Dori. and Wenyin Liu. "Sparse pixel vectorization: an algorithm and its performance evaluation," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.21, no.3, pp.202-215, Mar 1999.

[20] A. Bartolo, K. P. Camilleri, S. G. Fabri, J. C. Borg, "Line tracking algorithm for scribbled drawings," Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on , vol., no., pp.554-559, 12-14 March 2008.

[21] D. Dori; K. Tombre, "From engineering drawings to 3D CAD models: are we ready now?" Computer-Aided Design 27, 243-254, 1995.

[22] Y. Xuetao, P. Wonka, and A. Razdan, "Generating 3D building models from architectural drawings: A survey". IEEE Computer Graphics and Applications, 29 (1), 20-30, 2009.

[23] D. Ullman, S. Wood, and D. Craig, "The importance of drawing in the mechanical design process", Computers and Graphics 14(2):263-274, 1990.

[24] P. Varley and P. Company, "Automated sketching and engineering culture", in ed. B. Plimmer and T. Hammond, Proc. VL/HCC Workshop on Sketch tools for Diagramming, Herrsching am Ammersee, 15 Sep 2008, 83-92.

[25] P. Company, M. Contero, P. A. C. Varley, N. Aleixos, and F. Naya, "Computer-aided sketching as a tool to promote innovation in the new product development process", Computers in Industry. 60 (8), pp. 592–603, 2009

[26] L. Olsen, F.F. Samavati, M. Costa Sousa, J.A. Jorge, "Sketch-based modeling: A survey". Computers and Graphics 33 (1), 85-103, 2009.

[27] F. Kimura and H. Suzuki, "A CAD system for efficient product design based on design intent". CIRP Annals - Manufacturing Technology , 38 (1), 149-152, 1989.

[28] S. Yuan, L. Y. Tsui, and S. Jie, "Regularity selection for effective 3D objects reconstruction from a single line drawing", Pattern Recognition Letters 29 (10), 1486-1495, 2008.

[29] M. Li, F. C. Langbein, and R. R. Martin, "Detecting design intent in approximate CAD models using symmetry", Computer-Aided Design 42 (3) 183-201, 2010.

[30] S. E. Palmer, Vision Science, Photons to Phenomenology, Cambridge, MA: The MIT Press, 1999

[31] D. Hoffmann; Visual Intelligence. How we create what we see, New York: WW Norton & Company, 2000.

[32] S.J. Tate, G.E.M. Jared, "Recognising symmetry in solid models", Computer-aided Design 35 (7) 673-692, 2003

[33] A. Piquer, R.R. Martin, P. Company, "Skewed mirror symmetry for depth estimation in 3D line-drawings", GREC 2003, Lecture Notes in Computer Science 3088. 138-149, 2004.

[34] L.B. Kara, K. Shimada. "Sketch ‐ based 3D shape creation for industrial styling design". IEEE Computer Graphics and Applications, vol. 27(1), 60-71, 2007.

[35] P. Company, M. Contero, F. Naya and N. Aleixos "A study of usability of sketching tools aimed at supporting prescriptive sketches". Eurographics Symposium Proceedings. Sketch-Based Interfaces and Modeling (SBIM06). pp. 139-146. 2006

[36] R. Jarret and P. Su "Building tablet PC applications" Microsoft Press, 2002.

[37] L.B. Kara, K. Shimada, S.D. Marmalefsky "An evaluation of user experience with a sketch-based 3D modeling system" Computers & Graphics 31, 580–597, 2007.

[38] P. Company and P. A. C. Varley, "Operating modes in actual versus virtual paper-and-pencil design scenarios", Intelligent User Interfaces (IUI) Workshop on Sketch Recognition, Sanibel Island, Florida, 8th February 2009.

[39] R. Martin, P. Varley and H. Suzuki, "Perpendicularity as a key to interpreting line drawings of engineering objects", Proc. Digital Engineering Workshop: 5th Japan-Korea CAD/CAM Workshop, 115-120, 2005.

[40] E. S. Ferguson, "Engineering and the mind's eye", MIT Press, 1992.

[41] P. Company, N. Aleixos, F. Naya, P. A. C. Varley, M. Contero, and D. G. Fernandez-Pacheco, "A new sketch-based computer aided engineering pre-processor" Proceedings of the Sixth International Conference on Engineering Computational Technology. Civil comp Ltd. Paper-149, 2008.

[42] S. Murugappan and K. Ramani, "FEAsy: A sketch ‐ based interface integrating structural analysis in early design" ASME Conf. Proc. 2009, 743 (2009), DOI:10.1115/DETC2009 ‐ 87727.

[43] J. LaViola, "Advances in mathematical sketching: moving toward the paradigm's full potential", IEEE Computer Graphics and Applications, 27(1):38-48, 2007.

[44] Y. S. Suh, "Reconstructing 3D feature-based CAD models by recognizing extrusions from a single-view drawing", Proc. IDETC/CIE 2007.

[45] P. Company and P. A. C. Varley, "A method for reconstructing sketched polyhedral shapes with rounds and fillets", 10th International Symposium on Smart Graphics, Banff, Canada, 24th-26th June 2010.

[46] R. Juchmes, P. Leclerq, "A multi-agent system for the interpretation of architectural sketches", Sketch-Based Interfaces and Modeling. Eurographics Symposium Proceedings. SBM'04 pp. 53-61, 2004.

[47] D. G. Fernández-Pacheco, J. Conesa, N. Aleixos, P. Company, and M. Contero, "An agent-based paradigm for free-hand sketch recognition", Springer-Verlag Lecture Notes in Artificial Intelligence LNAI 5883, pp. 345-354, 2009

Data Unification on a Dataflow Visual Language for VJing

Atsutomo Kobayashi, Buntarou Shizuki and Jiro Tanaka
Department of Computer Science
University of Tsukuba
Tennodai 1-1-1, Tsukuba, Ibaraki, Japan
{*atsutomo,shizuki,jiro*}*@iplab.cs.tsukuba.ac.jp*

***Abstract*—We have been developing ImproV, a dataflow visual language system for VJing, which is a form of live performance involving simultaneous manipulation of imagery. The dataflow of ImproV represents a processing flow of video data and video effects that are familiar to performers. ImproV is designed to construct a combination of several video effects on the fly, enabling performers to create new video effects while VJing. We previously conducted a user study for the first prototype of this system. The results show that two data types of dataflow confuse performers and that connecting nodes is time consuming. Therefore, we unified the data types into a single type called `Video` to simplify dataflow and developed adjustment knobs to reduce the frequency of node connecting.**

The `Video` type, which is a sequence of images, is used to represent not only processing the target video but also parameters of video effects such as radius of blur and distance of translation. An adjustment knob on the input port is a new dataflow editing user interface for adjusting an effect parameter or opacity of a video flowing through the dataflow. In ImproV, all input ports on all nodes have an adjustment knob. These adjustment knobs allow performers to quickly fade video and adjust parameters.

***Keywords*-VJ; VJing; live video performance; DJ; DJing; image processing; end-user programming;**

I. Introduction

VJing is a form of live video performance and played mostly in night clubs. Performers of VJing are called VJs. Similar to DJing in a night club, VJing requires a VJ to keep playing videos improvisationally without interruption during the performance. The task includes choosing a video suitable to the mood and music, which dynamically changes during the performance, and seamlessly replacing the current video to the new one. Additionally, some visual effects (e.g., blur, and transformation) might be applied to the video before replacing.

To this end, a VJ uses one (or more in some cases) video mixer with multiple video inputs. When the VJ wants to change a video to another, he/she connects the new video source to an unused video input of the mixer while playing the original video with another video input, and seamlessly replaces the video with the new one by adjusting the knobs on the mixer and effectors. Lew [1] analyzed this task and found that it consisted of three steps:

STEP 1: Media retrieval
STEP 2: Preview and adjustment
STEP 3: Live manipulation

In traditional VJing, STEP 1 involves just choosing a pre-composed video the VJ prepared before the performance, and pre-defined video effects are chosen and applied in STEP 2. Then, the VJ starts showing the new video in STEP 3.

In contrast to such traditional VJing, we wanted to extend STEP 1 to enable a VJ to construct a new video effects structure by combining several video effects on the fly, giving him/her more chances to have the performance be truly improvisational. For this purpose, we developed ImproV, a datafow visual language system for VJing. With ImproV, the sole requirement for a VJ before a performance is to prepare simple videos. During the performance, the VJ can improvisationally create various effects from the prepared videos using ImproV's visual language.

We developed the first prototype of ImproV, whose language uses two data types, `Video` and `Value` [2]. To evaluate the first prototype, we conducted a user study with five professional VJs. For operational tests, the first author, who is a professional VJ, and an amateur VJ performed at actual musical events. Figure 1 shows one of the VJs using the first prototype at a jazz concert.

Figure 1. VJing with ImproV

As a result of the user study and operational tests, we found that while the dataflow diagram of ImproV is understandable for VJs, using two data types increases visual complexity of dataflow editing; thus, making the editing interface confusing. Moreover, connecting nodes is time consuming. To simplify dataflow editing, we unified the two data types into a single type called `Video`. To reduce the frequency of node connecting, we developed adjustment knobs located on each input port that enable VJs to quickly adjust the input values.

After a brief description of the first prototype of ImproV, we explain the user study of the prototype. We then describe the unification of the two data types into a single type called `Video` and give examples that characterize this type. Next, we explain the adjustment knobs. Finally, we discuss possibilities of applying the unification of data types and adjustment knobs to dataflow visual language systems for other domains.

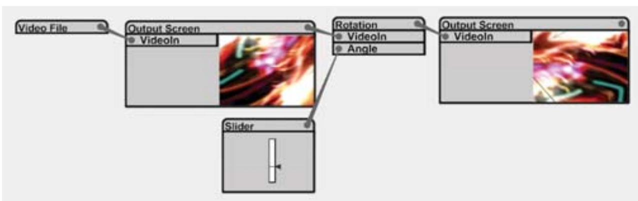


Figure 2. Dataflow of first prototype

II. First Prototype

ImproV is a video composing system. It provides a dataflow visual language, in which a node represents a video source, a video effect, or an output (or preview) screen. ImproV supports video files (including still images in Jpeg, GIF, and PNG), USB cameras, and audio inputs from sound cards. After data, video and audio, are captured, they are translated into images. All effects to the data are then processed in the graphics processing unit. Figure 2 shows an example of the dataflow composed on the first prototype of ImproV. All nodes have an output port on the top right and some have input port(s) listed on the left side.

In Figure 2, a video is rotated. The video file is imported using the `Video File` node, which reads its video file and repeatedly outputs the video stream frame-by-frame to its output port. The video is previewed using the `Preview Screen` node, which shows the content of the video on the dataflow editor, rotated using the `Rotation` node, and the rotation angle is specified using the `Slider` node. The `Angle` input port on the `Rotation` node accepts the `Value` type, which is a sequence of scalar values, and the other input port accepts the `Video` type, which is a sequence of images. Finally, the result of the `Rotation` node is output using the `Output Screen` node, which represents the display for the audience.

To create a node, a VJ chooses a node type from the GUI menu. Dragging from an output port to an input port connects the nodes, i.e., the action creates an edge between the nodes. Immediately after they are connected, the data begin to flow from the output port to the input port. In ImproV, output can fork to multiple inputs, and multiple outputs can be merged through mixer nodes corresponding to a blending mode, for example, alpha blending or addition blending.

III. User Study for First Prototype

We have conducted a user study for the first prototype of ImproV. The purpose of the user study was to confirm that the dataflow diagram is understandable for VJs, or those who have knowledge of video authoring, and to evaluate the operability of ImproV.

Five professional VJs (with experience of paid VJing) who all have knowledge of Adobe After Effects (AE) participated. The participants were first interviewed about their experience in VJing and skills with video authoring tools. Then, they participated in Experiment 1. After Experiment 1, they were told about ImproV and freely used ImproV for practice. Then, they participated in Experiment 2. Finally, they filled in a questionnaire.

Experiment 1

In Experiment 1, the participants were shown a dataflow diagram of ImproV and asked what the dataflow diagram expressed. Since Experiment 1 was conducted to confirm that a dataflow diagram is understandable for VJs, no information about ImproV was provided to the participants. To make the answers accurate, the participants were asked to construct a corresponding video processing structure using AE. Figure 3 is the dataflow diagram from Experiment 1.

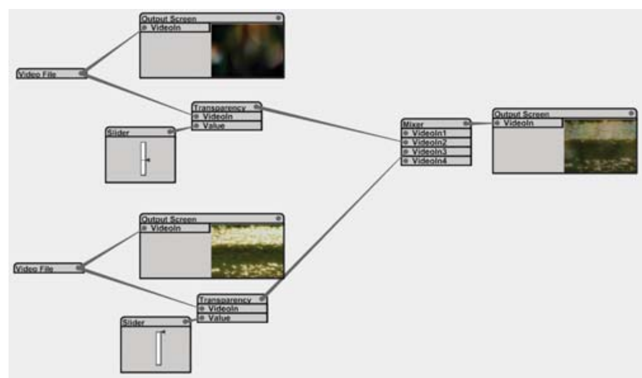


Figure 3. Dataflow diagram from Experiment 1

Experiment 2

In Experiment 2, the participants performed the following three tasks in both ImproV and AE:

- Task 1 Apply blur effect to the video displayed in the main output.
- Task 2 Merge two paths of video streams and switch them.
- Task 3 Apply blur effect to a part of the video displayed in the main output.

Task 1 is the simplest task that often used in VJing for applying an effect. Task 2 is the task that most often used in VJing for switching two videos. Task 3 is a complex task for creating a new effect for which Improve was developed.

The participants were able to ask questions about operating ImproV during the tasks. The completion times for the tasks were measured to compare the operability of ImproV and AE. The participants performing the tasks were recorded for later analysis.

Results

In Experiment 1, all participants answered correctly. This result indicates that the dataflow diagram of ImproV is understandable for VJs.

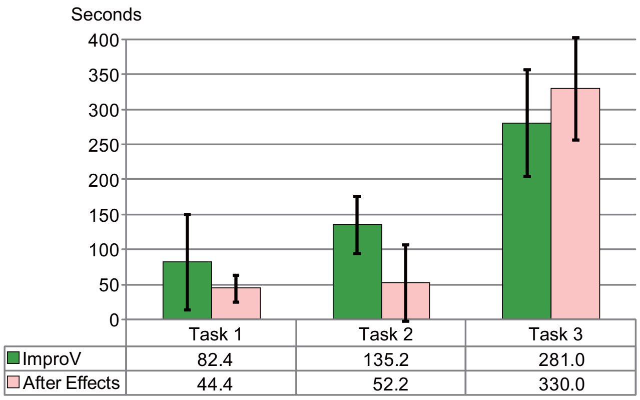


Figure 4. Result of Experiment 2

Figure 4 shows the completion time of each task using the two systems in Experiment 2. We performed an analysis of variance to examine the difference between ImproV and AE. There was a significant difference between ImproV and AE in Task 2 ($F = 7.169, p < 0.05$), and there were no significant differences in Tasks 1 and 3. The participants completed Task 3 slightly faster with ImproV than with AE. However, it was not significant. These results indicate that ImproV has near equivalent operability to AE in complex tasks, but exhibits problems in simple tasks. Therefore, we analyzed the video we recorded during the tasks and found that connecting nodes, dragging from output port to input port, took too much time.

The comments from the participants on the questionnaire also support the above results. The participants gave many positive comments regarding understandability, e.g., “It is intuitive with basic knowledge of video creation” and “The nodes are easy to understand”. Most of the negative comments mentioned the difficulties in connecting nodes, e.g., “I was having trouble connecting when I inserted an effect” and “I want sliders to be pre-attached to each new node”. The comments regarding complexity of the dataflow editor of ImproV were, e.g., “The dataflow becomes too complex while VJing” and “Managing the two data types is difficult”.

We analyzed these results and decided to unify the data types and to address the connecting problem.

IV. Video Type

From the results of the user study, we unified the data types into a single type called `Video`, which is a sequence of images. The `Video` type provides simple yet flexible ways to specify parameters of effects, still enabling represents a video as the processing subject of the dataflow.

In Figure 5, the video with triangles in the upper left is rotated at the center of the frame by using the `Rotation` node and its brightness is changed by using the `Brightness` node. Note that the rotation angle and brightness to be changed are designated by the *gray scale video* input to the `ValueIn` of both the `Rotation` and `Brightness` nodes. The resulting video from the `Rotation` and `Brightness` nodes is more rotated and darkened where the gray scale video is brighter, respectively.

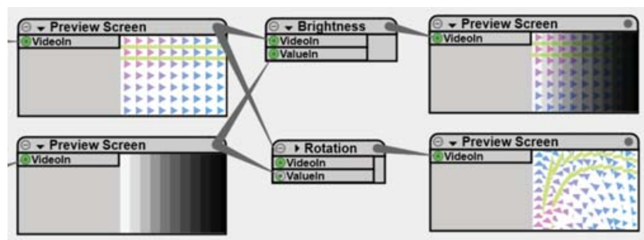


Figure 5. Setting parameters with `Video` type

The semantics is as follows. Each effect extracts a value for each pixel of each frame of the `Video`-type data and uses that value as the luminance value on the pixel position. More concretely, an effect is defined as f in Eq. (1). Out_{xy} is the pixel color of the resulting image at (x, y), n is the number of input ports, and V_n is the frame of nth input port.

$$Out_{xy} = f(x, y, V_1, V_2, V_3, \cdots, V_n) \tag{1}$$

For example, f of the `Rotation` node is defined using Eq. (2).

$$f(x, y, V_1, V_2) = V_{1_{RxRy}} \tag{2}$$

where

$$\begin{pmatrix} Rx \\ Ry \end{pmatrix} = \begin{pmatrix} \cos V_{2_{xy}} & -\sin V_{2_{xy}} \\ \sin V_{2_{xy}} & \cos V_{2_{xy}} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{3}$$

Giving effect parameters with `Video`-type values enables different values for different positions on the same frame to be set simultaneously. It also enables animation of parameter values according to the video that is input as the parameter.

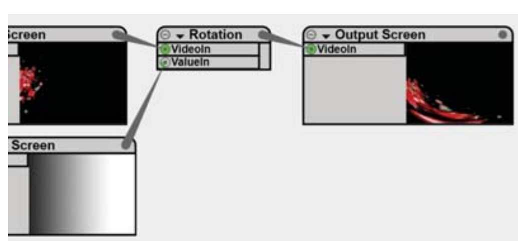


Figure 6. Dataflow of animation example

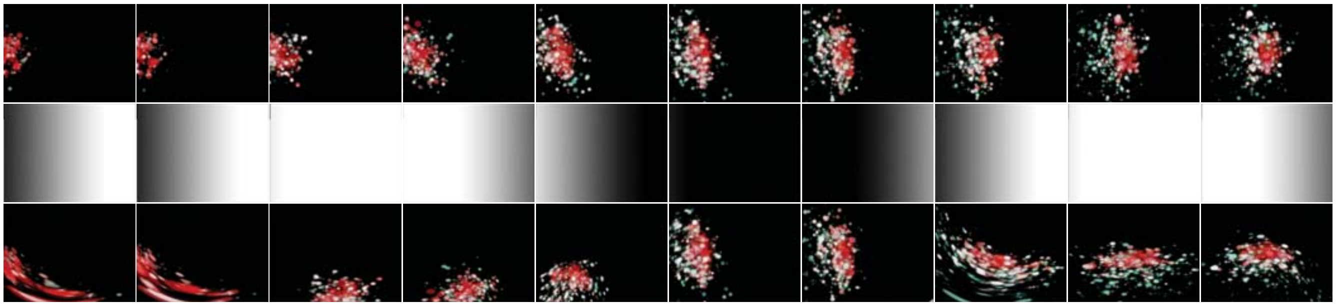


Figure 7. Example of animation

V. Examples using Video Type

In this section, we demonstrate three examples using the `Video` type for parameter animation, audio, and camera.

A. Animating Parameters

Designating a video effect parameter with a video can animate that parameter along with the input video's movement. Figure 7 shows ten frames of the input and output of the `Rotation` node shown in Figure 6. The top, middle, and bottom strips show the image sequence of `VideoIn`, `ValueIn`, and the output, respectively. In Figure 6, the rotation angle is designated by the gray scale video.

The resulting image sequence shows that the original video is modulated by the gray scale video. Although VJs still have to prepare some gray scale animated videos, those videos are versatile. Therefore, providing such videos as libraries is practical.

B. Audio Input

An `AudioIn` node captures audio signals from a sound card using PortAudio [11] and outputs a gray scale audio visualization. A VJ can designate the layout of the audio visualization by inputting `Video`-type data. An `AudioIn` node captures audio signals from the sound card then arranges the newer samples to where the pixels of the input video are brighter. Finally, it outputs the gray scale image according to the audio level.

For example, an `AudioIn` node outputs the waveform as vertical streaks since the horizontal gray scale video is input to the `AudioIn` node, as illustrated in Figure 8. Translating the horizontal color bars' y-axes according to the vertical streaks results in visualization of the waveform.

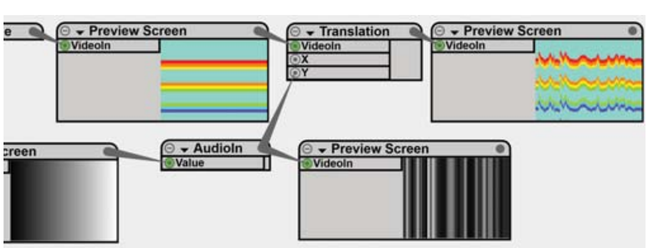


Figure 8. Example of audio visualization

Figure 9 shows a more complex example of audio visualization. By inputting circle gradation, an `AudioIn` node arranges the samples from center to outside. The output image from the `AudioIn` node results in ring streaks. The `Blur` node smoothes the streaks. The `Scale` node changes the scale of the video input in `VideoIn`. In Figure 9, the `Scale` node, by inputting the smoothed streaks to `ValueIn`, works as a ripple-like effect.

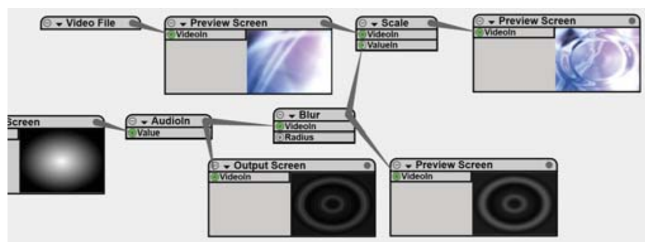


Figure 9. Example of complex audio visualization

C. Camera Input

With the `Video` type, a VJ can use a `Camera` node to control parameters. A `Camera` node captures the video from a USB camera. It was originally designed to be used to record an audience or DJs (or other performers). However, we found that using a captured video as another node's input parameter enables a VJ to use a USB camera as an instrument for controlling parameters.

The bottom left video of Figure 10 was captured using a USB camera. This video can be used as the parameter of the `Scale` node, which enlarges the original video where the parameter video is dark and shrinks the original video where the parameter video is bright. In the downer left of Figure 10, A VJ is rubbing the lens of the USB camera with his/her thumb. This produces a video similar to the animation of the gray scale video shown in Figure 6. In this way, the VJ can interactively control the area that he/she wants to enlarge by moving his/her thumb.

VI. Adjustment Knobs

A VJ can quickly set a parameter and adjust an input video value with an adjustment knob placed on each input port. Figure 11 shows what the adjustment knobs look like.

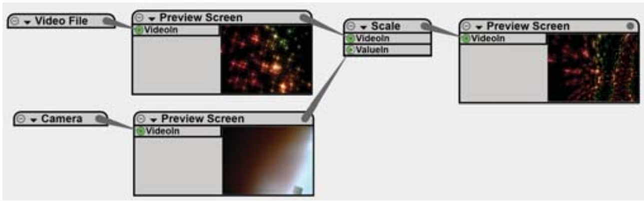


Figure 10. Controlling scale with USB camera

The Rotation node has two input ports, each of which is represented as double circles. The green fan between the two circles is the port's adjustment knob. The central angle of the fan indicates the value of the adjustment knob. A VJ can increase or decrease the value by dragging the mouse from the input port to up or down.

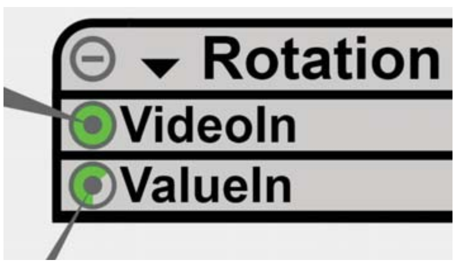


Figure 11. Adjustment knobs

An adjustment knob works differently depending on whether video input is given or not. In both cases, the value of the adjustment knob is used as opacity. If a video input is given to the input port, the port multiplies the value of its adjustment knob and the alpha channel value of the input video in each frame. In Figure 5, the rotation angle is restricted to 90° by designating the value of the adjustment knob to 25%.

If no video input is given to the input port, the port is treated as if a video of white frames is input. In this case, all pixels on the frame uniformly have the value of its adjustment knob. Figure 12 shows an example of such input ports. In this example, the Color node corrects the color of the input. The Saturation and Brightness are set around 50%, which means the same saturation and brightness as the original video. The Hue is set around 25%, which means the hue is 90°.

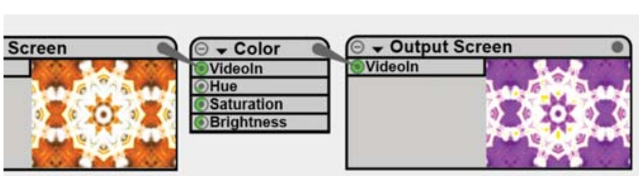


Figure 12. Correcting color with adjustment knobs

Adjustment knobs are designed as syntax sugar for the two idioms that an unconnected adjustment knob works as a Slider node which is a constant node in ImproV, and a connected adjustment knob works as a combination of Slider and Transparency nodes. Before the adjustment knob is developed, he/she has to create a Slider node and sometimes a Transparency node, and connect them. Creation and connection occurs many times when using nodes that have several input ports as parameters, such as the Color node. The same combination of Slider and Transparency nodes is also used for fading a video, which is often used in VJing.

To show the effectiveness of adjustment knobs, Figure 13 shows the corresponding dataflow of Figure 12 on an older version of ImproV. The meanings of these two dataflows are identical. The Slider node outputs a gray plane according to the value indicated by the slider on the node.

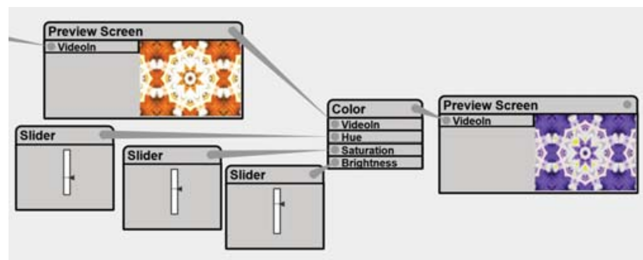


Figure 13. Correcting color with sliders in older version of ImproV

Figure 14 is a screenshot of the older version of ImproV. The transparency of the gray scale video is adjusted by a combination of Transparency and Slider nodes before the gray scale video is input to the Rotation node. The rotation angle is restricted by the Slider node as Figure 5.

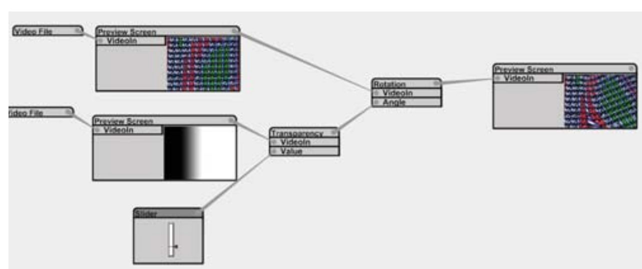


Figure 14. Adjusting video with Transparency node in older version of ImproV

The advantage of adjustment knobs is not only the reduction in the number of node creating and connecting operations, but is also the fluidness of the interaction flow, i.e., a VJ drags the mouse from an output port to an input port and then drags the mouse from the input port up or down.

VII. Discussion

In this section, we discuss the possibility of applying features of ImproV to other domains, i.e., audio and music.

It might be able to unify data types of a dataflow visual language for processing audio. An audio processing dataflow requires two data types, audio signal and parameter controlling signal. Both data types are sequences of linear values. Therefore, the abstract concept of data unification on a dataflow visual language can be applied to audio processing. However, note information in processing a musical score is difficult to represent by a sequence of linear values. We can define 0 as note-off and other values as note-on for the pitches of the values. However, such a definition may not suite a musician's intuition.

From the user interface point of view, adjustment knobs are also efficient for audio processing. Adjustment knobs in a dataflow visual language system for processing audio can be used to set linear values and adjust the value input to the input port. This reduces edge connection frequency and visual complexity. However, it depends on the characteristics of the data type. Adjustment knobs will work for many data types in a dataflow visual language system for any domain.

VIII. Related Work

There have been several dataflow visual language systems for supporting live performances. Editing dataflow during a performance is recognized as an established performance method in music. Bencina developed the audio processing system AudioMulch [3], which is based on a similar concept to that of ImproV in which the performer constructs effects on the fly. We are aiming at a similar abstraction level to AudioMulch with our system, in which professionals can understand and construct the dataflow without programming knowledge. The dataflow of AudioMulch also handles single data type, audio signal. However, parameter controlling is located in the user interface panel outside the dataflow editor. The `Video` type of our system can control parameters and be processed in the dataflow. It is also clear which knob corresponds to which parameter.

Dataflow visual language systems have been developed for VJing. Müeller et al. developed Soundium [5], a dataflow visual language system for VJing and explain the VJing method using Soundium [6]. Soundium has large function libraries and several data types which make dataflow too complicated. Parameter controlling is located outside the dataflow editor, as in AudioMulch.

ReacTable [4] by Jorda et al. uses tangible objects on a touch panel to construct an audio synthesizer's structure on the fly. VPlay [7] by Taylor et al. is a dataflow visual language system designed for multi-touch panels. Both systems have similar dataflow editing methods that bringing two nodes close together connects the nodes, and turning a node changes a parameter of the node. However, these editing methods sometimes result in an unintended dataflow structure. Moving a node can connect or disconnect edges.

Several systems, such as SPATIAL POEM [8], Rhythmism [9], and video-organ [10] for controlling the visual attributes in VJing have been developed. While our study mainly focused on STEP 1 of Lew's analysis, these studies addressed STEP 2 and can be used together with our system. Moreover, the `Video` type of our system enables VJs to use camera input to control parameters on the fly. This makes a camera a useful new instrument, which shows the flexibility of our system.

IX. Conclusions

ImproV is a dataflow visual language for VJing. We have conducted a user study for the first prototype of ImproV and found that two data types of dataflow confuse VJs and that connecting nodes is time consuming. To address these problems, we unified the data types into a single type called `Video` and developed adjustment knobs. This unification simplifies dataflow and makes various applications such as parameter animation, audio visualizing, and using cameras as controllers possible. The adjustment knobs enable quick interaction by reducing the frequency of constant node creation and connection. Dataflow is also easy to understand due to the reduction in visual complexity.

References

[1] M. Lew, "Live Cinema: designing an instrument for cinema editing as a live performance," in *NIME '04*, pp. 144–149.

[2] A. Kobayashi, B. Shizuki, and J. Tanaka, "ImproV: A system for improvisational construction of video processing flow," in *HCI International 2009 Part IV*, pp. 534–542.

[3] R. Bencina, "Oasis Rose the composition - real-time DSP with AudioMulch," in *Proceedings of the Australasian Computer Music Conference*, 1998, pp. 85–92.

[4] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner, "The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces," in *TEI '07*, pp. 139–146.

[5] P. Müeller, S. M. Arisona, S. Schubiger-Banz, and M. Specht, "Interactive media and design editing for live visuals applications," in *GRAPP '06*, pp. 232–242.

[6] P. Müeller, "Live visuals tutorial: part III," in *SIGGRAPH '07 courses*, pp. 127–151.

[7] S. Taylor, S. Izadi, D. Kirk, R. Harper, and A. G. Mendoza, "Turning the tables: an interactive surface for VJing," in *CHI '09*, pp. 1251–1254.

[8] J. Choi and S. H. Hong, "SPATIAL POEM: A new type of experimental visual interaction in 3D virtual environment," in *APCHI '08*, pp. 167–174.

[9] S. D. Tokuhisa, Y. Iwata, and M. Inakage, "Rhythmism: a VJ performance system with maracas based devices," in *ACE '07*, pp. 204–207.

[10] B. Bongers and Y. Harris, "A structured instrument design approach: the Video-Organ," in *NIME '02*, pp. 1–6.

[11] R. Bencina and P. Burk, "PortAudio - an open source cross platform audio API," in *ICMC '01*, pp. 263–266.

Sketch-Based Image Editing Using Clothoid Curves

Gunay Orbay, Mehmet Ersin Yumer, Levent Burak Kara*
Mechanical Engineering Department
Carnegie Mellon University
Pittsburgh, PA 15213
Email: *lkara@cmu.edu

***Abstract*—We present a sketch-based image editing technique that allows users to modify an image using a set of strokes drawn on the image. The proposed approach uses clothoid curves to beautify the sketched deformation curves. A deformation diffusion algorithm then spatially distributes the user specified deformations through out the image to produce smooth transformations from the original image to the resulting image. We demonstrate the technique on a variety of image types including photo-realistic images, real product images, and sketches.**

I. Introduction

Image editing tools are becoming increasingly more ubiquitous with the wide availability of a variety photo editing software [1]. However, such tools remain difficult to master due to the complex nature of their user interaction methods and the parameters that need to be controlled to achieve the desired effects. While many projects no doubt require advanced skills, experience, and appropriate task planning, even relatively simple modifications become difficult to implement with such tools.

This paper presents a new image editing method that enables users to apply desired image modifications through simple sketched strokes. The proposed method allows input images to be smoothly deformed into new ones with the specification of only a few strokes. As such, it advances existing image editing methods that uses fixed deformation masks, curve-based deformation handles or area-based deformation lattices. The proposed approach allows complex deformations to be specified through an intuitive interface, while ensuring globally smooth image deformations.

At the heart of our approach is an aesthetic curve construction, followed by a deformation diffusion algorithm. On the input image, users sketch the curves they want to deform, as well as the new shapes of these curves. They can also sketch the curves that are to be precisely preserved. To suppress the undesirable artifacts that may arise due to the free-form nature of our sketch-editing method, the user input is beautified into *clothoid* curve segments through automatic curvature profile processing. With this input, a spatial deformation vector field is calculated and subsequently applied to the original image that results in the final image.

Figure 1 shows the usage of our system. Shape variations can be intuitively and fluidly explored on different kinds of images including photo-realitic images, real product images, and sketches.

II. Related Work

The main focus in image editing have been in cut-and-paste type operations and image enhancing. Elder *et al.* [2] presented a contour based image editing technique where modifications are applied directly in the image contour domain. These modifications are then transferred to the final image using edge based image reconstruction [3]. This approach enables robust cut-and-paste type operations. Saund *et al.* [4] have introduced a perceptually-supported editing tool where they combine a variety of previously studied image editing methods [5], [6] deployed in an interactive environment. In an earlier work [6], the authors studied decomposing simple images into geometric primitive based shapes for enabling simple shape based editing. In 2002, Seitz *et al.* [7] introduced a set of image editing techniques which enable a sequence of images to be modified through operations applied to one of the images in the sequence. Their method makes use of the plenoptic function through which the modifications are propagated in a physically-consistent way. Perez *et al.* [8] presented a set of image editing operations based on second order discrete Poisson equations in image domain. Their approach enables seamless region mapping between images and modification of original illumination, color and texture in arbitrary regions of a given image, our method focuses on nonlinear warping of a single image through user defined strokes. Closely related to our method is the work by McCrae and Singh [9], where a sketching interface is described that replaces the input strokes with clothoid curve segments. This approach enables strokes to be beautified into smoothly varying curve segments, but does not aim to edit an underlying image or geometric space.

Our approach differs from the above works in a number of ways. Our target is rapid image editing where users' strokes can be used to derive a smooth deformation field on the image. Specifically, the curves dictated by the user are interpreted as clothoid curves that allow the regions corresponding to those curves to attain aesthetic curvatures. The remaining regions of the image are then modified based on the diffusion of the deformation vectors computed between the original image and the users' strokes. This approach provides close control over image editing, while allowing arbitrarily many constraints and target shapes to be specified by the user.

III. Overview and User Interaction

Our system takes as input an image and a set of strokes drawn on the image. Figure 1 shows the image editing process.

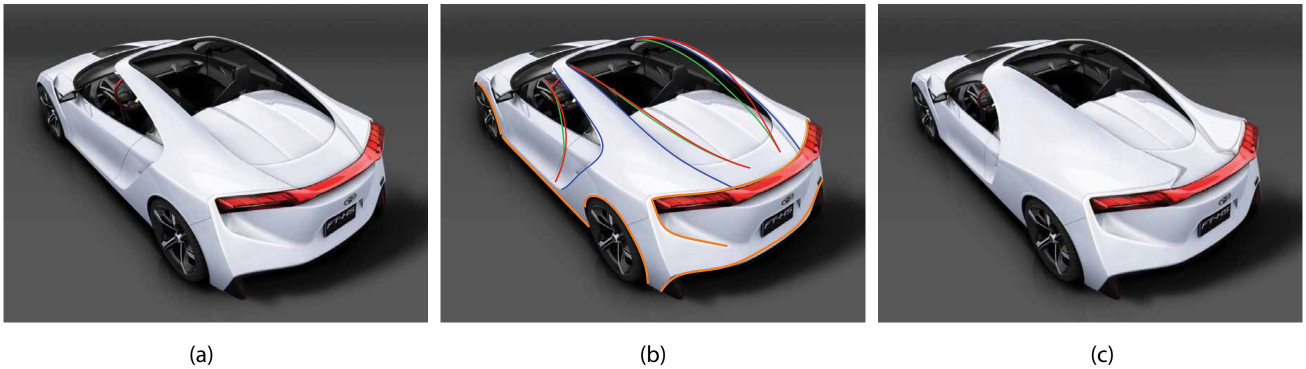


Fig. 1. (a) A car image, (b) The set of curves defining the desired modifications, (c) Final result.

Users may draw three types of strokes:

Anchor Strokes: Orange color. These strokes indicate regions of the image that remain unchanged. Any potential deformation ceases at these strokes.

Source Strokes: Blue color. These strokes demarcate curves in the image that will undergo user-specified deformations. Input source strokes are first modified under local image gradients that precisely aligns them with the salient image features.

Target Strokes: Green color. For each source stroke, a target stroke indicates the new desired shape. Each target stroke is automatically beautified into a clothoid curve using a functional regression on the stroke's original curvature profile.

Once processed, each kind of stroke is internally represented as a corresponding *curve* of the same type. Given the anchor, source, and target curves, a smooth vector deformation field is calculated on the image. The computed deformation field is then applied to the image, resulting in the modified image. The key property of the deformation is that the image attains the precise curve shape dictated by the target curves, while the remainder of the image deforms in visually smooth fashion.

IV. Preprocessing the Input

For image deformations to be visually meaningful, our approach seeks to satisfy two requirements. First, the source strokes must accurately identify the salient curves in the image. Second, the target strokes must be beautified into aesthetically pleasing curves to eliminate artifacts that may arise from a literal interpretation of the raw strokes.

A. Source Curve Construction

We use active contours [10] to align the raw source strokes with the intended image curves. During this process, an edge map is extracted from the image using Canny Edge Detection [11], and the curves are pulled toward these identified edges. The process is initialized by anchoring the end points of the source curve to the nearest stable points on the edge map, and using the regular active contours method on the resulting configuration as given by Kass *et al.* [10].

B. Target Curve Beautification

In this step, the target strokes are beautified into smooth curves through a curvature profile modification. Previous studies have proposed a set of differential properties and related functionals for generating aesthetic geometries [12], [13], [9]. Nearly all arguments regarding aesthetic properties involve an analysis of the curvature of the curves or surfaces. In this work, we use *clothoid curves* [9] for beautifying the target curves as they readily produce G^2 continuous curves with simply varying curvature profiles.

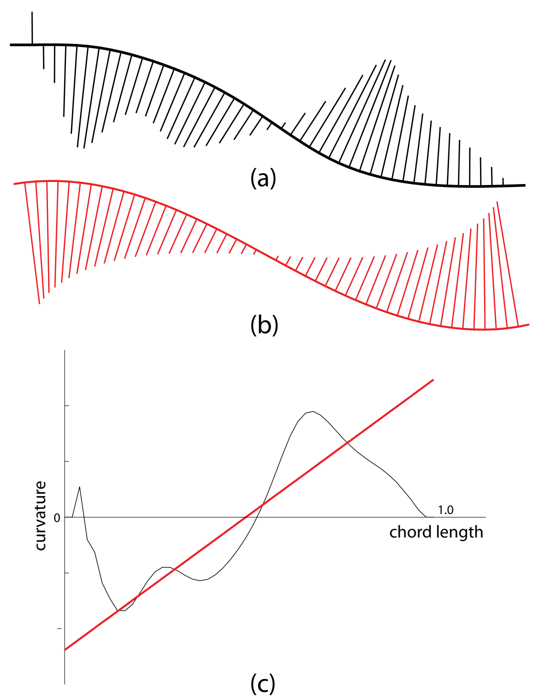


Fig. 4. Curvatures of target curves are rectified into clothoid profiles using linear regression. (a) The original target curve, (b) the rectified curve, (c) the linear regression line on the original curvature profile of the target curve.

The beautification process begins with the calculation of the curvature profile of a target stroke. Since clothoids' curvature profiles are linear with respect to the curve length parameter,

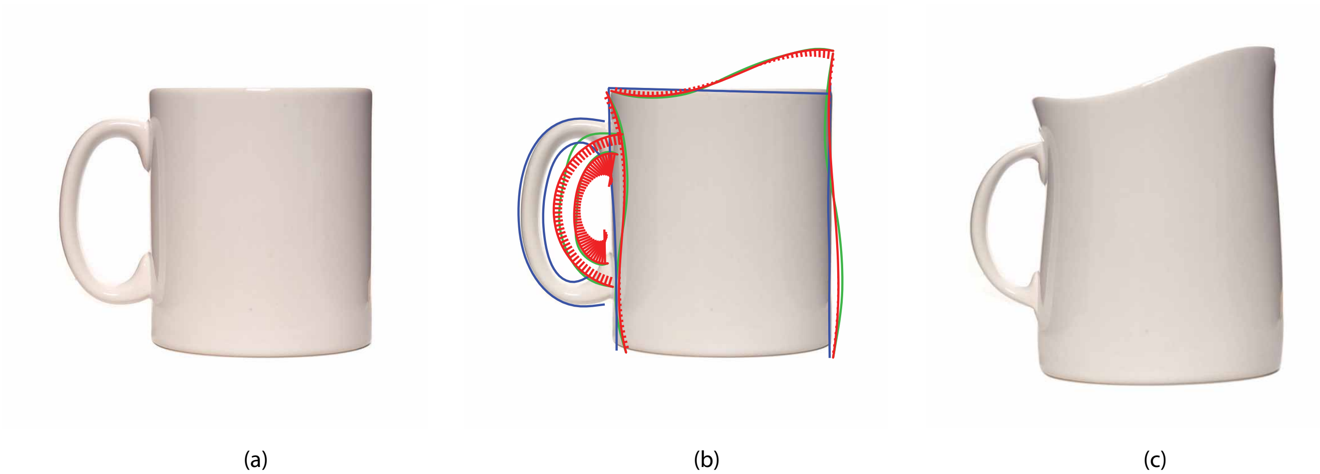


Fig. 2. (a) A mug image. (b) The set of source and target curves that define the intended modifications. Note that there are no anchor curves in this case, (c) Final result.

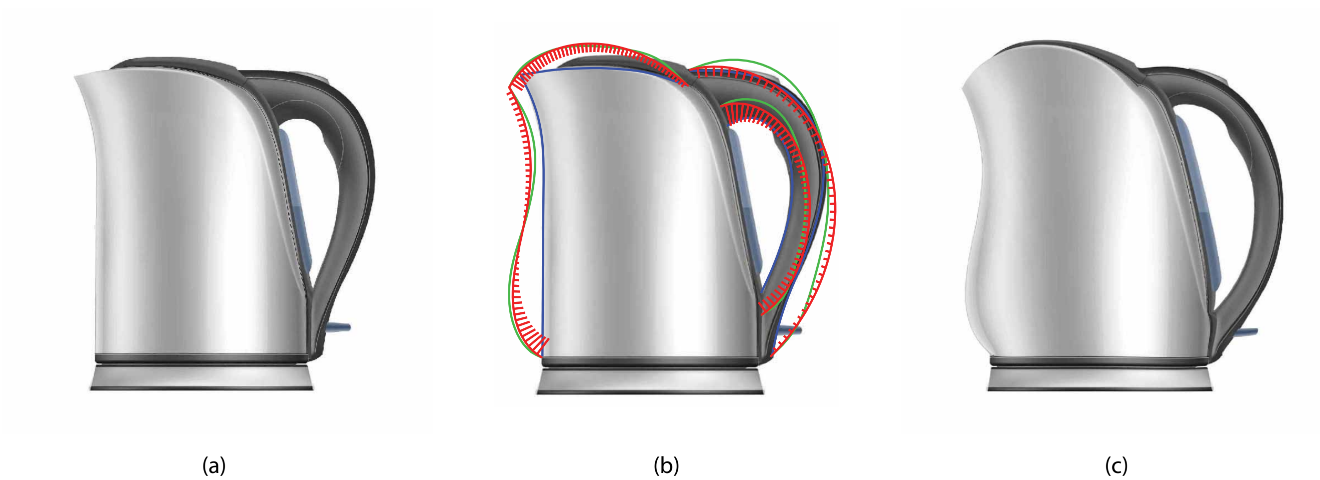


Fig. 3. (a) A kettle image. (b) The set of curves that define the intended modifications. (c) Resulting modification.

a linear regression to the original curvature profile produces a clothoid that best represents the original stroke. Figure 4(c) illustrates the idea.

Once the modified curvature profile is determined, the beautified curve is computed through double integration of the curvature profile. The first two derivatives of the curve with respect to the chord length define the tangent and the curvature as follows:

$$\vec{t}(s) = \frac{d\vec{C}(s)}{ds} \quad (1)$$

$$\vec{\kappa}(s) = \frac{d^2\vec{C}(s)}{ds^2} = \frac{d\vec{t}(s)}{ds} = \bar{\kappa}(s)\vec{n}(s) \quad (2)$$

where $\vec{C}(s)$ is the curve parametrized with the chord length, $\vec{t}(s)$ is its unit length tangent, $\vec{\kappa}(s)$ is its curvature, $\bar{\kappa}(s)$ is the magnitude of the curvature, and $\vec{n}(s)$ is the unit normal orthogonal to the tangent. Starting from an arbitrary position ($\vec{C}(0) = (0,0)$) and an arbitrary tangent ($\vec{t}(0) = (0,1)$), the tangent and the position is computed iteratively using Euler's method as:

$$\Delta\vec{t}(s) = \bar{\kappa}(s)\vec{n}(s) \quad (3)$$

$$\Delta\vec{C}(s) = \vec{t}(s)\Delta s \quad (4)$$

where Δs is the step in chord length. The resulting integration produces a curve with the desired shape and size, with its location undetermined. The curve is then situated between the start and end points of the original stroke through rigid body translations and rotations. In our experiments, this approach produced curves close to the input curves while successfully imposing linear curvature profiles.

Note that for each target stroke, a single clothoid is fit using a linear regression on the curvature profile. However, since the reconstruction through integration is insensitive to the underlying curvature functional, this approach can be trivially extended to a more precise beautification involving multiple piecewise continuous and linear curvature segments similar to the way described in [9], or other aesthetic curvature functionals that are of interest.

V. Curve-Based Image Deformation

The input to the image deformation step is a set of pre-processed source and target curves, and the set of anchor curves. This step calculates a smooth spatial deformation that transforms each source curve into the corresponding target curve, while preserving the anchor curves. For this purpose, we use an isotropic vector diffusion method to spatially diffuse the deformation vectors between the source and target curve pairs.

Although there are arbitrarily many vector fields that can produce the desired deformation, in this work we use a formulation similar to the gradient vector flow field (GVF) method [14]. This approach suitably diffuses the initial deformation vectors computed between each source-target curve pair, thus producing deformations that vary smoothly through the image space.

The GVF is defined as the continuous 2D vector field $\vec{V}(x,y) = [u(x,y), v(x,y)]$ that minimizes the following energy functional [14]:

$$\epsilon = \iint \underbrace{\mu(u_x^2 + u_y^2 + v_x^2 + v_y^2)}_{\text{Smoothness}} + \underbrace{\|\nabla f\|^2 \|\vec{V} - \nabla f\|^2}_{\text{Pointing objects}} dxdy \tag{5}$$

where $f(x,y)$ is a 2D scalar field, ∇ is the gradient operator in Cartesian coordinates, $\vec{u}(x,y)$ and $\vec{v}(x,y)$ are the x and y components of the vector field at point (x,y) while the subscripts denote partial derivatives. μ is the coefficient which controls the amount of diffusion. In this energy functional, the first term forces the local smoothness of the calculated vector field while the second term forces $\vec{V}(x,y)$ to converge to the original deformation vector in regions of high deformation. The vector field $\vec{V}(x,y)$ minimizing this energy functional can be calculated using calculus of variation by solving the following two decoupled Euler equations:

$$\begin{aligned} \mu\nabla^2 u - (u - f_x)(f_x^2 + f_y^2) = 0 \\ \mu\nabla^2 v - (v - f_y)(f_x^2 + f_y^2) = 0 \end{aligned} \tag{6}$$

where the Laplacian operator is defined as:

$$\nabla^2 u(x,y) = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \tag{7}$$

The two discretized equations can be solved independently to produce the vector field components in x and y directions. In this application, the discretization of the domain naturally comes from the pixelated image. In our implementation, we use an iterative multigrid solver that enables an efficient computation of the solution to the differential equations. We note, however, any conventional differential equation solver could be similarly adopted.

The boundary conditions to the above set of equations arise from the anchor curves (zero deformation), and the deformation vectors from each source curve to the corresponding target curves. These vectors are preserved throughout the computation. The remainder of the vector field is computed using the above two equations.

Prior to initializing the boundary conditions, we resample the target and source curves to have the same number of points along their paths. The sampling density is chosen such that the distance between adjacent sample points is always less than the pixel length. This ensures all pixels along the curves to attain a unique deformation vector. Moreover, a simple direction check is performed to map each end of the source curve to the intended end of the target curve, thereby eliminating sensitivity to drawing directions.

A problem arises if the computed deformation field is applied directly to the original image, as the deformation vectors do not guarantee a complete mapping from the original image to the yet-to-be computed deformed image. To overcome this difficulty, we utilize deformation vectors that originate from the target curves and point toward the source curves. The subsequent diffusion is then applied in this new configuration. This allows, for every pixel in the deformed image to point to a real-valued location in the original image (through the deformation vector) and attain a value as an interpolation from the original image. For smoothness, we use a bicubic spline interpolation for this purpose.

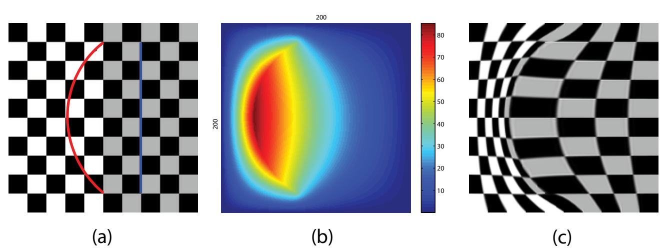


Fig. 5. Calculation of a spatial transformation field from source and target curves. (a) The displacement vectors from the source curves (blue) to the target curves (red) are diffused. (b) The calculated spatial deformation field in x direction. (c) The deformed image.

VI. Implementation and Examples

The proposed method is implemented in MATLAB. The user sketches the input curves on the original image. We are using a multigrid solver for the deformation field calculation. For a typical image of size 1024×768, the computation of the deformation field and its subsequent application to produce the deformed image takes on the order of 10 seconds on dual core 2.5 GHz laptop with 3 GB of memory. However, especially when deformations are large, it might be desirable to apply the deformation in successive smaller steps, which will proportionally increase the processing time.

In Figure 1, intended modifications are sketched in the form of anchor, source and target curves. The resulting deformation is shown in Figure 1(c). Note that the deformations cease at the regions marked by anchor curves. Figure 2 and 3 present similar modifications for a mug and a kettle, respectively. Note that there are discrepancies between the target strokes and their

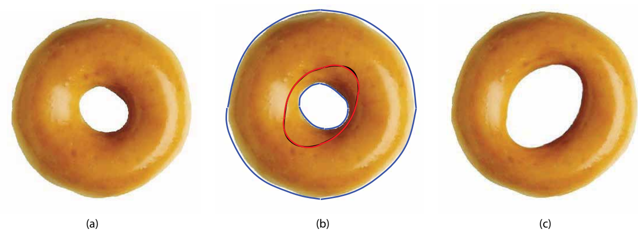


Fig. 6. Modifying the interior of a donut.

beautified versions. This arises as a result of using a single clothoid during curvature beautification. Such discrepancies can be trivially alleviated by regressing multiple clothoid pieces along the target curves instead of a single one. Figure 6 shows the editing an image with an interior loop. The loop is drawn in multiple strokes.

A sketched concept car is shown in Figure 7(a). This sketch is is first modified through a set of user drawn curves that define the modifications and constraints on the image (Figure 7(b)). The result of these modifications (Figure 7(c)) have gone through a second set of modifications (Figure 7(d)), resulting in the final image (Figure 7(e)).

VII. Conclusion

We present a sketch-based space deformation technique for aesthetic image editing. The proposed approach enables a stylus-based, fluid edits to an image, making the approach suitable for exploring variations to an underlying image. Our method uses a target curve beautification method using clothoid curves that ensures the deformation to be smooth and visually pleasing. Our current studies indicate that the approach works well for a variety of different image types. Future work will include extending the curvature-based stroke beautification to multiple piecewise continuous aesthetic segments rather than a single curve.

References

[1] Photoshop. Adobe Systems Incorporated. [Online]. Available: http://www.adobe.com/products/photoshop.html

[2] J. Elder and R. Goldberg, "Image editing in the contour domain," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 291–296, Mar. 2001.

[3] J. Elder and S. Zucker, "Space scale localization, blur, and contour-based image coding," in *IEEE CVPR*, 1996, pp. 27–34.

[4] E. Saund, D. Fleet, D. Larner, and J. Mahoney, "Perceptually-supported image editing of text and graphics," *ACM Transactions on Graphics*, vol. 23, no. 3, p. 728, Aug. 2004.

[5] S. Bagley and G. Kopec, "Editing Images of text," *Communications of ACM*, vol. 37, no. 12, pp. 63–72, 2001.

[6] E. Saund and T. P. Moran, "A perceptually-supported sketch editor," *Proceedings of the 7th annual ACM symposium on User interface software and technology - UIST '94*, pp. 175–184, 1994.

[7] S. Seitz and K. Kutulakos, "Plenoptic image editing," in *Sixth International Conference on Computer Vision IEEE*. Narosa Publishing House, 2002, pp. 17–24.

[8] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM SIGGRAPH 2003 Papers on - SIGGRAPH '03*, p. 313, 2003.

[9] J. McCrae and K. Singh, "Sketching piecewise clothoid curves," *Computers & Graphics*, vol. 33, no. 4, pp. 452–461, 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0097849309000843

[10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, Jan. 1988. [Online]. Available: http://www.springerlink.com/index/10.1007/BF00133570

[11] P. Bao, L. Zhang, and X. Wu, "Canny edge detection enhancement by scale multiplication," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 9, pp. 1485–1490, 2005. [Online]. Available: http://www.computer.org/portal/web/csdl/doi/10.1109/tpami.2005.173

[12] G. Farin, *Curves and Surfaces for CAGD*. Morgan-Kaufmann, 2002.

[13] N. Sapidis, *Designing Fair Curves and Surfaces*. SIAM, 1994.

[14] C. Xu and J. Prince, "Snakes, shapes, and gradient vector flow," *Image Processing, IEEE Transactions on*, vol. 7, no. 3, pp. 359 –369, Mar. 1998.

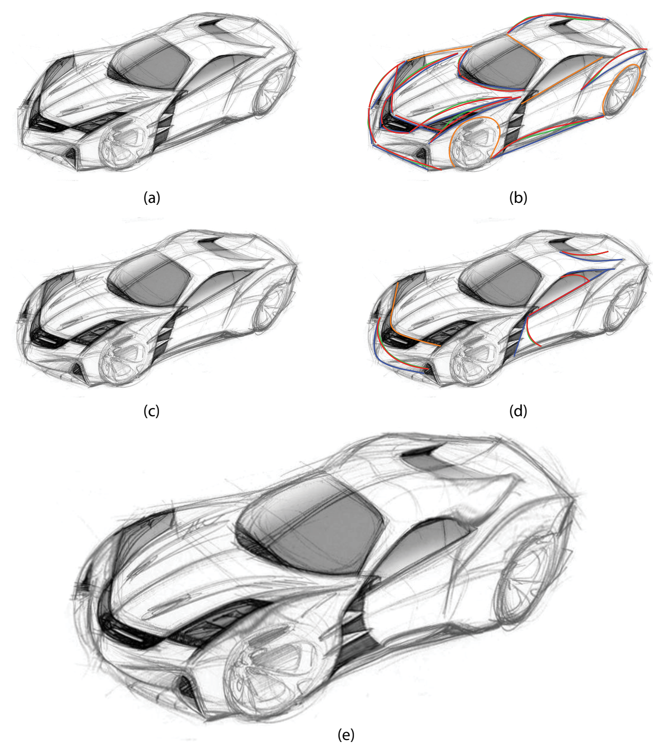


Fig. 7. The original car sketch (a), set of curves that define the first set of intended modifications (b), intermediate result (c), set of curves that define the second set of intended modifications (d), final result(e).

Recognising Sketches of Euler Diagrams Augmented with Graphs

Gem Stapleton Aidan Delaney Peter Rodgers Beryl Plimmer
University of Brighton University of Kent University of Auckland
{g.e.stapleton,a.j.delaney}@brighton.ac.uk p.j.rodgers@kent.ac.uk bpli001@ec.auckland.ac.nz

Abstract

Euler diagrams form the basis of many visual languages. Such languages are formed by augmenting Euler diagrams with graphs or shading. However, tool support for creating augmented Euler diagrams is generally limited to generic diagram editing software using mouse and keyboard interaction. A more natural and convenient mode of entry is via a sketching interface which facilitates greater cognitive focus on the task of diagram creation. Previous work has developed sketching interfaces for Euler diagrams. This paper presents the first sketch tool for Euler diagrams augmented with graphs and shading. The tool allows the creation of sketches of these diagrams via pen-based interaction. To effect the recognition process, we define heuristics for classifying strokes as either curve, node, edge, or shading. To evaluate our recognition engine, we asked 10 participants to each sketch six augmented Euler diagrams. Using the results of the study, we fine-tuned the recognizer, achieving a statistically significant improvement.

1. Introduction

The effective use of visual languages often relies on software support. For example, in software engineering, class diagrams can be automatically generated from source code, allowing the programmer to visualize the structure of their implementation. From a different perspective, other software that supports the use of visual languages facilitates the creation of diagrams by users. For instance, standard diagram editors, such as Visio, allow users to draw diagrams using traditional point, click and drag mouse interaction.

Recently, we have seen the development of software that allows users to create diagrams using a stylus, or pen, on a touchscreen device. Sketching diagrams is advantageous in that it allows the user to focus on the actual diagram creation rather than the interface of the editing tools, and it is a useful problem solving and communications technique [4]. Sketch recognition software developed to date has focused on user interface design and graph oriented diagrams [7]. With respect to user interface design tools, the sketched items are largely independent of each other. In graph oriented diagrams, the spatial positioning of nodes and edges is not of semantic significance.

We have devised sketch recognition software for Euler diagrams [2, 15] where the spatial relationships between sketched items is fundamental to their semantics. This lays the foundations for the development of sketch recognition software for visual languages that extend them such as the Euler diagram augmented with a graph in figure 1; this is a variation on a diagram seen in [11] and represents information from a semantic network. Example notations that extend Euler diagrams include the monadic family of diagrammatic logics such as Swoboda and Allwein's Euler/Venn system [14], Shin's Venn-II system [13], spider diagrams by Howse et al. [5] and Gil et al. [6] and Minoshima et al.'s extended Euler diagram logic [9]. These diagrams typically augment Euler diagrams with graphs or shading, usually both. Furthermore, diagrammatic logics with greater expressive power add even more syntax, such as arrows, to make semantically rich statements. Examples include Kent's constraint diagrams [8] and Oliver et al.'s ontology diagrams [10].

We report on the development of sketch recognition software for Euler diagrams augmented with graphs and shading, which provides a basis for more sophisticated sketching software for the diagram types just given. Section 2 overviews our sketching software. The recognition component of our tool is detailed in section 3. Section 4 presents a user study used to evaluate and improve the recogniser. The software can be downloaded from [12].

2. SpiderSketch

The tool, called SpiderSketch, that we have developed allows users to sketch Euler diagrams augmented with graphs and shading and then recognises each pen stroke. Formally, *augmented Euler diagrams* comprise a finite set of closed curves drawn in the plane, a finite set of nodes, a finite set of edges whose end points are incident with nodes, and a set of minimal regions (formed by the curves) that are shaded. A

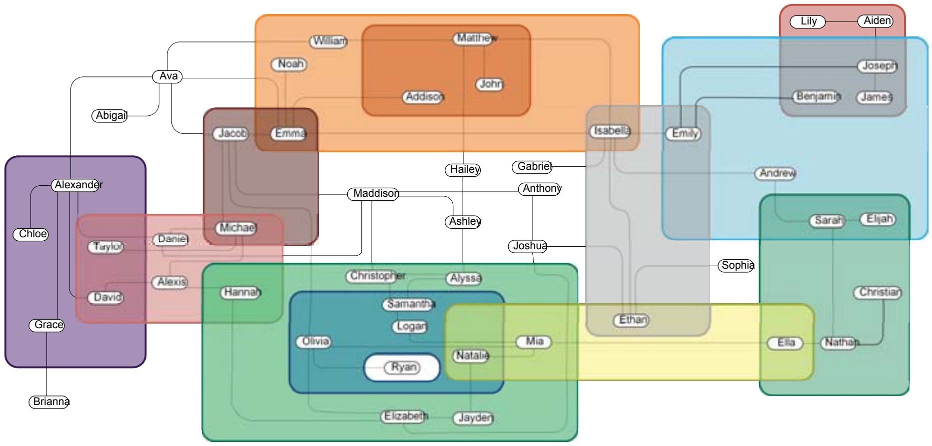


Figure 1. Visualizing semantic networks.

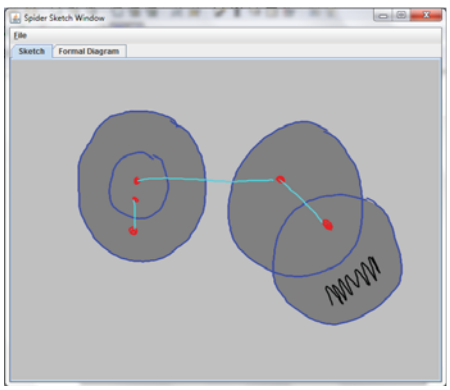


Figure 2. A screen-shot of SpiderSketch.

sketched diagram is a hand-drawn image that approximates an augmented Euler diagram.

In the sketching interface, users create sketches by drawing as on a piece of paper; figure 2 shows a screenshot. A stylus stroke is immediately rendered on the canvas. On completion of the stroke (stylus up event) it is passed to a recognizer that we describe in section 3. The recognizer result may be one of four classes: curve, node, edge, or shading. Each piece of syntax is immediately coloured to show the result of the recognizer. Curves are dark blue, nodes are red, edges are light blue, and shading is black. Online colouring allows one to readily check the association made by the software and, thus, have an opportunity to change the sketch if necessary. The sketch can also be saved.

3. The Recognizer

We have devised a number of heuristics that classify each stroke as one of the four types of syntax. The recogniser works online, classifying each stroke as it is created by the user. To design the heuristics, we identified features of each syntax type that distinguish it from the other syntax types.

1. Nodes are generally small and round, unlike the other pieces of syntax.
2. Shading is entered into the interface by a 'scribble' type action. This can be described as a stroke that is relatively long compared to the size of its bounding box and has many sharp turns. Potentially, this is similar to a node but we assume that shading takes up more space. Edges do not generally have sharp turns. Finally, curves are not particularly long compared to the size of their bounding box.
3. Edges are line segments that do not self-intersect. In addition, the end points are often far apart from one another, given the length of the line segment. By contrast, end points are generally close to each other for nodes or shading, which are 'densely packed' in to a space on the screen. In addition, curves should self-intersect since they are closed although users may not enter them that carefully; either way, the end points of a curve should be relatively close together.

The heuristics are described in full below and they are applied in a particular order. First, the recogniser decides whether the entered stroke, s, is a node. If s is not a node then the recongiser determines whether it is shading. Again, if s is not shading then it determines whether it is an edge. Finally, if s is not an edge then it is deemed to be a curve. Our heuristics are each reliant on a threshold value particular to it. For instance, we consider a stroke to be very small, and therefore a node, if the longest side of its bounding box is at most $T_{n,vs}$ pixels (n for node, vs for very small). Initially, we chose $T_{n,vs}$ to be 10 pixels. The threshold values can be adjusted to improve the recognition success rate. In full, the heuristics are as follows:

1. Any stroke that either (a) has a very small bounding box, or (b) is approximately square and has small bounding box is a node. By very small, we mean

$$BBlongestSide < T_{n,vs} \tag{1}$$

for some threshold value $T_{n,vs}$, where $BBlongestSide$ is the length of the longest side of the bounding box. By approximately square, we mean

$$\frac{BBshortestSide}{BBlongestSide} > T_{n,sq} \tag{2}$$

for some threshold value $T_{n,sq}$, where $BBshortestSide$ is the length of the shortest side of the bounding box. By small we mean

$$BBlongestSide < T_{n,s} \tag{3}$$

So, for a stroke, s, to be a node either (1) is true or both (2) and (3) are true.

2. Any stroke whose density is above a certain threshold, $T_{s,d}$, and contains at least $T_{s,t}$ sharp turns, typically called corners, is shading. We measure density by comparing the length of the stroke with the perimeter of its bounding box: the stroke must be significantly longer than the perimeter. In particular, we must have

$$\frac{\text{stroke length}}{\text{bounding box perimeter}} > T_{s,d}. \tag{4}$$

A corner is a single point, p, in the sketched line, l, where the angle, $\theta(p)$, formed by the two line segments on l whose end points are p satisfies

$$\theta(p) > T_{s,a} \tag{5}$$

for some threshold angle $T_{s,a}$. We count the number of corners and compare the number of them to a threshold value, $T_{s,t}$:

$$|\{p \in setP(l) : \theta(p) > T_{s,a}\}| > T_{s,t} \tag{6}$$

where $setP(l)$ is the set of non-end points on l. So, for a stroke, s, to be shading it must not be a node and both (4) and (6) must be true.

3. Any stroke whose end-points are far apart, given its length, and does not self-intersect is an edge. By far apart, we mean

$$\frac{\text{distance between end points}}{\text{length}} > T_e. \tag{7}$$

So, for a stroke, s, to be an edge it must not be a node or shading and (7) must be true.

Threshold	Value
$T_{n,vs}$	10 pixels
$T_{n,sq}$	0.55
$T_{n,s}$	15 pixels
$T_{s,d}$	0.5
$T_{s,a}$	60 degrees
$T_{s,t}$	1
T_e	0.2

Table 1. Initial threshold values.

Threshold	Value
$T_{n,vs}$	15pixels
$T_{n,sq}$	0.55
$T_{n,s}$	18 pixels
$T_{s,d}$	0.5
$T_{s,a}$	90 degrees
$T_{s,t}$	1
T_e	0.2

Table 2. Final threshold values.

4. Any other stroke is a curve.

The initial values we picked for these thresholds are in table 1. In section 4 we describe a study used to evaluate our recognition engine using these initial values. The results of the study were used to fine-tune the threshold values and the final values are shown in table 2. The process by which these final values were obtained is also described in section 4. To provide some insight into the recognition process, figures 3, 4, and 5 show a series of strokes, entered using the final threshold values, that show how the classification changes as properties of the strokes are altered.

4. Evaluating the Recogniser

We have conducted a study to evaluate the effectiveness of the recogniser component of our sketching tool. At the beginning of the study, each participant was given a short tutorial in how to use SpiderSketch. The researcher gave them a demonstration of the tool after which they were asked to draw some sketches, ensuring they were happy with how to

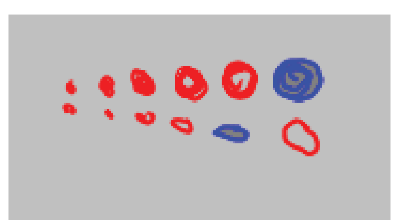


Figure 3. Recognising nodes.

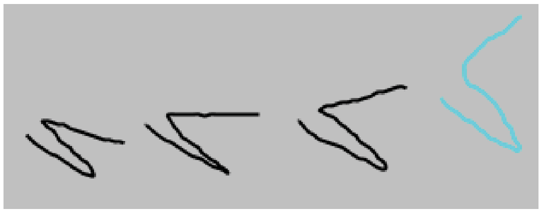


Figure 4. Recognising shading.

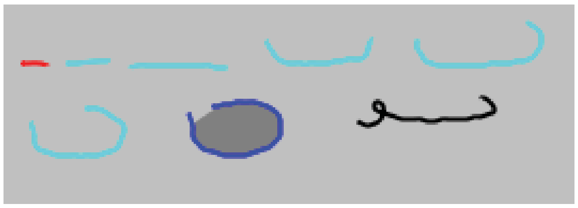


Figure 5. Recognising edges.

enter each type of syntax. In addition, they were told that it was a single stroke recogniser. In particular, they were informed that they had to remove the pen from the screen before drawing the next item. They were also told that if syntax was miscoloured then this was not their fault and they should not attempt to correct the stroke, because the study was about determining the correctness of the recognition engine. Each participant began the study when they were happy with using the tool.

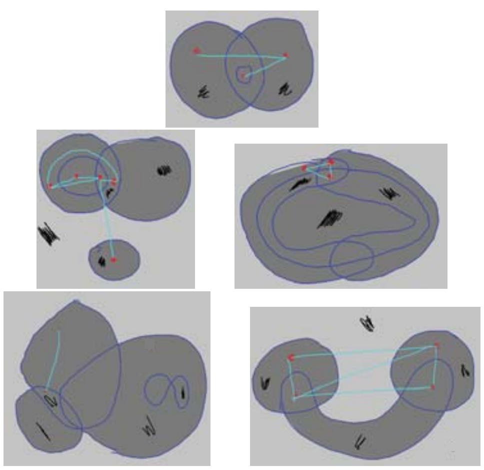


Figure 6. Diagrams used in the study.

For the study, each participant was asked to sketch six diagrams. The first five diagrams were given for them to copy and are shown in figure 6; each of these sketches was drawn by one of the 10 participants. For the sixth diagram, they were asked to draw any sketch with 3 curves, 3 nodes, 3 edges between the nodes, and 3 pieces of shading. An example diagram drawn by another participant can be seen

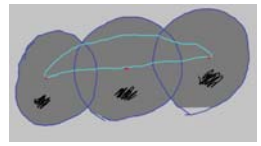


Figure 7. A sketch by a participant.

Syntax	Correct	Incorrect
Node	169	41
Not a node	587	4

Table 3. The node heuristics: initial classification.

in figure 7. Of note is that sometimes participants failed to copy the entire sketch and sometimes drew extra items in their sketches. Overall, there were roughly 200 strokes for each syntax type. Each sketch was saved for analysis purposes and was examined to determine whether it was correctly recognised. In particular, each stroke was assigned its correct syntax type by the researcher as well as the syntax type assigned by the recogniser.

Out of 801 sketched items, 747 were correctly classified, giving a success rate of 93.26% using the initial threshold values. Our task is now to determine whether the threshold values used in the heuristics can be adjusted to improve the recognition rate. Since the recogniser works hierarchically, we first examined the classification of nodes. Table 3 shows the number of strokes classified either as a node or not as a node using the initial values of the recogniser, broken down by whether that classification is correct. We can see that many of the $801 - 747 = 54$ original errors occurred in this part of the recognition process, with a total of $41 + 4 = 45$ errors. We extracted, from the saved sketches, all strokes incorrectly classified either as not being a node when it was a node (false negatives) or as being a node when it was not a node (false positives); these strokes can be seen in figures 8 and 9 respectively. In figure 8, some of the incorrectly recognised nodes are large and they break the condition that their bounding box is small. In figure 9, the strokes incorrectly recognised as nodes are generally small and it is perhaps unsurprising that errors occurred here. To improve the recogniser, we modified the values of $T_{n,vs}$ to

Syntax	Correct	Incorrect
Node	199	11
Not a node	576	15

Table 4. The node heuristics: final classification.

Figure 8. Strokes which should be nodes.

Figure 9. Strokes which should not be nodes.

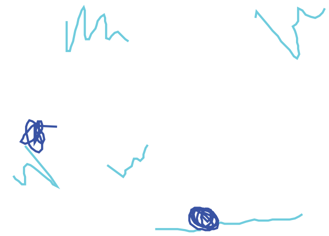


Figure 10. Strokes which should be shading.

Syntax	Correct	Incorrect
Shading	181	2
Not shading	393	0

Table 6. The shading heuristics: final classification.

15 and $T_{n,s}$ to 18, which were the best values we found after some exploratory data analysis. Altering the value of $T_{n,sq}$ did not seem to positively impact the recognition accuracy. Table 4 shows the numbers of correctly and incorrectly classifies strokes using these new threshold values for the nodes heuristics. There are now more false positives (increasing from 4 to 15) but many fewer false negatives (decreasing from 41 to 11).

Having changed the values of the thresholds associated with the node heuristics, we investigated classification errors associated with shading. At this point, we cannot overcome recognition errors due to the node heuristics. Thus, we now restrict our analysis to the only strokes that are correctly not recognised as nodes; this leaves $801 - 199 - 11 - 15 = 576$ strokes in the data set. At this point, there are no strokes which should have been recognised as shading, but were not recognised as such and only seven strokes that are shading but not recognised as such; see table 5 and figure 10 which shows the seven incorrectly classified strokes. Thus, we only attempt to improve the number of false negatives. Observing figure 10, we can see that some of these strokes do not have at least two sharp turns. Thus, we modified the shading threshold $T_{s,a}$ which is the angle used to determine whether a turn is sharp. The best value we found for this was 90 degrees. Table 5 shows the result of making

Syntax	Correct	Incorrect
Shading	176	7
Not shading	393	0

Table 5. The shading heuristics: initial classification.

this change, where no false positives were introduced and only 2 false negatives remain. These two incorrectly classified strokes can be seen in figure 11; since they are close to lines, it would be hard to distinguish them from edges.

Considering now the remaining data, i.e. the 393 strokes that are correctly identified as not being nodes or shading, the recogniser must determine whether each stroke is an edge. In fact, the recogniser has a 100% success rate here, correctly recognising 191 strokes as lines and the remaining 202 strokes are, by default, recognised as curves.

Using the final threshold values (table 2), of the 801 data points, 773 were correctly classified (totalling 11+15+2=28 errors), giving a final success rate of 96.5%. In order to determine whether the improvement is statistically significant, we carried out a McNemar χ^2 test (similar to a χ^2 test, but for paired data) using the data in table 7. Computing

$$\frac{(b-c)^2}{b+c} = \frac{(9-35)^2}{9+35} = 15.36$$

and comparing with a χ^2 table with one degree of freedom, we get a p-value of 0.0000886687. Therefore, the improvement is highly significant.

Figure 11. Strokes which should be shading after improving the recogniser.

		Final	
		Correct	Incorrect
Initial	Correct	$a = 738$	$b = 9$
	Incorrect	$c = 35$	$d = 19$

Table 7. Data for the McNemar test.

5. Conclusion

This paper presents the first tool, SpiderSketch, that supports users in the creation of augmented Euler diagrams. It uses a set of heuristics that determine whether each sketched item is a node, shading, an edge, or a curve. The recogniser was evaluated and improved using a sample of sketches produced by 10 participants. The improvements have been shown to be highly statistically significant. An accuracy rate of 96.5% is very high for a rules-based recogniser. Previous work [3] devised a recogniser, called CALI, for geometric shapes, possibly drawn using multiple strokes, such as lines, circles, rectangles and triangles. Whilst similar to our work here, it is our understanding that CALI would recognise each shape as some geometric shape, whereas for Euler diagrams it is important that some curves are recongised as just that, and not some particular shape.

Future work includes conducting another study, to re-evaluate the recongiser against new data, with the sketches drawn by different participants. This will provide insight into how accurately the recogniser works after the changes we have made. We acknowledge that there is a danger of over-fitting the recogniser to the data set considered in this paper. In addition, we plan to improve the recognition by taking into account context. We want to explore a blend of online recognition with contextual re-recognition. In our case, even after the improvements we made, there were still misclassifications occurring between nodes and shading in particular. Since each edge should end next to a node, we can use this contextual information to reclassify items as more of the sketch is drawn. In addition, we also plan to extend the recogniser to more complex diagrams that include arrows and mathematical symbols. We expect context to play a bigger role in fixing classification errors in these more complex diagrammatic notations. Further, if we wanted to include text recognition then we would expect to use a more sophisticated recogniser such as RATA [1].

A further avenue of work is to extend the functionality of the tool so that is supports users given it's understanding of the sketch. In particular, we plan to include a feature that automatically produces a 'formal' version of the sketched diagram; by this, we mean a beautified sketch that looks as though it was drawn using an editing tool. Moreover, we want to incorporate syntax matching, so that any semantic differences introduced when converting sketch to formal can be automatically identified and rectified, as we have done for Euler diagrams in [15].

Acknowledgement This research is supported by EPSRC grant EP/H048480/1 and a Royal Society of New Zealand Marsden Grant.

References

[1] S. Chang, B. Plimmer, R. Blagojevic. Rata.ssr: Data mining for pertinent stroke recognizers. In *Sketch Based Interface Modeling*. ACM, 2010.

[2] A. Delaney, B. Plimmer, G. Stapleton, P. Rodgers. Recognising sketches of Euler diagrams drawn with ellipses. In *Visual Languages and Computing*, pp 305–310. Knowledge Systems Institute, 2010.

[3] M. Fonseca, C. Pitmentel, J. Jorge. CALI: An Online Scribble Recongiser for Calligraphic interfaces. AAAI Spring Symposium, 2002.

[4] G. Goldschmidt. *Visual and Spatial Reasoning in Design*, chapter The Backtalk of Self-Generated Sketches, pp 163–184. University of Sydney, 1999.

[5] J. Howse, G. Stapleton, and J. Taylor. Spider diagrams *LMS Journal of Computation and Mathematics*, 12(3):299–324, 2005.

[6] J. Gil, J. Howse, S. Kent. Formalizing spider diagrams *IEEE Symposium on Visual Languages*, pages 130-137, 1999.

[7] G. Johnson, M. Gross, J. Hong. *Computational Support for Sketching in Design*. Now Pub. Inc., 2009.

[8] S. Kent. Constraint diagrams: Visualizing invariants in object oriented modelling. In *Proceedings of OOPSLA97*, pp 327–341. ACM, 1997.

[9] K. Mineshima, M. Okada, Y. Sato, R. Taakemura, iagrammatic Reasoning System with Euler Circles: Theory and Experiment Design. In *Diagrams*, Springer, pp 188-205, 2008.

[10] I. Oliver, J. Howse, G. Stapleton, E. Nuutilal S. Torma, Visualising and Specifying Ontologies using Diagrammatic Logics. In *5th Australasian Ontologies Workshop*, CRPIT, pp 87-104, 2009.

[11] N. Riche, T. Dwyer Untangling Euler diagrams. IEEE Transactions on Visualization and Computer Graphics, 16(6):1090-1099, 2010.

[12] *www.cem.brighton.ac.uk/users/ges9/SketchingEuler Diagrams/SketchingEulerDiagrams.html* . 2011.

[13] S.-J. Shin. *The Logical Status of Diagrams*. 1994.

[14] N. Swoboda, G. Allwein. Heterogeneous reasoning with Euler/Venn diagrams containing named constants and FOL. *Euler Diagrams 2004*, ENTCS, 2005.

[15] M. Wang, B. Plimmer, P. Schmieder, G. Stapleton, P. Rodgers, A. Delaney. SketchSet: Creating Euler diagrams using pen or mouse Accepted for *VL/HCC*, 2011.

Exploiting Visual Language Technologies in Structural Analysis

Vincenzo Deufemia[1], Massimiliano Giordano[1], Dario Nicola Pica[2], Giuseppe Polese[1]
[1]*Facoltà di Scienze MM.FF.NN., Università di Salerno*
Via Ponte don Melillo, 84084 Fisciano(SA), Italy
{deufemia,mgiordano,gpolese}@unisa.it

[2]*Soft.Lab S.r.l.*
Via Borgo II, 82030, Ponte (BN), Italy
d.pica@soft.lab.it

Abstract

In this paper we propose a visual language based framework to effectively tackle the problem of software based structural analysis in different application domains. In particular, the framework includes grammar based parser generation modules to easily adapt structural analysis software packages to evolving standards of specific application domains. Moreover, it includes visual analytics paradigms to enhance the software based structural analysis processes. To demonstrate the feasibility of the proposed framework we have implemented some of its modules and instantiated them in the context of the evaluation of earthquake-resistant masonry buildings.

1. Introduction

Structural analysis comprises the set of physical laws and mathematics required to study and predict the behavior of structures [3]. The subjects of structural analysis are engineering artifacts whose integrity is judged largely based upon their ability to withstand loads; they commonly include buildings, bridges, aircraft, ships and cars.

In the last decades, Computer Aided Design (CAD) and Computer-Aided Structural Engineering (CASE) software packages have been developed to provide automatic support for structural analysis [8]. These packages share many common characteristics across different application domains of structural analysis, such as the extensive use of computer graphics and complex mathematical models.

Structural analysis software producers strive to keep their products competitive, since they need to continuously adapt them to emerging computer graphics technologies and to evolving standards of specific application domains. Thus, they need frameworks and software platforms assisting them in the evolution of their software products.

In this paper we propose a visual oriented framework to support the effective generation of structural analysis software modules. We have also implemented two main modules of it, one for verifying compliance of designed structures to domain specific standards, and one to configure visual data mining activities during the design and evaluation of structures. In particular, in the former we have devised a grammar-based approach for modeling norms, which enabled us to exploit automatic parser generation techniques and to effectively embed them within structural analysis software packages. Moreover, the second module aims to enhance structure design through visual data mining, so as to stimulate more a proficient involvement of engineers in the structural analysis process, and to let them better convey their experience.

To demonstrate the feasibility of both the proposed approaches, we have implemented them in the context of the evaluation of earthquake-resistant masonry buildings.

2. The Proposed Framework

Fig. 1 shows the architecture of the proposed framework. It is composed of three main modules: a *Computer Aided Design* module, which includes the components forming a CAD system, a *Visual Language Parser Generator* (VLPG) that generates structural analysis components from grammars modeling safety standards, and a *Visual Data Mining Configuration* (VDMC) module that allows engineers to visually select algorithms and metaphors to configure a knowledge discovery process based on a proper visualization of the information computed during structural analysis.

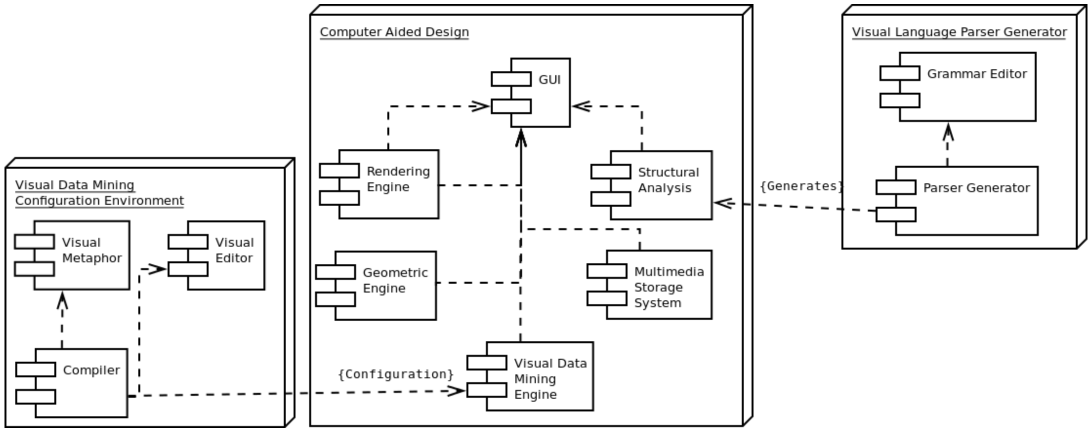


Figure 1. The architecture of the proposed framework.

The role of the VLPG is to guarantee the possibility of modeling safety norms through a rigorous formalism. In this way, norms specification is maintained outside of the source code, which allows to keep structural analysis packages compliant with latest versions of norms without changing their source code.

The VDMC module has the goal to facilitate the introduction of knowledge discovery processes within structural analysis. Potentially, structural analysis can have conspicuous benefits from the usage of knowledge discovery process, allowing designer to share past experiences. However, each different application domain of structural design needs specific data mining algorithms, and visual metaphors tailored to domain. For this reason, we needed a software platform supporting the configuration of the whole knowledge discovery process. In addition, since structural analysis packages can work with different mathematical resolution engines, this module also allows the selection of these components.

In the following we provide details of the VLPG and VDMC modules by using the masonry buildings case study to illustrate concepts and examples.

3. Generation of Structural Analysis Components through Visual Grammars

In this section we describe how an approach proposed for the analysis of visual languages can be adapted to the domain of structural analysis for engineering design. In particular, we show how to use visual language grammar models and parser generators to derive structural analysis software module capable of checking compliance of structures to safety constraints.

As a parse generator we use *VLDesk* [4], a system that inherits and extends to the visual field concepts and techniques of compiler generation tools like YACC [6]. In fact, *VLDesk* is based on the formalism of eXtended Positional Grammars (XPGs), which represents a direct extension of context-free string grammars [1], where more general relations other than concatenation are allowed [4]. The idea behind the definition of such formalism has been to overcome the inefficiency of visual languages parsing algorithms by searching suitable extensions of the well-known LR technique.

Our idea is to exploit the XPG formalism to specify structural analysis constraints, and then to use the generated LR-based parser to verify the compliance of designed structures to such constraints. In this way software packages for structural analysis can be easily kept up to date with evolving safety norms, without having to modify source code and to recompile it.

In the following we show how to model anti-seismic construction norms for the selected case study. In particular, we first need to define the symbols of the domain, e.g., openings, walls, and floors, and then need to specify the grammar productions by enclosing constraints between symbols and procedures to verify properties on the whole design model. The XPG formalism conceives the symbols of a visual language as graphical objects, each associated with a set of *syntactic attributes* [4] used to relate symbols between them. Fig. 2 shows some symbols used for modeling masonry buildings, together with their syntactic attributes, which correspond to their coordinates in the plane. Fig. 3 shows an example of model representing the front of a building.

Fig. 4 shows a portion of grammar productions modeling buildings satisfying anti-seismic construction norms, which we have defined for the evaluation of earthquake-resistant masonry buildings. The productions alternate symbols with relations and have

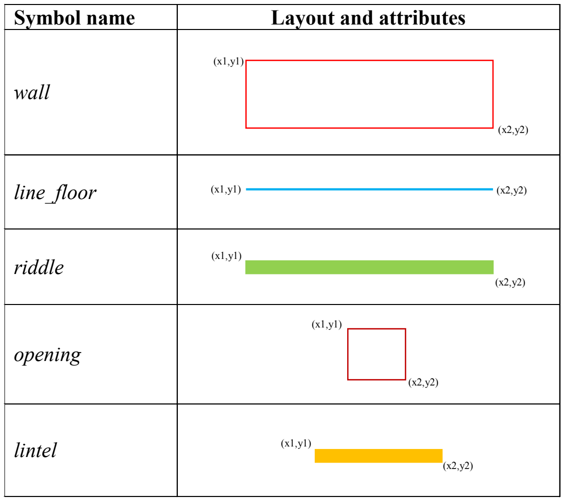


Figure 2. Some visual symbols for the building design domain.

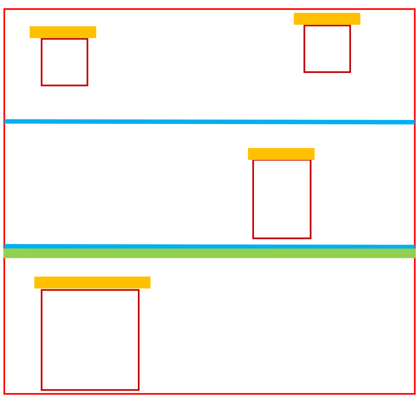


Figure 3. A model representing the front of a building.

associated Δ rules, which are used to compute the attributes of the symbol on the left-hand side of the production from the attributes of the symbols on the right-hand side. The grammar defines a building structure as a set of floors containing several openings. The floors structure is composed of a wall with several lines representing a floor with an optional riddle underneath it (productions 2-6). The openings are composed of a set of non-intersecting opening symbols each having a lintel above it (productions 7-9). The semantic checks associated to the first production are invoked once the whole engineering model has been recognized. In the proposed example, the checks regard further anti-seismic constraint verifications on the model.

Fig. 5 shows the procedures we defined for the detection of the bearing walls and the verification of some properties on them. Such procedures exploit the information computed during the parsing, e.g. the position of the floors in the designed models, in order to determine the position and the size of the bearing walls contained in the model, and verify that they comply with anti-seismic construction norms. As an example, the procedure *Check_Bearing_Wall_Constraints* verifies the property: a bearing wall has to be 30 percent larger than the floor height.

```
(1) STRUCTURE → FLOORS <contains, not_intersect> OPENINGS
        {
        Bearing-wall=Campute_Bearing_Wall();
        Check_ Bearing_Wall _Constraints(Bearing-wall);
        Visualize_ Bearing_Wall ();
        }

(2) FLOORS→ FLOORS’ <contains> INTERFLOORS
        Δ:{ FLOORS.Interfloors= FLOORS’. Interfloors ∪ {INTERFLOORS.x1,
        INTERFLOORS.y1, INTERFLOORS.x2, INTERFLOORS.y2;}

(3) FLOORS→ wall
        Δ:{ FLOORS.x1 = wall.x1; FLOORS.y1 = wall.y1; FLOORS.x2 = wall.x2; FLOORS.y2 =
        wall.y2;}

(4) INTERFLOORS → line_floor <isUnder, Height(0.2,0.4)> RIDDLE
        Δ:{ INTERFLOORS.x1 = line_floor.x1; INTERFLOORS.y1 = line_floor.y1;
            INTERFLOORS.x2 = line_floor.x2; INTERFLOORS.y2 = line_floor.y2;}

(5) INTERFLOORS → line_floor
        Δ:{ INTERFLOORS.x1 = line_floor.x1; INTERFLOORS.y1 = line_floor.y1;
            INTERFLOORS.x2 = line_floor.x2; INTERFLOORS.y2 = line_floor.y2;}

(6) RIDDLE → riddle
        Δ:{ RIDDLE.x1 = riddle.x1; RIDDLE.y1 = riddle.y1;
            RIDDLE.x2 = riddle.x2; RIDDLE.y2 = riddle.y2;}

(7) OPENINGS → OPENINGS’ <not_intersect> OPENING
        Δ:{ OPENINGS.ap = OPENINGS’.ap ∪ OPENING.ap }

(8) OPENINGS → OPENING
        Δ:{ OPENINGS.ap = OPENING.ap }

(9) OPENING → opening <isAbove, Height(0.1,0.2),isCentered,isLarger(0.15,0.3)> lintel
        Δ:{ OPENING.ap = {opening.x1, opening.y1, opening.x2, opening.y2}}
```

Figure 4. A portion of the grammar modelling anti-seismic norms.

```
Compute_Bearing_Wall(Floors, Openings){
        interfloor = getMin(Floors);
        interfloor-prev = (0,0,0,0);
        i = 0;
        while (interfloor != ∅){
                openings_0 = interfloor-prec;
                openings = computeOpenings(Openings, interfloor);
                openings_(openings.size+1)=(interfloor.x2,interfloor.y2)
                for (j=1; j<=openings.size+1;j++) do
                        bearing-wall_i[j] = (openings_(j-1).x2, interfloor.y1,openings_j.x1,
interfloor-prec.y1);
                if(i>0) bearing-wall_i = bearing-wall_i ∩ bearing-wall_(i-1);
                interfloor-prec  = interfloor;
                interfloor = next(interfloor);
                i++;
        }
        return bearing-wall;
}

Check_Bearing_Wall_Constraints(bearing-wall){
        for each bw in bearing-wall do{
                if(bw_i[j] < 0,3 * height_interfloor(i))
                        alert(“The bearing-wall ” + j + “ at floor “+ i + “ is large less
than ” + 0,3*h_interfloor);
        }
}
```

Figure 5. Semantic actions associated to the grammar for computing bearing walls and verifying an anti-seismic norm.

4. Configuration of Visual Data Mining Processes

Visual data mining (VDM) is a combination of traditional data mining techniques and visualization methods [5], and it represents a useful mean to support engineers during the activities of structural analysis. As an example, given a design building, the VDM can be used to retrieve the building structures analyzed by engineers in the past, which highlighted the same structural issues; but VDM can also be used to find out the visual patterns common to several building structures.

Since VDM algorithms embed concepts far from engineer's background, their definition requires the use of means that properly assist both experienced and naïve users. Inspired by the work presented in [2], here, we present a visual language to provide users with a simple and intuitive mean for the configuration of VDM processes and for the selection of the right visual metaphor for data analysis.

The visual language used to configure VDM algorithms is the flowchart notation, where the boxes represent the processing operations, i.e., the data mining algorithms chosen from a repository, and the arcs represent the execution flow. Fig. 6 shows an example of visual sentence, which defines a visual data mining process for the construction of a decision tree based on the similarities among bearing walls. In particular, the process prescribes (1) the comparison of all the bearing walls previously computed from the building models, after having selected a proper similarity algorithm; (2) the clustering of the comparison results based on the chosen algorithm, (3) the construction of a decision tree based on the clusters having a size greater than 4.

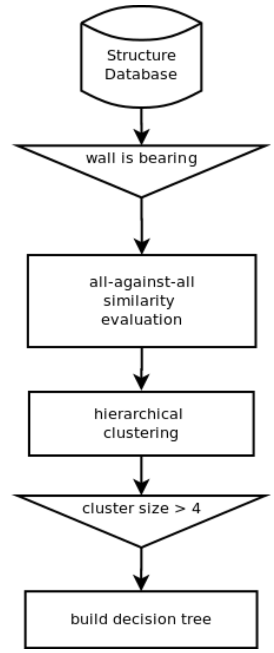


Figure 6. A visual sentence defining a visual data mining algorithm

5. Conclusions

The availability of advanced analysis tools based on finite element and matrix structural analysis concepts has enabled engineers and architects to model, analyze and design innovative, complex and unusual building and industrial structures [7]. In this paper, we have presented a framework for supporting the construction of structural analysis software packages. In particular, the framework addresses the construction of those package modules that need frequent updates to reflect changes in the application domain, or changes across different application domains. The framework exploits known paradigms developed in the visual language research community, and it has been used experimentally in the domain of masonry building.

Acknowledgments

This research has been supported in part by the Campania Regional Government through the "POR Misura 3.17 - Ricerca industriale e sviluppo competitivo – Metadistretto ICT".

6. References

[1] Aho, A.V., R. Sethi, and J.D. Ullman, *Compilers Principles, Techniques, and Tools*, Addison-Wesley Series in Computer Science, 1987.

[2] S. Bimonte, R. Laurini, F. Ferrucci, and G. Polese, "Prototype of a Visual Language for Spatial Data Mining based on the 'Miner Trip' Metaphor: VisMiner", in *Proc. of HCC 2003*, pp. 76-83.

[3] M. L. Camilleri, *Structural Analysis*, Nova Science Publishers, 2009.

[4] G. Costagliola, V. Deufemia, and G. Polese, "A Framework for Modeling and Implementing Visual Notations with Applications to Software Engineering", *ACM TOSEM*, 13(4), 2004, pp. 431-487.

[5] U. M. Fayyad, G. Grinstein, and A. Wierse, *Information Visualisation in Data Mining and Knowledge Discovery*, Morgan Kaufmann Publishers, San Francisco, 2002.

[6] S. C. Johnson, "YACC: Yet Another Compiler Compiler", Bell Laboratories, Murray Hills, NJ, 1978.

[7] M. Mahendran, "Applications of Finite Element Analysis in Structural Engineering", in *Proceedings of CAE*, 2007, pp. 38-46.

[8] O. Ostrouwsky, *Engineering Drawing with CAD Applications*, ELBS Pub., 1993.

Improving Shape Context Matching for the Recognition of Sketched Symbols

Gennaro Costagliola, Mattia De Rosa, Vittorio Fuccella
University of Salerno
Via Ponte Don Melillo
84084 Fisciano (SA), Italy
{gencos, matderosa, vfuccella}@unisa.it

Abstract

In this paper we present an approach to recognize multi-stroke hand drawn symbols, which is invariant with respect to scaling and supports symbol recognition independently from the number and order of strokes. The approach is an adaptation of the algorithm proposed by Belongie et al. in 2002 to the case of sketched images. This is achieved by introducing a new step in which the original Shape Context point-to-point cost matrix is updated according to stroke related information. The approach has been evaluated on a set of symbols from the Military Course of Action domain and the results show that the new recognizer outperforms the original one.

1 Introduction

The spread of systems equipped with touch screens has recently attracted the interest of researchers on the recognition of sketched diagrams. Several methods for the recognition of hand-drawn symbols have been proposed. The earliest of them [19] were only able to recognize unistroke symbols while a number of multi-stroke recognition approaches are now available.

In this paper, we present an approach to recognize multi-stroke hand drawn symbols. Since the approach is an adaptation of the image-based matching algorithm proposed by Belongie et al. [2], it is invariant with respect to scaling and is independent from the number and order of the drawn strokes. Furthermore, it has a better recognition accuracy than the original one when applied to hand drawn symbols. This is due to the exploitation of information on stroke points, such as temporal sequence and occurrence in a given stroke.

Briefly, the algorithm proposed by Belongie et al. calculates the matching cost between two shapes as the minimum weighted bipartite graph matching between two equally sized sets of sampled points from both shapes. This is done by calculating a matrix of matching costs between each couple of points of the two symbols and selecting the resulting best match. The cost of matching of two points is calculated by evaluating the difference between their *shape contexts*, which are suitable shape descriptors introduced by the authors.

Our improvement lies in re-calculating the cost matrix and, as a consequence, the total cost of matching between the two symbols. The cost matrix is re-calculated by considering the symbols as sequences of sampled points and detecting the longest subsequences of points in each sketched symbol stroke whose mapping on the template symbol produces still subsequences in any template symbol stroke. Once the subsequences are detected their lengths are used to decrease the matching costs of the involved points proportionally. This provides a further check on the structural similarity of the symbols that the original algorithm does not take into consideration that proves to enhance the accuracy when applying the shape context descriptors to sketch recognition.

The approach has been evaluated on a set of more than 100 symbols extracted from the *Military Course of Action Diagrams* (COA) domain [6]. In the experiment we have compared the performance of the previous algorithm to that of our enhanced version. We report a *top 1* recognition rate of 95.7% on symbols sampled at 128 points. This rate is for an improvement over the previous algorithm of 3.7%. The recognition grows up to 99% when considering the top 3 interpretations. The proposed improvement does not introduce signficant delays in the running time of the recognition procedure.

The rest of the paper is organized as follows: the next section briefly describes the approach presented in [2]; section 3 describes the approach. The evaluation is presented in section 4; section 5 contains a brief summary of the state of the art in hand-drawn symbol recognition; finally, section 6 offers our conclusions and outlines the future work.

2 Background: Symbol Recognition Through Shape Context

In this section we describe the symbol recognition procedure introduced by Belongie et al. [2] as used in our approach.

2.1 Feature descriptor

The shape context descriptor is first proposed by Belongie et. al [2] and is a global feature descriptor that has been successfully used in various application fields [2, 9, 10, 18]. It is a point based descriptor, thus a significant set of points must be sampled from the original figure. The shape context of a point is built by dividing a point's surrounding area into bins with different angles and increasing distances. The shape context of a point P_i is defined by computing the number of other points that are located in each of its surrounding bins,

$$h_i(k, l) = \#\{x \neq P_i : (x - P_i) \in bin(k, l)\} \tag{1}$$

where k and l are the spatial coordinates of the bin and # refers to set cardinality. As an example, in Figure 1, the bin containing the lowest point of the right side of the symbol A contains $bin(5, 20) = 4$ points. In other words, a point's shape context is a log-polar histogram which defines the relative distribution of other points. Different points in one shape have different shape context and similar points in two similar shapes have similar shape context. Belongie et al. [2] adopts the Chi-square distance to measure the difference between the shape context features of two points P_i and Q_j.

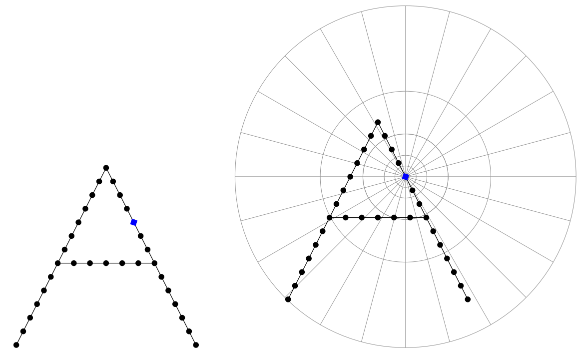
$$C_{ij} = \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{N} \frac{[h_i(k, l) + h_j(k, l)]^2}{h_i(k, l) + h_j(k, l)} \tag{2}$$

h_i and h_j are the shape context histograms of P_i and Q_j.

2.2 Matching

In order to compare a sketched symbol s against a template symbol t equation (2) must be calculated for each pair of sampled points P_i of s and Q_j of t. This will then produce a $n \times n$ matrix C where n is the number of points sampled from each symbol. Based on this cost matrix, the algorithm constructs a permutation π which allows to map each point P_i of s to a point $Q_{\pi(i)}$ of t such that the total matching cost H is minimized:

$$H(\pi) = \sum_{i}^{n} C_{i\pi(i)} \tag{3}$$



(a) Sampled points (b) Shape context of a point

Figure 1. The shape context descriptor

In the next section, we describe our approach and will make use of a function Π to map a sequence of points seq = $< P_1, \ldots, P_m >$ to the sequence of points $\Pi(seq) =< Q_{\pi(1)}, \ldots, Q_{\pi(m)} >$, with $m \geq 1$.

3 The Approach

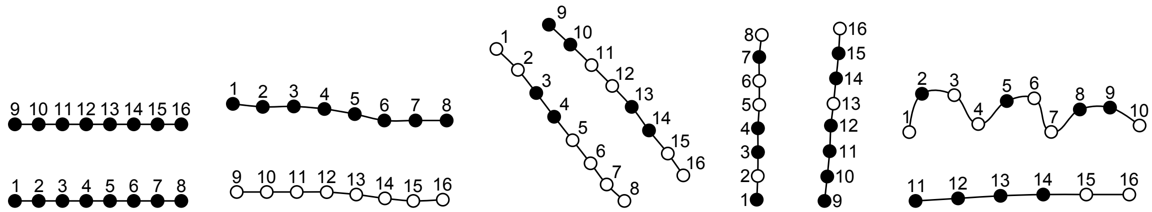
Sketched and template symbols consist of strokes starting with a pen-down and ending with a pen-up events. All the symbols are pre-processed to produce a set of a fixed number n of sampled points. These points are distributed among the strokes proportionally to their length. The $k-th$ stroke of a symbol is then represented as a sequence of sampled points $P_1, P_2, \ldots, P_{m_k}$ with $\sum_k m_k = n$.

Given a sketched symbol s and a set of template symbols $\{t_1, \ldots, t_l\}$, in order to find a template symbol that best matches s we first compute the matrix C and the mapping π as shown in section 2 on each pair (s, t_j). Then, based on C we compute an updated version $C'_{i\pi(i)}$ for $i = 1 \ldots n$ to calculate the new matching cost H' associated to t_j and s. Finally, the template symbol with the minimum cost will be presented as the best match.

Our contribution is the method to calculate the new $C'_{i\pi(i)}$. Basically, given a sketched symbol s and a template t we consider, for each stroke in s, the longest subsequences that are mapped by Π to sequences (also inverted) in any stroke of t. This provides a further check on the structural similarity of the two symbols that the shape context descriptor does not take into consideration. To reflect this property in the final cost computation, we proportionally decrease the matching cost of points participating in longer subsequences. More formally,

$$C'_{i\pi(i)} = C_{i\pi(i)} * (1 - 2 \times \log_{2n} |lsub(i)|) \tag{4}$$

where $|lsub(i)|$ denotes the size of the longest subsequence in a sketched symbol stroke containing the point P_i such



(a) Template (b) D.1 (c) D.2 (d) D.3 (e) D.4

Figure 2. A template and four hand drawn sketched symbols

$(i,\pi(i))$	(9,1)	(10,2)	(11,3)	(12,4)	(13,5)	(14,6)	(15,7)	(16,8)	(1,9)	(2,10)	(3,11)	(4,12)	(5,13)	(6,14)	(7,15)	(8,16)	$\sum$
$lsub$	s1								s2								
C	3.27	3.90	4.07	7.00	7.00	8.67	5.73	5.67	6.00	5.40	6.67	7.00	7.00	5.74	5.00	3.93	**92.05**
C'	-0.65	-0.78	-0.81	-1.40	-1.40	-1.73	-1.15	-1.13	-1.20	-1.08	-1.33	-1.40	-1.40	-1.15	-1.00	-0.79	**-18.41**

(a) Mapping between the D.1 and the template symbol

$(i,\pi(i))$	(4,1)	(3,2)	(5,3)	(6,4)	(7,5)	(8,6)	(16,7)	(15,8)	(2,9)	(1,10)	(10,11)	(9,12)	(11,13)	(12,14)	(14,15)	(13,16)	$\sum$
$lsub$	s1		s2				s3		s4		s5		s6		s7		
C	10.95	12.00	6.50	6.93	7.27	7.33	6.70	5.84	5.84	6.70	7.00	7.90	6.93	7.00	12.00	10.95	**127.85**
C'	6.57	7.20	1.30	1.39	1.45	1.47	4.02	3.50	3.50	4.02	4.20	4.74	4.16	4.20	7.20	6.57	**65.49**

(b) Mapping between the D.2 and the template symbol

$(i,\pi(i))$	(1,1)	(3,2)	(4,3)	(2,4)	(9,5)	(10,6)	(11,7)	(12,8)	(7,9)	(5,10)	(6,11)	(8,12)	(16,13)	(14,14)	(15,15)	(13,16)	$\sum$
$lsub$	s1	s2		s3	s4				s5	s6		s7	s8	s9		s10	
C	8.33	6.73	7.84	8.97	8.60	7.07	6.60	8.00	6.67	7.17	6.34	7.40	8.52	7.10	6.73	8.00	**120.08**
C'	8.33	4.04	4.70	8.97	1.72	1.41	1.32	1.60	6.67	4.30	3.80	7.40	8.52	4.26	4.04	8.00	**79.09**

(c) Mapping between the D.3 and the template symbol

$(i,\pi(i))$	(1,1)	(11,2)	(12,3)	(13,4)	(14,5)	(10,6)	(15,7)	(16,8)	(2,9)	(4,10)	(3,11)	(5,12)	(7,13)	(6,14)	(8,15)	(9,16)	$\sum$
$lsub$	s1	s2				s3	s4		s5	s6		s7	s8		s9		
C	6.55	2.24	2.00	1.93	2.00	10.57	1.27	1.50	7.81	4.93	9.33	6.60	7.27	4.83	5.57	3.67	**78.08**
C'	6.55	0.45	0.40	0.39	0.40	10.57	0.76	0.90	7.81	2.96	5.60	6.60	4.36	2.90	3.34	2.20	**56.19**

(d) Mapping between the D.4 and the template symbol

Figure 3. The mapping costs for the symbols in Figure 2

that $\Pi(lsub(i))$ or $inverse(\Pi(lsub(i)))$ is still a sequence in any template symbol stroke (containing the point $Q_{\pi(i)}$).

3.1 An example

Figure 2 shows a template symbol (a) and four sketched symbols (b, c, d, e) sampled at 16 points. On the sketched symbols, consecutive points belonging to the same subsequence are shown with the same color (white or black). Each table in Figure 3 shows the mapping π between one of the four sketched symbols and the template symbol. Each column shows the mapping between sketched and template symbol points, the subsequence to which the sketched point belongs and the cost of the mapping for both C and C' with the exception of the last column that shows the total cost of the mapping.

It can be noted that the C total cost for D.1 (92.05) is worse than the corresponding one for D.4 (78.08) despite the fact that D.1 is obviously the most similar sketched symbol for the template. On the other hand the C' total cost for D.1 (-18.41) is by far the lowest one.

4 Evaluation

The approach has been tested on a set of 113 symbols extracted from the *Military Course of Action Diagrams* domain [6]. COA diagrams are used to depict battle scenarios. Our set is an extract chosen by picking only a single variant of a symbol from the whole set. The domain, in fact, includes thousands of unique symbols. Nevertheless, many symbols are variants of a basic symbol, differing from it in a few details. The whole set of symbols is shown in figure 4. The above described symbol set was chosen in order to test the approach with a large (more than 100) set of symbols. A single template has been defined for each symbol. The number of strokes of the template symbols ranges from

Algorithm	top1	top2	top3	top4	top5	top6
C	91.0%	94.0%	95.6%	96.7%	97.0%	97.4%
C'	92.2%	97.1%	98.1%	98.4%	98.7%	98.9%
Diff.	+1.2%	+3.1%	+2.5%	+1.7%	+1.7%	+1.5%

Table 1. Result with 64 sampled points

Algorithm	top1	top2	top3	top4	top5	top6
C	92.0%	95.3%	96.3%	97.0%	97.6%	97.7%
C'	95.7%	98.2%	99.0%	99.1%	99.2%	99.2%
Diff.	+3.7%	+2.9%	+2.7%	+2.1%	+1.6%	+1.5%

Table 2. Result with 128 sampled points

2 for the simplest ones to 19 of the most complex one, with an average value of 5.9 (s.d. = 2.9) strokes.

In literature several systems for facilitating the input of COA diagrams have been described [4, 8, 13]. In [13] an accuracy of about 90% when considering the top 3 interpretations on a set of 485 symbols has been obtained.

In the experiment, we performed the recognition process on a set of symbols hand drawn by 5 different users. Each user drew all of the 113 symbols 4 times each. Thus, we collected a set of $5 \times 4 \times 113 = 2260$ sketched symbols. The following apparatus was used: the host system was a *Dell Precision T5400* workstation equipped with an *Intel Xeon* CPU at 2.50 GHz running *Microsoft Windows XP* operating system and the *Java Run-Time Environment 6*. The interactive device was a *Sympodium ID250* Interactive Pen Display, attached through both USB and RGB cables to the workstation.

In the experiment we compared the performance of the previous algorithm to that of our enhanced version. The recognition procedure was executed at two different point sampling rates (64 and 128) for both the sketched and template symbols. Shape contexts with 5 concentric circles and 12 sectors were used. For each symbol drawn by a user, the matching cost with each of the 113 template symbols was computed. Then we ordered the list of the templates increasingly by the similarity to the unknown symbol and we considered the position of the correctly matching template.

For $n = 1, \ldots, 6$, we report the ratio of matching templates falling in the top n positions. Tables 1 and 2 report the results of the above described trials for 64 and 128 sampling rates, respectively. In both tables, the first and the second row report the performance of the original and the improved algorithms, respectively; the third row reports the improvement obtained with the latter over the former.

As we can see, the improved algorithm has better performances in all cases. The improvement is more marked with a sampling rate of 128 points. In this case, a top 1 recognition rate of 95.7% is obtained. This rate improves that of the previous algorithm of 3.7%. The top 3 recognition rate is 99.0%.

5 Related Work

The methods and techniques treated in this paper fall within the sketch recognition and symbol matching research fields. In sketch recognition, the recognition of symbols can be regarded as the fists step (also called *low-level recognition*) of the process of recognizing more complex diagrams which use a predefined set of symbols (*high-level recognition*) [25].

Even though in most cases methods proposed for image recognition can be used [2, 24], several specialized methods for the recognition of *sketchy* images have been proposed. The earliest recognizers were only able to recognize unistroke symbols. In a pioneering work [19], a feature-based recognition approach is proposed. In this approach a stroke is characterized by 13 features including its length, size of the bounding box, average speed of the stylus, etc. A statistical pattern matching is used to compare the unknown stroke to those gathered in a training phase. Many unistroke symbol recognizers (e.g. text entry applications [15]) use elastic matching [20], a common pattern recognition-based approach to calculate a distance between two strokes. It basically works by evaluating the distances between corresponding points extracted from the two strokes. A recently proposed approach [22] has results comparable to those obtained through *elastic matching*, but enables accurate recognition with a few number of templates and can be easily implemented on any platform without requiring the inclusion of external libraries.

As for multi-stroke symbol recognition, many approaches have been tried, including image-based techniques, visual language parsing and graph matching.

In image-based techniques, the symbol is treated as a rasterized image. The advantage of such an approach is its independence on stroke order and number. E.g., in [14] the initial image is framed and down sampled into a 48 x 48 square grid. The recognizer exploits common classifiers in image-based matching, such as *Hausdorff* distance (and an ad-hoc defined variant of it), *Tanimoto* and *Yule* coefficients. The distances obtained by different classifiers are then combined together in order to obtain a unified measure.

As for grammar-based techniques, they rely on a grammar which defines the rules to determine if a given input belongs to the language generated by the grammar. In adjacency grammars, as those used in [17], primitive types are the terminal symbols and productions describe the topology of the symbols. Furthermore, a set of adjacency constraints (e.g. incident, adjacent, intersects, parallel, perpendicular) define the relation between two primitives.

The approaches based on graph matching usually represent symbols through *Attributed Relational Graphs* (ARG),

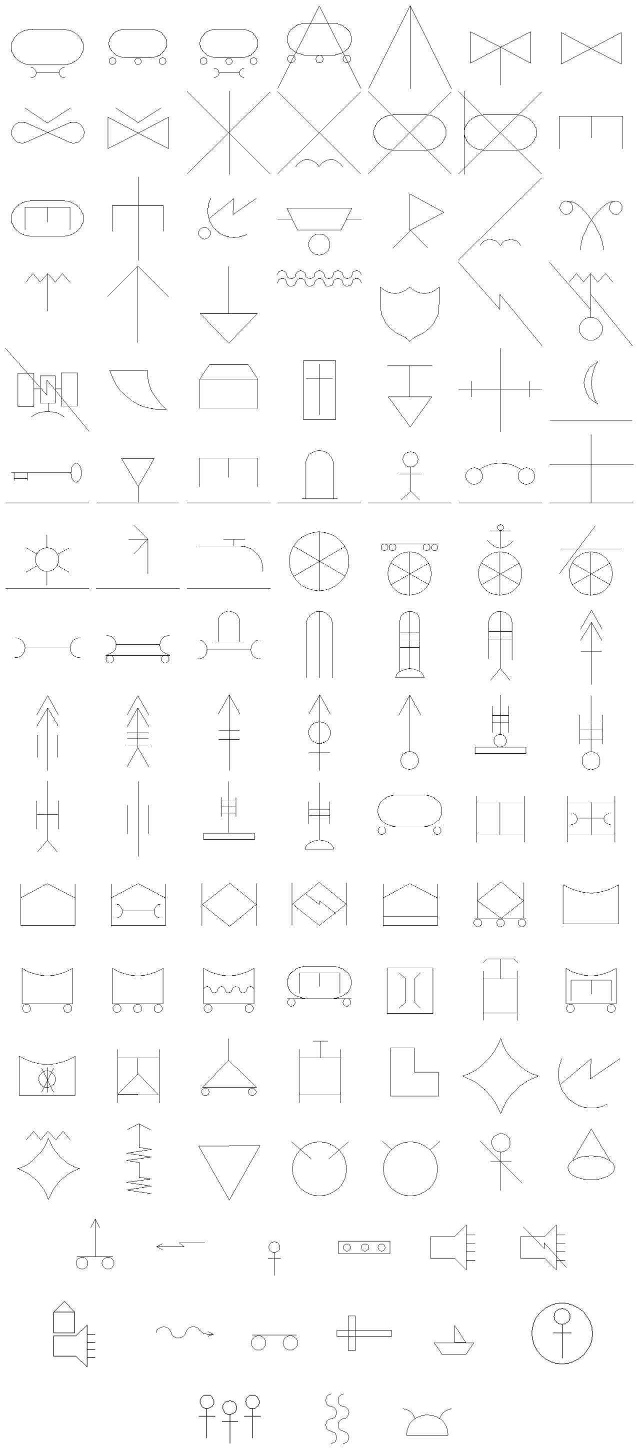


Figure 4. The 113 template symbols

which gives a structural description of the symbol [21]: the nodes in the graph are associated to the primitives composing the symbol, while the edges are associated to spatial relations between the primitives. ARGs are used for hand drawn symbol recognition in [16]. The authors solve the efficiency issues related to graph matching by exploiting approximate techniques.

Casella et al. [3], use an agent-based system for interpreting sketched symbols. The method exploits the knowledge about the domain context for disambiguating the symbols recognized at a lower level. The approach presented in [1] is an extension to the recognition of multi-stroke symbols of the *$1* approach proposed in [22]. It preserves the *minimalism* of its predecessor and relies on its single stroke recognition functionality. In [5] a stroke sequence of a symbol is firstly transformed to a string and then the *Levensthein* distance is used to compare two strings. The characters of the string are obtained by coding the directions of successive sampled points. This approach is clearly dependent on stroke order.

A statistical approach is used in [11]. Nine features are extracted from the unknown symbol and compared to those of the models. The best match is chosen through a statistical classifier. The approach also solves the problems related to the partitioning of the sketched elements (symbols, connectors, etc.) but requires the availability of at least five training examples per class.

General-purpose frameworks offering a whole set of functionalities for defining sketch-recognition environments, have also been proposed [12]. LADDER [12], in which a symbol is defined through a set of rules, which describe its shape principally on the basis of the characteristics of the primitives (length, slope, etc.) and relations among them (angle, presence of intersections, etc.). CALI [7] exploits a naive Bayesian classifier to recognize multi-stroke geometric shapes. A statistical analysis of various geometrical features of the shapes is performed, such as the convex hull, the largest triangle and largest quadrilateral that can be inscribed within the hull, the smallest area enclosing rectangle that can be fitted around the shape.

6 Conclusions

In this paper we have presented an adaptation to the case of sketched symbol of the algorithm for shape matching proposed by Belongie et al. in 2002. The effectiveness of the algorithm has been proven by testing it on a set of more than 100 symbol classes. We have found that the proposed enhancement improves the recognition rate of the original algorithm, obtaining an accuracy of 99% when considering the top 3 interpretations on symbols sampled at 128 points.

Future work include the test of the algorithm on larger sets of symbols, such as hand written oriental characters. It is worth noting that we have not tested the performance of the approach on rotated symbols. Point matching through *shape context* has been lately extended to rotated shapes [23], but their management in our approach is left to future work.

7 Acknowledgement

This research has been partially supported by the grant "cofinanziamento per attrezzature scientifiche e di supporto, grandi e medie (2005)" from the University of Salerno.

References

[1] L. Anthony and J. O. Wobbrock. A lightweight multistroke recognizer for user interface prototypes. In *Proceedings of Graphics Interface 2010*, GI '10, pages 245–252, Toronto, Ont., Canada, Canada, 2010. Canadian Information Processing Society.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:509–522, April 2002.

[3] G. Casella, V. Deufemia, V. Mascardi, G. Costagliola, and M. Martelli. An agent-based framework for sketched symbol interpretation. *Journal of Visual Languages & Computing*, 19(2):225–257, 2008.

[4] P. Cohen, L. Chen, J. Clow, M. Johnston, D. Mcgee, J. Pittman, and I. Smith. Quickset: A multimodal interface for distributed interactive simulation. In *Proceedings of the UIST'96 demonstration*, pages 217–24, 2003.

[5] A. Coyette, S. Schimke, J. Vanderdonckt, and C. Vielhauer. Trainable sketch recognizer for graphical user interface design. In *Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction*, INTERACT'07, pages 124–135, Berlin, Heidelberg, 2007. Springer-Verlag.

[6] T. D.U. Commented app-6a-military symbols for land based systems, 2005.

[7] M. Fonseca and J. Jorge. Using fuzzy logic to recognize geometric shapes interactively. In *Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference on*, volume 1, pages 291–296 vol.1, May 2000.

[8] K. D. Forbus, J. Usher, and V. Chapman. Sketching for military courses of action diagrams. In *Proceedings of the 8th international conference on Intelligent user interfaces*, IUI '03, pages 61–68, New York, NY, USA, 2003. ACM.

[9] M. Frenkel and R. Basri. Curve matching using the fast marching method. In *In EMMCVPR*, pages 35–51, 2003.

[10] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *EUROPEAN CONFERENCE ON COMPUTER VISION*, pages 224–237, 2004.

[11] L. Gennari, L. B. Kara, T. F. Stahovich, and K. Shimada. Combining geometry and domain knowledge to interpret hand-drawn diagrams. *Computers & Graphics*, 29(4):547–562, 2005.

[12] T. Hammond and R. Davis. Ladder, a sketching language for user interface developers. In *ACM SIGGRAPH 2007 courses*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.

[13] T. Hammond, D. Logsdon, B. Paulson, J. Johnston, J. Peschel, A. Wolin, and P. Taele. A sketch recognition system for recognizing free-hand course of action diagrams. In *Twenty-Second IAAI Conference on Artificial Intelligence*, 2010.

[14] L. B. Kara and T. F. Stahovich. An image-based, trainable symbol recognizer for hand-drawn sketches. *Comput. Graph.*, 29:501–517, August 2005.

[15] P.-O. Kristensson and S. Zhai. Shark2: a large vocabulary shorthand writing system for pen-based computers. In *Proc. of UIST '04*, pages 43–52, NY, USA, 2004. ACM.

[16] W. Lee, L. Burak Kara, and T. F. Stahovich. An efficient graph-based recognizer for hand-drawn symbols. *Comput. Graph.*, 31:554–567, August 2007.

[17] J. Mas, G. Sanchez, J. Llados, and B. Lamiroy. An incremental on-line parsing algorithm for recognizing sketching diagrams. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition - Volume 01*, pages 452–456, Washington, DC, USA, 2007. IEEE Computer Society.

[18] G. Mori and J. Malik. Estimating human body configurations using shape context matching, 2002.

[19] D. Rubine. Specifying gestures by example. *SIGGRAPH Comput. Graph.*, 25:329–337, July 1991.

[20] C. C. Tappert. Cursive script recognition by elastic matching. *IBM J. Res. Dev.*, 26:765–771, November 1982.

[21] W.-H. Tsai and K.-S. Fu. Error-correcting isomorphisms of attributed relational graphs for pattern analysis. *Systems, Man and Cybernetics, IEEE Transactions on*, 9(12):757–768, 1979.

[22] J. O. Wobbrock, A. D. Wilson, and Y. Li. 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, pages 159–168, New York, NY, USA, 2007. ACM.

[23] S. Yang and Y. Wang. Rotation invariant shape contexts based on feature-space fourier transformation. In *Proceedings of the Fourth International Conference on Image and Graphics*, ICIG '07, pages 575–579, Washington, DC, USA, 2007. IEEE Computer Society.

[24] W. Zhang, L. Wenyin, and K. Zhang. Symbol recognition with kernel density matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2020–2024, Dec. 2006.

[25] R. Zhao. Incremental recognition in gesture-based and syntax-directed diagram editors. In *Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*, CHI '93, pages 95–100, New York, NY, USA, 1993. ACM.

FcBD: An Agent-Based Architecture to Support Sketch Recognition Interfaces

Danilo Avola, Paolo Bottoni, Alexandru Dafinei, Anna Labella
Department of Computer Science, Sapienza University of Rome
Via Salaria 113, 00198 Rome, Italy
{avola, bottoni, labella}@di.uniroma1.it, a_dafinei@yahoo.it

Abstract

Sketch-based interfaces are increasingly used for interaction with desktops or mobile devices, favouring natural interaction through the expression of commands and concepts with graphical symbols. However, developing this kind of human-oriented interfaces is still challenging, due to: the limited number and type of graphical symbols definable and recognizable in a device-independent way, the computational cost incurred for recognizing complex symbols, and the different users' styles in tracing symbols. ***FcBD****'s agent-based architecture supports the definition and real-time recognition of sets of 2D graphical symbols, independently of the input style.* ***FcBD****'s layered organisation favours abstraction from the concretely employed device, and proposes an approach to symbol recognition based on associating with each recognition agent disjoint sets of primary and secondary features.*

1 Introduction

Multimodal interfaces exploit one or more human-oriented modalities (e.g. gesture, speech) to allow users natural interaction with desktop or mobile devices. Users increasingly access heterogeneous applications and services via different modalities according to environmental (e.g. indoor, outdoor) or device-dependent (e.g. screen size, computational capacity) characteristics. In this context, sketch-based interfaces provide a simple interaction tool that allows users to graphically express concepts and commands.

The development of a sketch-based interface meets three major problems: *multi-domain definition*, *real-time recognition* and *tracing style interpretation*.

The first problem concerns the construction of engines able to define and recognize a wide range of domains (i.e. sets of graphical symbols, called libraries). In fact, building a domain-specific sketch-based interface is a time-consuming operation that often requires the re-engineering and/or re-designing of the whole interface. Moreover, as libraries are extended with new symbols, recognition conflicts grow more relevant.

As for the second problem, the computational complexity of the recognition algorithms usually depends on both the number of symbol constituents (e.g. a circle inside a square is composed of two elements) and the number of strokes composing each basic element (e.g. a square can be drawn using four strokes/lines). Hence, a generic software architecture should support recognition processes on symbols of arbitrary complexity and its implementation should be as independent as possible of specific devices, applications, or services.

The last problem concerns the style adopted by users when sketching a symbol: 1) users may draw any given symbol with different numbers of strokes; 2) strokes can be drawn in different directions (e.g. leftwards or rightwards); 3) over-tracing and/or cross-tracing phenomena lead to drawing each stroke in solid, bold, or dashed styles. A style-independent recognition protocol is therefore desirable.

This paper describes the ***F****eature* ***c****alculation* ***B****id* ***D****ecision* (**FcBD**) agent-based architecture for sketch recognition. **FcBD** is equipped with two types of agents: *feature evaluators* and *symbol recognizers*, which are defined with respect to a specific library of symbols. Every time a user draws a symbol, the active evaluators work concurrently, each computing a specific feature on the drawn sketch. Recognizers read from the pool of features thus obtained only those needed to discriminate the symbol for which they are responsible, and make a bid for classifying the symbol under scrutiny. In case of conflict, a mediator agent progressively applies three solution strategies based on an adaptive choice or relaxation of features.

As feature calculation is largely independent of the set of symbols to be recognised, **FcBD** provides a rich set of evaluators, with a simple protocol for including new ones, and defines a procedure to associate features with symbols. Similarly, the definition of a new library of symbols, or the addition of a symbol to an existing library, requires a training period to establish the best combination of feature-symbol assignments. The recognition process is performed in real-time, independently of the user's input style.

The paper is structured as follows. Section 2 discusses related work on sketch-based interfaces. Section 3 presents **FcBD**'s architecture and a set of novel features. Section 4 shows some experimental results on a case study and Section 5 concludes the paper.

2 Related Work

Due to the vastity of the literature on sketch recognition, we focus only on works which are more directly related to our approach. The existing approaches can be generically classified as *feature-based* or *language-based*, identifying processes driven by mathematical and/or geometrical features, or by description language specifications, respectively. Our approach tends to take advantage of both classes.

Surveys on works in the first class are [8, 9], which also provide background on the identification of relevant features. In these works the authors propose a robust and extensible approach through which to identify a wide range of freehand drawn 2D shapes. They exploit geometrical properties and different sets of mathematical operators, as well as fuzzy logic mechanisms to identify the most common geometrical 2D entities.

Another meaningful work in the same class, showing different mathematical feature extraction methods, is [14]. The authors introduce five approaches through which to characterize and identify a sketched shape. The recognition processes are treated like image-processing problems (i.e. raster) and not as sketch-understanding issues (i.e. stroke processing). A further set of useful feature-based works is described in [7, 15]. In the first work the authors propose a new technique for gesture recognition which adopts a manifold learning approach combining temporal and spatial information, while handling multiple strokes with weighted distances. The second paper details the comparison of three machine learning methods for adaptive sketch recognition, where geometrical and dynamical features are examined to provide the more suitable set of readable measurements. Finally, the work in [3] combines image-based and time-based processing methods with the aim of providing a sketch recognition framework, combining knowledge on how objects look (image data) and how they are drawn (temporal data).

For the second class of works, we consider papers which have influences some basic architectural choices of our work. In [1, 2], the authors present SketchREAD, a multi-domain sketch recognition engine capable of recognizing freely hand-drawn diagrammatic sketches. The description of the graphical symbols is performed by a hierarchical shape description language taking into account the different geometrical and spatial constraints among the various components of a same symbol. Actually, LADDER provides the basic description language on which different authors have shaped their sketch definition approach (see [11, 12]) and still represents the reference point for advanced multi-domain and multi-purpose sketch recognition frameworks. LADDER's robustness and reliability are further highlighted in [13], where results of a on user tracings have been used to refine the shape description engine. Based on a variant of LADDER, the work in [4] introduces the description language SketchML and develops a number of features which have been reused in this work.

A multi-domain agent-based approach is adopted in the AgentSketch environment (see [6]) where different features, as well as domain-dependent information, are used in the recognition process, and domain contextual information is employed to solve ambiguities in symbol discrimination. While some aspects of the agent model are common to **FcBD** and AgentSketch, we are interested in symbol recognition centered on the agent's domain-independent knowledge of a specific symbol. Moreover, our disambiguation approach is only based on the agent's knowledge related to its own symbol without using contextual information.

3 The Framework Architecture

This section describes **FcBD** architecture, as depicted in Figure 1, which supports a MultiAgent System adaptable to sketch recognition in different domains. Each agent is able to autonomously recognise its own symbol independently of input style, device, or contextual information. Moreover, we adopted a modular and open architecture, so as to easily incorporate machine learning mechanisms to enable agents to refine their own recognition criteria. However, such refinement mechanisms have not been yet put in place in the current implementation.

FcBD is structured into five layers which process the set of strokes drawn by the user up to producing a univocal assignment of such a set to a symbol type from a library (or an error message).

The first layer (*Data Structure Builder*) produces a

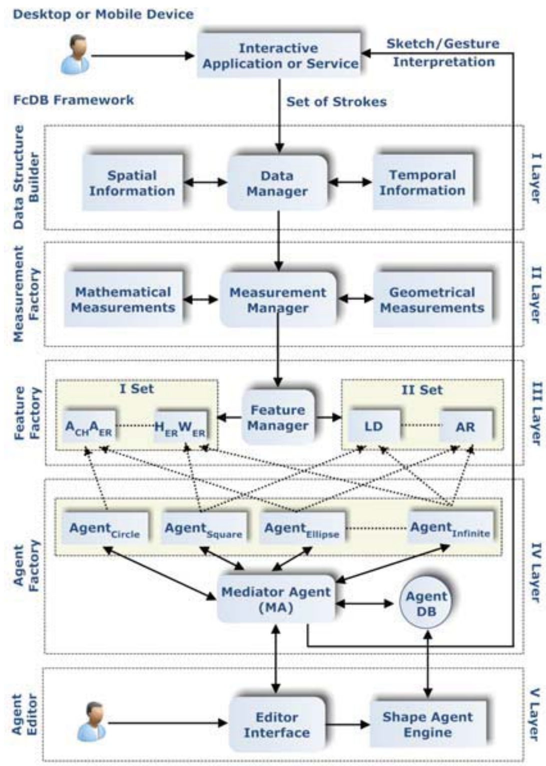


Figure 1. FcBD Framework Architecture.

data structure including spatial and temporal information for each stroke drawn by the user. A *Data Manager* performs a set of normalization operations (e.g. interpolation, noise filtering) in order to obtain a data structure independent of the physical input acquisition device. In this way, **FcBD** can in principle be used to recognize symbols which were physically drawn using different devices, e.g. on a tablet PC, or by performing gestures in front of a digital camera.

In the second layer, a *Measurement Factory* computes mathematical and geometrical measurements on each stroke, based on the data structures produced by the previous layer. Measures are defined according to a set of *special polygons*, well known from discrete geometry [5, 10], and to the mathematical properties of a stroke (e.g. length, degree of curvature). The main task of the *Measurement Manager* is to coordinate the process on each analyzed stroke.

The *Feature Factory* layer represents the core of the system, through which all the feature evaluator agents are activated. These agents evaluate derived features (e.g. ratios, correlations) from the measurements obtained from the previous layer. **FcBD** provides two different sets of features. The first set, based on [8, 9], comprises some style-independent features for identifying the main properties of a shape. The second set contains novel features conceived to complement the features in the first set. The two sets together provide a rich description of 2D graphical symbols. The main task of the *Feature Manager* is to produce and index the features to make them available to the next layer.

The *Agent Factory* in the fourth layer provides the execution environment for recognizer agents. For every symbol belonging to the library, a specific agent is endowed with knowledge about the set of features (the *feature vector*) which are collectively most suitable to univocally identify instances of that symbol, and the typical intervals in which the feature measures fall for such instances. All the recognizer agents have the same simple behavior: first, they perform the matching process between the measures extracted from the user sketch and the intervals associated with the features relevant to the agent; second, they return a numerical value indicating their confidence in the matching. Actually, every agent is responsible for evaluating if a sketch belongs to its family of symbols, composed of the different orientations for a specific symbol (e.g. north or south arrow-head). When two or more recognizer agents identify the user sketch as belonging to their symbol family, i.e., they bid for the sketch with high confidence, then the mediator agent acts to solve the ensuing conflict. In particular, it compares confidence degrees and asks agents with high comparable degrees to take into account features not previously considered.

The *Agent Editor* layer allows off-line addition of new symbols and related agents: for every new library symbol, a skilled user can define a new recognizer agent by supervising the selection of the specific set of features (and their typical intervals) associated with the agent. As all agents only differ for the associated sets of features, no programming is involved in this phase, unless new features need to be added to the library, for which specific algorithms have to be defined. Moreover adding new recognizers does not require any modification of the mediator, since the ambiguity resolution mechanism is independent of any specific symbol.

The architecture of **FcBD** is designed with a view to computational efficiency. Once the set of strokes from the device is obtained, both spatial and temporal information and geometrical and mathematical measurements are computed in two consecutive steps, with all the computational processes occurring simultaneously for each step. Note that the evaluation of second-level measures can be considered computationally irrelevant since it only applies simple operations (typically ratios) to the measurements obtained in the previous step. Moreover, recogniser agents work in parallel and in an asynchronous way. Experimental tests (as the one discussed in Section 4), also on different complex ad-hoc test symbols, have achieved real-time recognition performance on PCs and on mobile simulators.

A final point should be spent on **FcBD** description language. This is currently described through a mathematical abstraction provided by the set of features that identify a particular shape. Although it currently represents an effective practice for shape description, we plan to integrate some formal language like the one in [12], in order to fully enable easy library expansion and interaction with different modalities.

3.1 Feature Sets

As discussed before, features in **FcBD** fall into two sets. The first, Set_I, contains measures (e.g. area, perimeter) derived from special polygons (e.g. convex hull, largest quadrilateral) which characterise 2D shapes in generic terms, computed on the set of pixels covered by the sketch (i.e. similarly to raster image processing). This achieves independence from the user input style, thus allowing **FcBD** to recognise symbols in a robust way with respect to possible user inaccuracies. The second set, Set_{II}, contains a set of measures (e.g. density, angle), derived from strokes's properties (e.g. length, distance), which are original to **FcBD**.

Table 1 presents two features from Set_I (for the complete set see [9]) and the whole of Set_{II}. Features in Set_I provide a set of basic measurements, while those in Set_{II} are adopted to increase accuracy and to identify the direction in which a symbol in a family (e.g. line, closed arc) is drawn. In case of conflicts, or if no agent places a bid, agents are asked to use features in Set_{II} to try increase their confidence degree.

Table 1. Feature Sets: Set_I and Set_{II}.

I Set	Description	Main Purpose
● $H_{ER}W_{ER}$	height enclosing rectangle / width enclosing rectangle	shape thinness; shape orientation.
● $A_{CH}A_{ER}$	area convex hull / area enclosing rectangle	rectangle similarity.
II Set	**Description**	**Main Purpose**
■ IOP	inner points / outer points	pattern of the points; pattern of the clusters.
■ LD	*line density*: pixel variations of the sketch	shape symmetry; shape orientation.
■ CLS	*closure*: proximity between start and end points of the sketch	conflict support; shape derivation.
■ AR	*angle ratio*: angles related to the skeleton of the sketch	change direction; shape identification.
■ $P_{CH}P_S$	enclosing rectangle perimeter / sketch perimeter	shape twists; shape derivation.
■ $A_{ERS}A_{BB}$	sketch enclosing rectangle area / sketch bounding box	shape rotation; shape orientation.

After the user has produced a free-hand sketch $\overline{s}$, every agent establishes its confidence in the membership of $\overline{s}$ in its family, based on the values of the features associated with the agent through the **FcBD** editor. Each agent k thus produces an integer value $d_k \in \{0, \ldots, 100\}$. The value d_k is computed by the function AG_k according to formula (1).

$$AG_k(\overline{s}) = \sum_{i=1}^{n} W_k(f_i) \times B_k(f_i(\overline{s})) \qquad (1)$$

For each feature $f_i \in Set_I \cup Set_{II}$, $W_k(f_i)$ associates it with a fixed numerical value (between 15 and 45), indicating its weight in discriminating instances of the symbol associated with agent k. These weights have been established using the editor and according to empirical observations, but could be modified by adding machine-learning capabilities to agents. In (1), $B_k(f_i(\overline{s}))$ is a binary value (0 or 1) which states whether f_i evaluated on the set of strokes $\overline{s}$ is significant for agent k. This value is derived from an experimental analysis which established the significant interval of values for f_i when drawing instances of the shape for which agent k is responsible.

Once d_k is obtained for each agent (based only on features in Set_I), the mediator establishes a ranking among symbols. If no d_k is above a confidence threshold (currently set at 55), then no symbol is recognized. In this case, the mediator asks each agent to start a new evaluation considering features in Set_{II} and adding the newly computed values $W_k(f_i) \times B_k(f_i(\overline{s}))$ to the original d_k. If no agent reaches the threshold, then **FcBD** returns a failure. Otherwise, **FcBD** chooses the highest ranking agent (i.e. with the highest d_k). When two or more agents rank first, then a conflict occurs.

3.2 Conflict Solution

FcBD deals with conflicts by employing three strategies, which are consecutively applied until one succeeds (or even the last one fails). The conflict solving task is given to the mediator agent which will interact with each involved agent.

In particular, if only one agent produces a d_k above the threshold (no conflict), the mediator must just identify the related symbol. When several agents exceed the threshold with the same highest confidence value (conflict), the mediator enters the resolution mode and starts applying the solution strategies. In the first strategy each agent considers those features in Set_{II} that are relevant to its family of symbols, possibly incrementing the confidence value. If the conflict is not resolved yet, the second strategy considers the features in Set_I that were not previously used by the conflicting agents, but which have been identified as relevant to the specific conflict. The identification of conflict-specific discriminant features is based on experimental training. Again, the new d_{k_i} values for each conflicting agent k_i are compared.

The final strategy allows agents to relax (by 20%) the numerical intervals for the features in Set_I that are most discriminant for the family associated with the agent. If all of the strategies fail, **FcBD** communicates that no symbol has been recognized.

Technically, each recogniser agent is implemented as an autonomous thread that processes the available set of strokes and transmits (before becoming idle) the results to the mediator. In case of conflict, the mediator will notify and wake up the involved recogniser agents to apply the required strategy.

4 Experimental Results

This section discusses some experimental results on a domain of symbols specifically designed to be easily realizable via simple gestures, as well as through traditional sketchpads or tablets (see Table 2). Symbols occur in families of rotations, thus testing the capabilities of **FcBD** to resolve both symbols and their orientation. Figure 2 presents a snapshot of the **FcBD** interface while recognising an *infinity* symbol in vertical orientation. Table 2 shows, for each symbol family F_{shape_k} (e.g. *closed arc*) and feature f_i (e.g. $A_{lg}A_{er}$), the interval (e.g. 0.55-0.78) and the assigned weight (e.g. 25%) in the recognition of a related symbol. A cross indicates that f_i is not used in the first phase of the recognition process, but can be used for conflict solution, depending on which families are in conflict.

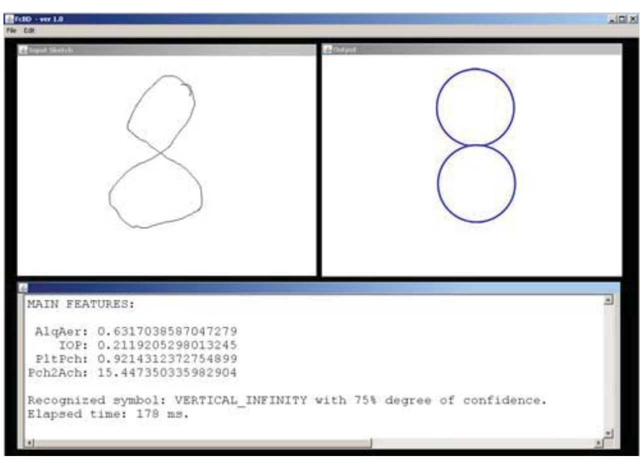


Figure 2. The FcBD environment in action.

In the first row A, H, W and P indicate *Area*, *Height*, *Weight* and *Perimeter*, respectively, while *lq*, *er*, *lt*, *ch* and *bb* denote geometrical shapes: *largest quadrilateral*, *enclosing rectangle*, *largest triangle*, *convex hull* and *bounding box*.

The interval ranges related to the features belonging to the first set are more relaxed compared to the original ones in [9]. Indeed, rather than considering these features as discriminant for specific families of symbols, we have used them to set a drawn sketch in a general class of shapes, distinguishing for example between angular or rounded shapes.

In Figure 3 we present the recognition rate for the symbols in Table 2, summarizing results on 200 attempts for each symbol. One can observe that symbols are directly recognized in the first phase for most of the time (76% to 95%). When conflicts occur (e.g. on *circle* or *ellipse*) the system is often able to find the solution (70% to 90%), while in a small number of cases the system fails. In particular, there are cases in which the framework correctly classifies the sketch in a family of symbols but does not resolve the orientation correctly (e.g. *closed arc* with north orientation recognized as oriented southwards), or incorrectly assigns a sketch to a different family of similar shape. In Figure 3 these cases are indicated as *Confusion Family* and *Confusion*, respectively. Finally, in about 5% of the cases the conflicts are not solved and the system does not recognize any symbol.

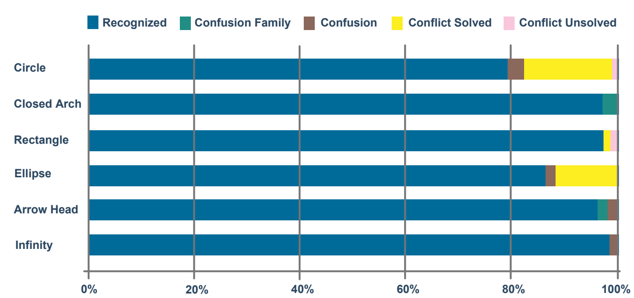


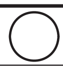
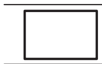
Figure 3. Success Rate on Test Domain.

During the experimental session, we have also measured **FcBD**'s time performance on the different symbols. Results indicate the possibility of usage for online recognition, as times range from a minimum of 150 ms (*line*) to a maximum of about 200 ms (*circle*), including the conflict-solving process.

5 Conclusion

Sketch-based interfaces provide an intuitive and natural way to graphically express concepts and/or commands to interact with desktop or mobile devices.

Table 2. Features and Associated Weights Within the Recognition Process.

Domain	$A_{lq}A_{er}$	$A_{lt}A_{ch}$	$A_{ch}A_{er}$	CLS	$H_{er}W_{er}$	IOP	$P_{lt}P_{ch}$	$A_{lq}A_{ch}$	$[P_{ch}]^2 A_{ch}$	Others
	[0.50 - 0.60] X	[0.40 - 0.45] X	[0.77 - 0.85] X	1 X	[0.80 - 1.0] 20%	[0.002 - 0.009] 25%	[0.84 - 0.88] X	[0.66 - 0.68] 25%	[12.0 - 13.5] 30%	$P_{ch}P_s$, LD, AR
	[0.55 - 0.78] 25%	[0.49 - 0.79] X	[0.50 - 0.90] X	1 20%	[0.60 - 0.99] X	[0.002- 0.050] 35%	[0.85 - 0.98] X	[0.73 - 0.96] 20%	[13.00 - 18.00] X	$A_{bb}A_{er}$,LD, AR
	[0.75 - 0.96] 25%	[0.45 - 0.60] X	[0.86 - 0.98] 25%	1 X	[0.40 - 0.99] X	[0.003 - 0.005] 25%	[0.80 - 0.92] X	[0.85 - 0.99] 25%	[14.0 - 20.0] X	$P_{ch}P_s$, AR
	[0.50 - 0.70] X	[0.40 - 0.50] 25%	[0.73 - 0.85] X	1 X	[0.34 - 0.80] 25%	[0.003 - 0.004] 25%	[0.83 - 0.89] X	[0.65 - 0.70] 25%	[13.50 - 18.50] X	$A_{bb}A_{er}$, $P_{ch}P_s$
	[0.50 - 0.54] X	[0.82 - 0.99] 35%	[0.35 - 0.60] X	0 30%	[0.50 - 0.99] X	[0.080 - 0.300] X	[0.97 - 0.99] 35%	[0.85 - 0.99] X	[18.0 - 23.0] X	$A_{bb}A_{er}$, LD
	[0.50 - 0.75] 25%	[0.45 - 0.55] X	[0.70 - 0.93] X	1 X	[0.40 - 0.65] X	[0.150 - 0.60] 45%	[0.84 - 0.93] 15%	[0.74 - 0.87] X	[15.0 - 19.2] 15%	$A_{bb}A_{er}$, LD, AR

We have presented **FcBD**, an agent-based architecture supporting the definition of libraries of 2D symbols and the derivation of specialised sketch-recognition systems tuned to the different libraries. **FcBD** supports an original strategy for conflict resolution, based on revision of features not directly considered for classification and adaptable to the specific symbols in conflict. **FcBD** is completely domain-independent. We have proved it on a set of symbols devised to be easily traceable with different devices (e.g. tablets, gestures in front of a camera, or movements with a smartphone), but the individual features it adopts have been already tested on different domains (e.g. ERD, DFD, UML [4]).

Current work is in four directions. First, we want to perform an extensive testing of the new architecture with respect to the above mentioned domains. Second, we need to provide a formal definition of the agent-description language to favor end-user development of domain-specific symbols. Third, an image processing layer is being developed for the integrated capture and recognition of user gestures. Finally, we want to facilitate the use of the editor for occasional users.

References

[1] C. Alvarado. Multi-domain sketch understanding. In *ACM SIGGRAPH 2007 courses*. ACM, 2007.

[2] C. Alvarado and R. Davis. Sketchread: a multi-domain sketch recognition engine. In *Proc. UIST 2004*, pages 23–32. ACM, 2004.

[3] R. Arandjelović and T. M. Sezgin. Sketch recognition by fusion of temporal and image-based features. *Pattern Recogn.*, 44:1225–1234, June 2011.

[4] D. Avola, A. Del Buono, G. Giorgio, S. Paolozzi, and R. Wang. SketchML a representation language for novel sketch recognition approach. In *Proc. PETRA '09*, pages 31:1–31:8. ACM, 2009.

[5] J. E. Boyce, D. P. Dobkin, R. L. Drysdale, III, and L. J. Guibas. Finding extremal polygons. *SIAM J. Comput.*, 14:134–147, February 1985.

[6] G. Casella, V. Deufemia, V. Mascardi, G. Costagliola, and M. Martelli. An agent-based framework for sketched symbol interpretation. *J. Vis. Lang. Comput.*, 19:225–257, April 2008.

[7] H. Choi, B. Paulson, and T. Hammond. Gesture recognition based on manifold learning. In *Proce. SSPR & SPR '08*, pages 247–256. Springer, 2008.

[8] M. J. Fonseca, A. Ferreira, and J. A. Jorge. Sketch-based retrieval of complex drawings using hierarchical topology and geometry. *Computer-Aided Design*, 41(12):1067–1081, 2009.

[9] M. J. Fonseca and J. A. Jorge. Experimental evaluation of an on-line scribble recognizer. *Pattern Recognition Letters*, 22(12):1311–1319, 2001.

[10] H. Freeman and R. Shapira. Determining the minimum-area encasing rectangle for an arbitrary closed curve. *Commun. ACM*, 18:409–413, July 1975.

[11] T. Hammond and R. Davis. Ladder: A language to describe drawing, display, and editing in sketch recognition. In *Proc. IJCAI-03*, pages 461–467, 2003.

[12] T. Hammond and R. Davis. LADDER, a sketching language for user interface developers. *Comput. Graph.*, 29:518–532, August 2005.

[13] T. Hammond and R. Davis. Creating the perception-based ladder sketch recognition language. In *Proc. DIS '10*, pages 141–150, New York, NY, USA, 2010. ACM.

[14] R. S. Tumen, M. E. Acer, and T. M. Sezgin. Feature extraction and classifier combination for image-based sketch recognition. In *Proc. SBIM '10*, pages 63–70. Eurographics, 2010.

[15] L. Zhang and Z. Sun. An experimental comparison of machine learning for adaptive sketch recognition. *Applied Mathematics and Computation*, 185(2):1138–1148, 2007.

Supporting Visual Information Extraction from Geospatial Data

Giuseppe Della Penna and Sergio Orefice
Department of Computer Science
University of L'Aquila, Italy
Email: {giuseppe.dellapenna,sergio.orefice}@univaq.it

Daniele Magazzeni
Department of Sciences
University of Chieti-Pescara, Italy
Email: magazzeni@sci.unich.it

***Abstract*—The Spatial Relation Query (SRQ) tool is a graphical software system, supported by a SQL-like query language, that enables users to perform information extraction driven by the visual appearance and the spatial arrangement of the information. The tool has been initially designed to support visual information extraction from web pages. Indeed, its former underlying spatial relation formalism relied on the bounding boxes of the graphical objects, which is a suitable choice for the web domain. In this paper we present a theoretical extension of the SRQ spatial composition framework that has been enhanced to work directly on the contours of the graphical objects. This allows us to apply the tool to more general contexts, such as GIS applications.**

I. Introduction

The process of knowledge acquisition from generic information domains has its central phase in the Information Extraction (IE), which aims to extract from the located documents relevant information that appear in certain semantic or syntactic relationships. In particular, IE tries to process the relevant information found on the documents in order to make it available to structured queries. Most often, information extraction systems are customized for specific application domains, and require manual or semi-automatic training sessions.

In [1] we have proposed a general IE approach based on the visual appearance of the information, conceived as its user-perceived rendering. This allows one to shift the IE problem from the low level of code (e.g., raster graphics, vector drawing, wordprocessor formatted text, web page, etc.) to the higher level of visual features, providing a paradigm of the kind "what you see drives your search" that supports a natural query formulation. The approach is based on the box spatial relation theory [2], such that graphical objects are syntactically described and manipulated through their bounding boxes whatever the shape of the object is, and on a SQL-like language, which allows users to write queries based on the visual arrangement of the information in an intuitive way. These formalisms have been implemented in a full-featured graphical software system, the SRQ (Spatial Relation Query) tool, which has been profitably used on a wide variety of applications within the IE web page domain. An early attempt to apply the tool to geospatial data has been proposed in [3]. However, in that work the underlying box spatial relation theory could deal with only very simple applications.

Other approaches in the recent literature that make visual information extraction include, for example, techniques that focus on specific application areas like, e.g., the work in [4] where the authors propose a machine learning methodology which allows one to automatically extract specific field of PDF documents, or the approaches based on visual web page analysis [5]–[8] which exploit the visual web page representation. However, the latter works focus on information extraction specifically targeted on tasks like record boundary detection [5], web page segmentation [7], visual web table extraction [6], or visual similarities detection (e.g., the recognition of repetitive patterns) [8]. Moreover, the major shortcoming of all the approaches above exploiting the visual appearance of information is the lack of an automatic counterpart supporting the visual information extraction.

In the last years, several studies have focused on the adoption of visual languages and visual techniques in many critical activities related to Geographic Information Systems (GIS) [9], ranging from design of geographic databases, to interoperability support, to decision-making support (see, e.g. [10]–[13]). Motivated by the increasing attention of researches aiming at bridging the gap between advanced geographic management techniques and practical problem-solving, in this paper we present the extension of the SRQ framework proposed in [1], [3] to the GIS context. To this aim, the spatial relation theory has been reformulated in order to work on polygonal contours. Indeed, the previous bounding box syntactical model has some intrinsic limitations that make it difficult to apply in frameworks where graphical objects are represented by complex figures, such as in GIS applications. The new theory has been embedded in the SRQ tool, which now is able to be experimented also on complex and meaningful applications in the geospatial data domain.

II. The Spatial Composition Framework

Before describing the new spatial relation formalism, let us briefly recall the basic notions of the *box* syntactical model as defined in [2]. In general, a syntactical model describes a family of visual languages based upon the nature of their graphical objects and composition rules, and it is formally defined by the quadruple (*graphical image*, *syntactical image*, *syntactical attributes*, *spatial relations*), where the first three components characterize the graphical objects by means of their *graphical* (namely the graphical image) and *logical* (namely the syntactical image and attributes) parts. In particular, the syntactical image is a suitable approximation of the graphical image that makes easier its syntactical manipulation, whereas the syntactical attributes are specific points on the syntactical image used by the spatial relations to compose graphical objects and form visual sentences.

Formally, in the box syntactical model, graphical objects have general geometric figures as their graphical images, and their bounding boxes as the corresponding syntactical images. The syntactical attributes of a graphical object are the upper left and lower right points of the bounding box , and the spatial relations are specified as follows.

Definition 1: Given a graphical object a with syntactical attributes (x, y) and (x_1, y_1), a generic spatial relation *REL*, with respect to a, defines two functions $UL_{REL}(x, y, x_1, y_1)$ and $LR_{REL}(x, y, x_1, y_1)$ that map the coordinates of the upper-left (x, y) and lower-right (x_1, y_1) points of the bounding box of a onto sets of points.

Then, given a graphical object b with syntactical attributes (h, k) and (h_1, k_1), we have that $aRELb$ holds if and only if $(h, k) \in UL_{REL}(x, y, x_1, y_1)$ and $(h_1, k_1) \in LR_{REL}(x, y, x_1, y_1)$.

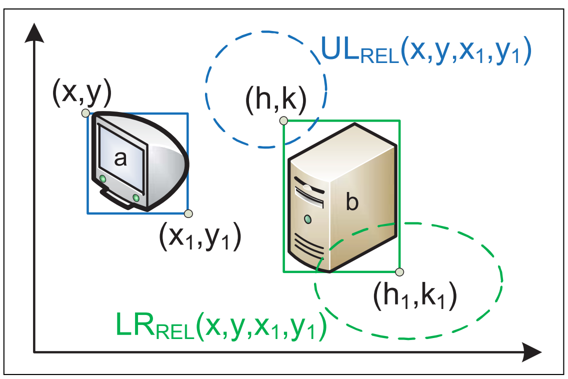


Fig. 1. A generic spatial relation *REL* in the box syntactical model

In other words, $aRELb$ holds if and only if the upper left and the lower right points of the bounding box of b are contained respectively within two areas calculated on upper left and lower right points of the bounding box of a through the functions UL_{REL} and LR_{REL} (see Figure 1).

Definition 1 yields three types of possible spatial arrangements: inclusion, intersection and spatial concatenation, where the term "concatenation" refers to any spatial arrangement of graphical objects not intersecting their areas. As an example, the relations *INTERSECT*, *INCLUDE*, *UP*, *DOWN*, *LEFT* and *RIGHT*, taken from [2], are defined in Table I. These relations model the general types of *spatial overlapping* and *spatial concatenation*.

So far, the box syntactical model has been used as the underlying formalism of the SRQ tool to support visual information extraction from web pages in [1] and from simple maps in [3]. However, this model has some intrinsic limitations, since the graphical objects are syntactically described and manipulated

TABLE I
DEFINITION OF GENERAL BOX SPATIAL RELATIONS

INTERSECT	$UL_{INTERSECT}(x, y, x_1, y_1) = \{(m, n)\|m \leq x_1, n \geq y_1\}$ $LR_{INTERSECT}(x, y, x_1, y_1) = \{(m, n)\|m \geq x, n \leq y\}$
INCLUDE	$UL_{INCLUDE}(x, y, x_1, y_1) = \{(m, n)\|x \leq m < x_1, y_1 < n \leq y\}$ $LR_{INCLUDE}(x, y, x_1, y_1) = \{(m, n)\|x < m \leq x_1, y_1 \leq n < y\}$
UP	$UL_{UP}(x, y, x_1, y_1) = \{(m, n)\|n < y\}$
DOWN	$LR_{DOWN}(x, y, x_1, y_1) = \{(m, n)\|n > y_1\}$
LEFT	$UL_{LEFT}(x, y, x_1, y_1) = \{(m, n)\|m > x_1\}$
RIGHT	$LR_{RIGHT}(x, y, x_1, y_1) = \{(m, n)\|m < x\}$

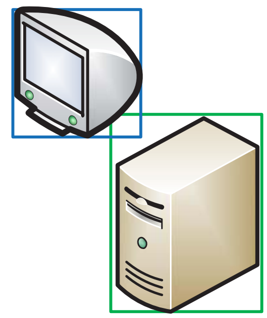


Fig. 2. An example of image intersection problem

through their bounding box, whatever the shape of the object is. As a consequence, it is difficult to apply it in frameworks where graphical objects are represented by complex figures, such as in general GIS applications. Indeed, in the box syntactical model, the syntactical image (i.e., the bounding box) of a graphical object matches its graphical image (i.e., its real shape), only for rectangular figures, whereas the bounding box for any other kind of object typically includes areas that are not actually part of the object itself (see, e.g., the objects in Figure 2). Due to this fact, the application of spatial relations may sometimes lead to results that are actually incorrect. For example, Figure 2 shows the typical *image intersection problem*, where the spatial relation *INTERSECT* would hold between the two (disjoint) shapes, since their bounding boxes actually intersect. To overcome this limitation, it is necessary to conceive the syntactical image of a graphical object as its graphical image. This idea is at the basis of the extension to new *contour syntactical model*.

A. The Contour Syntactical Model

The graphical objects of the visual languages characterized by the new model have "simple polygons" as their *graphical images*. In geometry, a polygon is a region of the plane bounded by a finite collection of line segments forming a closed curve. A simple polygon is a polygon where no pair of edges cross each other.

Figure 3(a) shows a simple polygon, whereas the contour in Figure 3(b) is an example of non simple polygon since it contains self loops.

Moreover, in order to support a more precise manipulation,

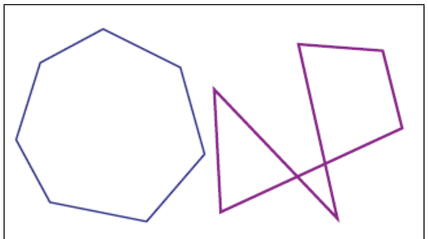


Fig. 3. Polygons

the logical part of graphical objects has to be appropriately extended. Thus, the other novelty of the contour syntactical model is in the *syntactical image* component of the graphical objects, which now coincides exactly with the contour of the graphical image.

Then, a natural choice for defining the graphical object syntactical attributes in the new model is to consider the smallest number of points of the syntactical image that allow one to completely enclose the graphical object (as it was also implicitly done in the box syntactical model with the upper left and lower right points).

Formally speaking, given a graphical object $O = (G, S, A, R)$ with syntactical image S, we write $p \in S$ to indicate a generic point of S and p_x, p_y to indicate the x and y-coordinate, respectively, of p. Moreover, we shall indicate with $UP(S)$, $DP(S)$, $LP(S)$ and $RP(S)$ the upmost, downmost, leftmost and rightmost point, respectively, of S. For instance, $UP(S)$ is formally defined as $p \in S | \forall p' \in S, p'_y \leq p_y$.

In particular, in the contour syntactical model, the syntactical attributes A of a graphical object with syntactical image S are the set of four points $UP(S)$, $DP(S)$, $LP(S)$ and $RP(S)$, when used by spatial concatenation relations, or the set of all the points of the contour S, in the case of inclusion and intersection relations. The corresponding setting of spatial relations is formally defined in the following subsection.

B. Contour Spatial Relations

The new characterization of the logical part of graphical objects in the contour model leads to the following reformulation of the spatial relations.

Definition 2: A generic spatial relation *REL* between two graphical objects a and b, with syntactical attributes A and A', respectively, is defined by means of the function $set_{REL}(A, p) = X \in P(R^2)$ with $p \in A'$ that associates each point in A' to a set of points X. Then, $a\ REL\ b$ holds iff $\forall p \in A', p \in set_{REL}(A, p)$.

In other words, each syntactical attribute of b must be contained in the corresponding set of points X calculated by the syntactical attributes of a, as appropriately defined by the function set_{REL}.

Let a and b be two graphical objects with syntactical image S and S' and syntactical attributes A and A', respectively. Then, in order to define the four basic concatenation spatial relations (which model the disjoint spatial arrangement) in the contour model, it is enough to consider $A = \{UP(S), DP(S), LP(S), RP(S)\}$ and $A' = \{UP(S'), DP(S'), LP(S'), RP(S')\}$, as shown in Table II. Note that, to shorten the notation, in the table we assume that set_{REL} returns R^2 if not differently specified.

TABLE II
DEFINITION OF BASIC CONCATENATION CONTOUR SPATIAL RELATIONS

UP	$set_{UP}(A, UP(S')) = \{(m,n) \| n < DP(S)_y\}$
DOWN	$set_{DOWN}(A, DP(S')) = \{(m,n) \| n > UP(S)_y\}$
LEFT	$set_{LEFT}(A, LP(S')) = \{(m,n) \| m > RP(S)_x\}$
RIGHT	$set_{RIGHT}(A, RP(S')) = \{(m,n) \| m < LP(S)_x\}$

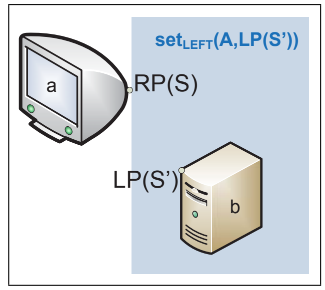


Fig. 4. Spatial relations *LEFT* in the contour syntactical model

For example, Figure 4 shows a visual arrangement where the *LEFT* relation holds between the objects a and b.

To complete the set of basic contour spatial relations, let us define the inclusion and intersection relations that model the overlapping of graphical objects. Let a and b be two graphical objects with syntactical image S and S' and syntactical attributes A and A', respectively. In order to define the two basic relations *INCLUDE* and *INTERSECT* in the contour model, we have to consider $A = S$ and $A' = S'$. Then, we define such relations as shown in Table III.

These two relations use the auxiliary relation $CONTAIN(p)$, defined through the function $set_{CONTAIN(p)}(A, p) = Area(S)$, where $Area(S)$ is the set of internal points of the syntactical image S. In other words, $a\ CONTAIN(p)\ b$ iff the point $p \in S'$ is inside the area of S.

Thus, a graphical object a includes another graphical object b iff all the points of the contour of b are enclosed by the contour of a. On the other hand, a graphical object a intersects another graphical object b iff at least one of the points of the contour of b is enclosed by the contour of a.

III. The SRQ Tool

The overall organization of the SRQ software system is shown in Figure 5. The current prototype application has been developed in Java with the support of *ANTLR* [14] for the spatial relation compiler and query interpreter. The application

TABLE III
DEFINITION OF THE OVERLAPPING CONTOUR SPATIAL RELATIONS

$a\ INCLUDE\ b$	$\Leftrightarrow$	$\forall p \in S' : a\ CONTAIN(p)\ b$
$a\ INTERSECT\ b$	$\Leftrightarrow$	$\exists p \in S' : a\ CONTAIN(p)\ b$

```
SELECT [FIRST | LAST |
  ( CLOSEST [LEFT | RIGHT | UP | DOWN]
  '(' <variable> ')' )]
  CONTOUR | <property> [',' <property>...]
FROM <uri>
WHERE <boolean expression using spatial relations and
  the CONTOUR variable>
HAVING <boolean expression on property values>
ORDER BY <property>
  | XPOS (UV | DV | LV | RV)
  | YPOS (UV | DV | LV | RV)
  | ZPOS (UV | DV | LV | RV)
  [ASC | DESC] [ ',' ... ]
WITH <variable> '='
  [<uri>|<coordinates>|<sub-query>]
  [',' <variable> '=' ...]
```

Fig. 6. Overall structure of a SRQL query

uses a plugin system to interface with several different sources and build the corresponding *visual data model*.

In particular, the SRQ tool is composed of two interacting modules: the *Spatial Relation Editor* and the *Spatial Relation Query Executor*. These modules share the access to a *spatial relation library* which contains the basic building blocks of the query language.

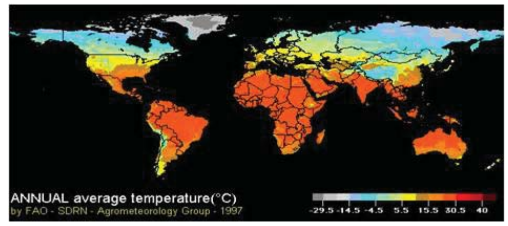
The Spatial Relation Editor module allows one to create and manage the spatial relation library, which defines the spatial relations used in the queries. The library itself is written in an appropriate *Spatial Relation Definition Language* (SRDL) that, for sake of simplicity, will not be described here.

The Spatial Relation Query Executor allows one to compose and execute queries. The composition of queries can be accomplished through the *Spatial Relation Query Language* (SRQL) language, which exploits the spatial relation theory given in Sections II-A, II-B to create queries based on the visual arrangement of the data and is designed to be as similar as possible to the well known SQL.

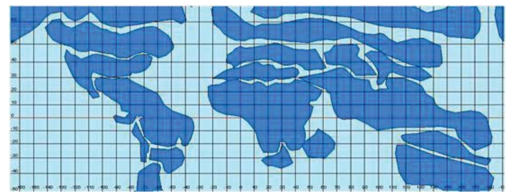
Before applying an SRQL query, the tool analyzes the information resource and generates the corresponding *user image*, i.e., an image which shows how the information actually looks like to the user. Then, it builds an appropriate visual data model, namely the *contour model*, that implements the notion of graphical object as the 4-tuple formally defined in Section II-A. This model encapsulates all the distinguished objects from the visual appearance, possibly associating a set of *attributes*, deducible from the source information, to each object. Depending on the particular IE data domain, such attributes characterize specific properties of the objects, which may be also visual (e.g., color, contained text, etc.).

The basic structure of a SRQL query is shown in Figure 6. More details on the syntax of the clauses composing a query can be found in [3].

Once a query has been written, it can be saved for later reuse, and analyzed/executed. Indeed, the system first performs some syntax checks on the query, possibly providing meaningful error messages that allow the user to improve the query definition, and then the query is executed on the contour model. After a successful query, the user image of the data is rendered in the application interface, the graphical



(a) Original map



(b) Traced shapefile

Fig. 7. Average temperatures

objects selected by the query (if any) are highlighted, and the corresponding attributes are appropriately returned. Finally, it is worth noting that the query results can be exported in an XML file to be exploited for further analysis and manipulations.

IV. A CASE STUDY

In this section we describe an application of the SRQ tool on real-world GIS data concerning the malaria diffusion. Malaria is an infectious disease widespread in tropical and subtropical regions, including parts of the Americas, Asia, and Africa. Each year, in these regions, malaria kills millions of people.

It is well known that the distribution of this disease is closely related to climatic and environmental factors like temperature, humidity and rainfall. Indeed, malaria is naturally transmitted by the bite of a female Anopheles mosquito which lives in wet and warm places.

First of all, we use SRQ to highlight such a relation by combining and analyzing climate data to find which nations are potentially at risk of malaria. The results of this analysis are then validated against the map shown in Figure 9(b), which illustrates the actual diffusion of the disease as reported by the World Health Organization [15].

In particular, the data sources for this case study are represented by the three maps taken from FAO [16] which contain, respectively, the average humidity, annual rainfall and temperature in the world. As an example, Figure 7(a) shows the temperature map. We use the MapWindow [17] software to trace such maps and build the corresponding shapefiles. Figure 7(b) shows the temperature shapefile, where we report the most significant map areas, associating each of them with the semantic data given by the corresponding legend.

Finally, these shapefiles have been layered over the world nations shapefile to obtain the composite map to be used as the source for the query to find the nations at risk of malaria,

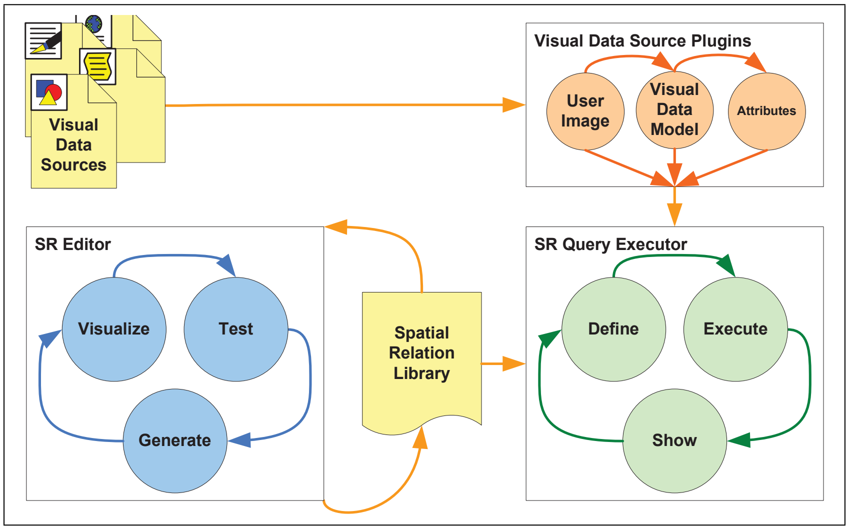
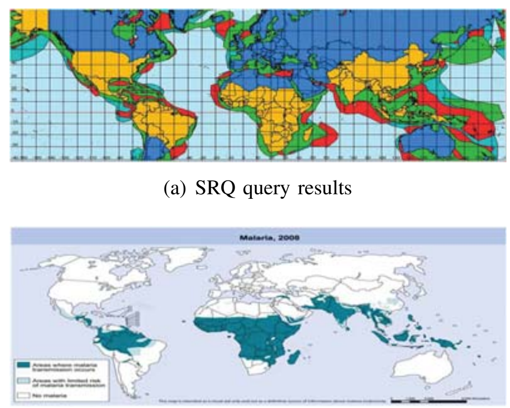


Fig. 5. SRQ tool architecture

```
SELECT *
 FROM local:maps
 WHERE CONTOUR INTERSECT ANY rainyArea
 HAVING layer = 'world'
 WITH rainyArea = SELECT CONTOUR
  FROM local:maps
  WHERE CONTOUR INTERSECT ANY warmArea
  HAVING layer = 'rain' AND rain_mm >= 1500
  WITH warmArea = SELECT CONTOUR
   FROM local:maps
   WHERE CONTOUR INTERSECT ANY humidArea
   HAVING layer = 'temperature' AND
       degrees_avg >= 20
   WITH humidArea = SELECT CONTOUR
   FROM local:maps
   HAVING layer = 'humidity' AND rate = 5
```

Fig. 8. SRQL query to find the world nations at risk of malaria



(a) SRQ query results

(b) WHO data

Fig. 9. Areas at risk of malaria due to the current climatic conditions

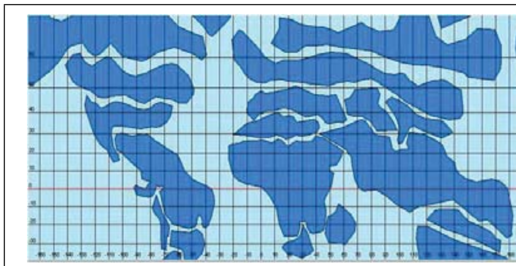


Fig. 10. Expected temperature shapefile

which is formulated in the SRQL language as illustrated in Figure 8.

The query uses the *INTERSECT* relation to intersect the world nations with the areas, taken from the corresponding layers, that represent rainy (i.e., average annual rain greater than or equal to 1500mm), warm (average temperature above 20C) and humid (humidity index 5) regions. In particular, each of these classes of regions is extracted by a subquery which works on the corresponding map layer.

The query results are shown in Figure 9(a), where the extracted areas are highlighted in yellow. Such results are very similar to the WHO data in Figure 9(b). The only visible difference is that the SRQ ones include more regions of the United States than those marked in the WHO map. However, we should expect this discrepancy, since the southern regions of the United States would be actually exposed to malaria, but the social and economic conditions of such regions (which were not considered by the query) allowed to eradicate the disease through an adequate health care and environmental control.

Encouraged by the positive validation of the SRQ analysis with respect to the current climatic data, we can try to obtain a forecast of the future diffusion of malaria. Indeed, the well known *global warming* is constantly changing the climatic conditions of many world areas, and this may affect the

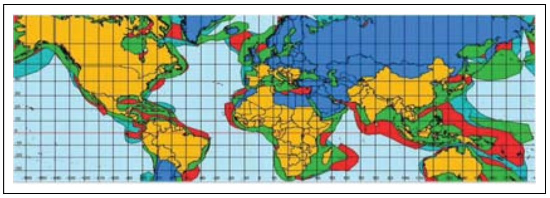


Fig. 11. Expected malaria diffusion due to global warming

incidence of malaria. This forecast actually represents the valuable contribution of SRQ for this case study.

To show what may happen, we have taken a map published by the UK's National Weather Service [18], which reports the estimated temperature growth in the future. We combined the data from this map with the ones in the temperatures shapefile (Figure 7(b)), to obtain a new shapefile shown in Figure 10, which reports the expected future average temperatures in the world. Finally, we have run again the query in Figure 8 using this modified data together with the same "rain" and "humidity" layers.

The query results, highlighted in yellow in Figure 11, indicate that the malaria could appear in the future in some areas of the southern Europe as well as in Canada and Australia. Interestingly, there have been actually some cases of malaria reported in the last years in Canada [19], [20] and Australia [21].

V. Conclusions

In this paper we have presented an extension of the *bounding box spatial relation formalism* in order to work directly on polygonal contours. The new theory has been also implemented within the SRQ tool, a graphical software system that provides an automatic support to the visual information extraction from different sources. In particular, the SRQ tool is suitable to be used both by novel or experienced users. The former users may take advantage of the user-friendly interface and the natural query paradigm of the tool to accomplish their searches. The latter ones may fully exploit the integrated power of the tool and of the SRQL language by both appropriately customizing the spatial relations and extensively using visual, textual and structural constraints in the queries.

So far, the SRQ tool has been mainly applied to web case studies. Thanks to the new *contour spatial relation formalism*, we are experimenting the extended version of the tool on geospatial applications. Other than extensively using the tool in GIS applications, in the future we also plan to apply it to new, challenging domains like, e.g., biomedical applications. Indeed, the contour model is suitable to correctly represent the complex spatial relations existing among the human body organs, and then the SRQ tool may be of help for diagnostic analysis as well as for medicine and surgery didactic activities.

Moreover, we are investigating some extensions of the spatial relation theory in order to capture further aspects of particular visual data sources, e.g., the notion of *time*. For example, it could be interesting to make queries that address the evolution of GIS datasets, i.e., to reason about *spatial-temporal data*.

References

[1] G. Della Penna, D. Magazzeni, and S. Orefice, "Visual extraction of information from web pages," *Journal of Visual Languages and Computing*, vol. 21, no. 1, pp. 23–32, 2010.

[2] G. Costagliola, A. De Lucia, S. Orefice, and G. Polese, "A classification framework to support the design of visual languages," *Journal of Visual Languages and Computing*, vol. 13, no. 6, pp. 573–600, 2002.

[3] G. Della Penna, D. Magazzeni, and S. Orefice, "A graphical tool to support visual information extraction," in *Proceedings of DMS 2009 (VLC workshop)*. Knowledge Systems Institute, 2009, pp. 329–334.

[4] Y. Aumann, R. Feldman, Y. Liberzon, B. Rosenfeld, and J. Schler, "Visual information extraction," *Knowledge and Information Systems*, vol. 10, no. 1, pp. 1–15, 2006.

[5] H. Zhao, W. Meng, Z. Wu, V. Raghavan, and C. Yu, "Fully automatic wrapper generation for search engines," in *WWW '05: Proceedings of the 14th international conference on World Wide Web*. New York, USA: ACM, 2005, pp. 66–75.

[6] W. Gatterbauer, P. Bohunsky, M. Herzog, B. Krüpl, and B. Pollak, "Towards domain-independent information extraction from web tables," in *WWW '07: Proceedings of the 16th international conference on World Wide Web*. New York, USA: ACM, 2007, pp. 71–80.

[7] X. Gu, J. Chen, W. Ma, and G. Chen, "Visual based content understanding towards web adaptation," in *Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems AH'02. LNCS*, vol. 2347. Springer-Verlag, 2002, pp. 164–173.

[8] Y. Yang and H. Zhang, "HTML page analysis based on visual cues," in *ICDAR '01: Proceedings of the Sixth International Conference on Document Analysis and Recognition*. Washington, DC, USA: IEEE Computer Society, 2001, pp. 859–864.

[9] M. Sebillo, G. Tortora, and G. Vitiello, "Introduction to the special issue on visual languages and techniques for human-gis interaction," *Journal of Visual Languages and Computing*, vol. 18, no. 3, pp. 227–229, 2007.

[10] V. M. B. Cavalcanti, U. Schiel, and C. de Souza Baptista, "Querying spatio-temporal databases using a visual environment," in *Proc. AVI '06*. ACM Press, 2006, pp. 412–419.

[11] E. Giaccardi and D. Fogli, "Affective geographies: Toward a richer cartographic semantics for the geospatial web," in *Proc. AVI '08*. ACM Press, 2008, pp. 173–180.

[12] Z. Sha and M. Zhang, "Development of web-based decision support system for field-based crop management," in *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*. New York, NY, USA: ACM, 2007, pp. 1–4.

[13] G. Andrienko, N. Andrienko, P. Jankowski, D. Keim, M. J. Kraak, A. MacEachren, and S. Wrobel, "Geovisual analytics for spatial decision support: Setting the research agenda," *Int. Journal of Geographical Information Science*, vol. 21, no. 8, pp. 839–857, 2007.

[14] ANTLR Parser Generator, "http://www.antlr.org/."

[15] World Health Organization, "Malaria, http://www.who.int/malaria."

[16] FAO, "Global climate maps, http://www.fao.org/WAICENT/FAOINFO/SUSTDEV/EIdirect/CLIMATE/EIsp0002.htm."

[17] Idaho State University Geospatial Software Lab, "Mapwindow open source gis, http://www.mapwindow.org/."

[18] UK National Weather Service, "Climate projections - surface air temperature - annual, http://www.metoffice.gov.uk/climatechange/science/projections/air_annual.html."

[19] P. Wilton, "Malaria may be on move to "tropical" canada," *Canadian Medical Association Journal*, vol. 158, no. 2, 1998.

[20] J. Laliberte, "Global warming crisis hits fever pitch," *National Review of Medicine*, vol. 3, no. 1, 2006, http://www.nationalreviewofmedicine.com/issue/2006/01_15/3_patients_practice01_1.html.

[21] B. Malkin, "Global warming will cause malaria epidemic in australia and pacific islands," *Telegraph.co.uk*, 2008, http://www.telegraph.co.uk/earth/environment/globalwarming/3490944/Global-warming-will-cause-malaria-epidemic-in-Australia-and-Pacific-Islands.html.

A More Expressive 3D Region Connection Calculus

Chaman L. Sabharwal, Jennifer L. Leopold, Nathan Eloe

Department of Computer Science
Missouri University of Science and Technology
Rolla, MO - 65409, USA
chaman@mst.edu, leopoldj@mst.edu, nwe5q8@mst.edu

Abstract Spatial qualitative analysis in 3D provides engineers with the ability to visually evaluate interference between parts, complicated geometric features, required tolerances, mechanism motion, and various aspects of assembly structure. Yet, despite the fact that 3D technology has been well developed for several years, many mechanical engineering designers currently do not use 3D CAD technology, and research in qualitative spatial reasoning continues to focus on 2D perspectives of 3D objects. In part, this is due to human cognitive limitations and the increased computational complexity of 3D modeling. However, given the greater wealth of information that can be obtained when additional dimensions are considered, and the enormous amount of 3D image data that are available, research efforts in this area should continue. Herein we present an extension of our previous work to combine quantitative and qualitative reasoning in 3D. The enhanced model is more expressive in that it now distinguishes several types of obscuration. Further, it differs from other region connection calculi used for qualitative spatial reasoning in that it determines the composite (topological plus obscuration) relation between two objects in 3D, not 2D.

Keywords-qualitative spatial reasoning; region connection calculus; 3D

I. Introduction

Computational spatial reasoning is useful for many applications, such as geographic information systems, image processing, and robotics. As with any knowledge representation and reasoning system, having additional dimensions of information often facilitates more accurate and informative analyses. Yet research in and implementations of qualitative spatial reasoning mostly have been limited to 2D, even when higher dimensions of data are available. In part, this is due to human cognitive limitations and the increased computational complexity that results when additional information is considered. However, given the enormous amount of 3D image data that continue to become available, as well as the increasing sophistication of mechanical and software systems that require spatial awareness, it is important that efforts to produce 3D spatial reasoning systems continue.

Herein we present an extension of our previous efforts to combine quantitative and qualitative reasoning in 3D [1]. This new model increases the expressive power by considering additional types of obscuration. Further, it differs from other region connection calculi in that it determines the composite (topological plus obscuration) relation between a pair of objects in 3D, not just 2D.

II. Related Work

A. Region Connection Calculi

Much of the foundational research on qualitative spatial reasoning is based on a region connection calculus that describes 2D regions by their possible relations to each other. Most notable are the RCC-5 and RCC-8 models [2, 3]. The RCC-5 relations are: *disconnected* (DC), *partial overlap* (PO), *equal* (EQ), *proper part* (PP), and *proper part inverse* (PPi). RCC-8 differs from RCC-5 in that it adds a relation for *externally connected* (EC), distinguishes PP as two relations, *tangential proper part* (TPP) and *non-tangential proper part* (NTPP), and separates the inverse relation PPi into TPPi and NTPPi.

RCC-23 [4] extends RCC-8 to 23 relations in order to accommodate concave regions in 2D. This expanded set of relations is based on primitives for connection and convex hull (i.e., primitives for *inside*, *partially inside*, *outside*, and an inverse for each of those three relations). In addition to the PO, EQ, TPP, TPPi, NTPP, and NTPPi, relations of RCC-8, RCC-23 distinguishes eight relations for DC (based on the aforementioned convexity predicates), and nine relations for EC (also based on the convexity predicates).

RCC-62 [5] is even more expressive than RCC-23; whereas RCC-23 considers a concave region as one whole part, RCC-62 decomposes such a region into an outside, boundary, interior, and inside. The resulting 62 relations are based on a 16-Intersection that compares one object's outside, boundary, interior, and inside with those of another object. However, like RCC-8 and RCC-23, RCC-62 only describes the relationship between regions considering two dimensions.

In contrast to RCC-8, RCC-23, and RCC-62, the LOS-14 [6] and ROC-20 [7] models qualify the spatial relation between 2D regions in terms of the obscuration that occurs between them. LOS-14 differs from ROC-20 in that it is restricted to objects that do not overlap in 3D. The 14 relations of LOS-14 describe the spatial relationship between two 2D objects in terms of whether or not one completely or partially hides the other using a qualitative depth relation between the objects; one object can be in front of another, but the two cannot be of equal depth. ROC-20 extends LOS-14 by allowing objects to be concave (and hence accommodates mutual occlusion, adding 6 relations to the LOS-14 set). ROC-20 defines all spatial relations in terms of a combination of occlusion (i.e., *NonOccludes, PartiallyOccludes, MutuallyOccludes,* and *TotallyOccludes*) and, as applicable, an RCC-8 relation.

B. RCC-3D

In the aforementioned RCC-based models, the complete spatial relationship, not just the degree of obscuration, is only with respect to a particular 2D viewpoint. Over the past two years, the authors have investigated a different approach for spatially reasoning over 3D objects. An extensive discussion of that early model (named RCC-3D) is beyond the scope of this paper; see [1] for those details. Briefly, RCC-3D defines 13 spatial relations based on an 8-Intersection of the 3D interior, exterior, and boundary of the two objects, and a 3-Intersection of the 2D interior and exterior of a particular projection of the two objects in the view reference plane. Employing a naming convention similar to that used for RCC-8, although using *c* to represent *converse* rather than *i* for *inverse*, the 13 RCC-3D relations are: DC, DC_pp, DC_p, EC, EC_Pp, EC_P, PO_Pp, PO_P, TPP_P, TPP_Pc, EQ_P, $NTPP_P$, and $NTPP_P$c; the subscript *P* in the relation name represents the particular 2D projection plane, and partial obscuration (as opposed to complete obscuration) is indicated by adding a (non-subscripted) *p* to the end of the relation name.

As has been done for other region connection calculi, a conceptual neighborhood graph (CNG) was defined for RCC-3D to identify those transitions that can occur when the geometry of one object in a pair is changed gradually. For RCC-3D the topological distance between relations is computed as the number of intersections that change from empty to not empty (or vice versa), from one relation to another. This distance is expressed using the format *x+y* where *x* is computed by considering the 8-Intersection in 3D (hence, the inter-relation distance), and *y* is calculated based on the 3-Intersection in 2D projection plane P (the intra-relation distance).

A Prolog program was written to generate a composition table (CT) for RCC-3D, providing the ability to answer questions such as, "given three objects, A, B, and C, and knowledge of relations R_1(A, B), and R_2(B, C), what can be said about the relation between A and C." A prototype RCC-3D reasoner was then implemented, and tested to compare automated determination of spatial relationships between 3D reconstructions of anatomy (in OBJ file format) against the determination of experts who had examined the original specimens manually. The preliminary results [1] were very promising, and work commenced to create an application that would allow the user to check for inconsistency in relations between abstract temporal "states" of the objects, and determine possible relations that would have had to occur to transition from one state to another; the integration of that functionality together with a visual user interface was named VRCC-3D [8].

III. VRCC-3D+

Although the early RCC-3D and VRCC-3D models showed potential for studies involving spatial data, the authors realized that ambiguous analyses could occur in certain situations. As shown in Fig. 1, it is not possible to determine intersection or obscuration in 3D from only the 2D projections; that is, we cannot say that, just because the projections intersect, the objects intersect in 3D space. More precisely, in VRCC-3D, if we know R_1(A, B), R_2(B, C), Obs_1(A, B, P), and Obs_2(B, C, P) (where R_i is one of the spatial relations and Obs_i is one of the obscuration relations), then we can only determine R_3(A, C) and Obs_3(A, C, P); we cannot tell whether A is in front of C, or C is in front of A. In projections, the information about the distance from the projection plane is lost.

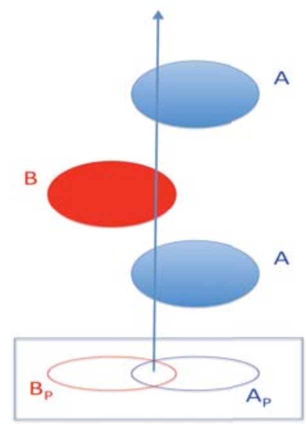


Figure 1. Shown are two configurations of object A relative to object B. From the intersection of projections A_P and B_P (where P is a 2D projection plane), it is not possible to determine: (1) if A and B intersect in 3D space, and (2) if A is in front of B, or B is in front of A.

To address this problem, the enhanced model described herein, VRCC-3D+, uses a depth parameter to distinguish additional types of obscuration. The result is a more expressive region connection calculus; the number of obscuration relations increases from 2 to 11, and the number of composite relations increases from 13 to 34 from VRCC-3D to VRCC-3D+, respectively.

A. Characterization of Base Relations

As with our earlier models, the eight base relations of VRCC-3D+ (which comprise the entire set of relations for RCC-8) are distinguished based on an 8-Intersection; see Table I. This is similar to the 9-Intersection presented in [9], with two notable exceptions: (1) the intersection between the exterior of one object and the exterior of another object is excluded since it will be non-empty for any pair of bounded objects (i.e., the intersection of the exteriors of two objects is not informative), and (2) the intersection predicates are computed for the 3D (not 2D) interior, boundary, and exterior of the objects.

	IntInt	IntBnd	IntExt	BndInt	BndBnd	BndExt	ExtInt	ExtBnd
DC	∅	∅	¬∅	∅	∅	¬∅	¬∅	¬∅
EC	∅	∅	¬∅	∅	¬∅	¬∅	¬∅	¬∅
EQ	¬∅	∅	∅	∅	¬∅	∅	∅	∅
NTPP	¬∅	∅	∅	¬∅	∅	∅	¬∅	¬∅
NTPPc	¬∅	¬∅	¬∅	∅	∅	¬∅	∅	∅
PO	¬∅	¬∅	¬∅	¬∅	¬∅	¬∅	¬∅	¬∅
TPP	¬∅	∅	∅	¬∅	¬∅	∅	¬∅	¬∅
TPPc	¬∅	¬∅	¬∅	∅	¬∅	¬∅	∅	∅

TABLE I. 8-Intersection Characterization Table (Int = Interior, Bnd = Boundary, Ext = Exterior; ¬∅= non-empty intersection, ∅ = empty intersection)

B. *Characterization of Obscuration*

Like ROC-20, VRCC-3D+ distinguishes non-occlusion, partial occlusion, and complete occlusion. However, for the VRCC-3D+ model it is assumed that all objects are opaque, so the ROC-20 notion of mutual occlusion (e.g., A obscures B, and B obscures A from a particular view reference point) is not supported. Instead VRCC-3D+ defines an equal occlusion, whereby the qualitative distance between two objects is represented with a predicate, *InFront,* that is computed with respect to a particular 3D viewpoint. Hence, the obscuration part of a VRCC-3D+ relation is determined by examining the three intersections between the 2D interior and exterior of each object (as had been done for the earlier model, VRCC-3D). But now an additional fourth parameter is incorporated into those definitions to represent the relative qualitative depths of the objects.

The names of the VRCC-3D+ obscuration types are of the form *xObsy*, where *Obs* refers to the intersection between the projections. Possible values for the prefix *x* are: *n* (no occlusion), *p* (partial occlusion), *e* (equal occlusion), and *c* (complete occlusion). The suffix *y* in the relation name refers to the value of *InFront*. When comparing two objects A and B, the suffix is blank when A is in front of B; a suffix of *_e* indicates that the two objects are equidistant from the view plane; the suffix *_c* indicates that B is closer to the view plane than A. There are a total of eleven occlusion relations, which are characterized in Table II. It should be noted that the particular 2D projection plane is not specified for each obscuration type listed in Table II because the results would be the same regardless of the reference plane; this will be the case for all tables that subsequently appear herein.

Not every type of obscuration is applicable to every base relation in VRCC-3D+. For example, it is not possible to have two objects, A and B, where A partially overlaps B in 3D space, but the projection of A does not overlap the projection of B for any viewpoint. All the possible combinations comprise a set of 34 relations for VRCC-3D+, as shown in Table III.

	IntInt	IntExt	ExtInt	InFront
pObs	¬∅	{∅,¬∅}	¬∅	Y
pObs_c	¬∅	¬∅	{∅,¬∅}	N
pObs_e	¬∅	¬∅	¬∅	E
cObs	¬∅	¬∅	∅	Y
cObs_c	¬∅	∅	¬∅	N
eObs_e	¬∅	∅	∅	E
eObs_c	¬∅	∅	∅	N
eObs	¬∅	∅	∅	Y
nObs_e	∅	¬∅	¬∅	E
nObs_c	∅	¬∅	¬∅	N
nObs	∅	¬∅	¬∅	Y

TABLE II. 3-Intersection + Depth Characterization Table (Int = Interior, Ext = Exterior; ¬∅= non-empty intersection, ∅ = empty intersection; Y = yes, N = no, E = equal)

	nObs	nObs_c	nObs_e	pObs	pObs_c	pObs_e	eObs	eObs_c	eObs_e	cObs	cObs_c
DC	X	X	X	X	X		X	X		X	X
EC	X	X	X	X	X		X	X		X	X
PO				X	X	X	X	X		X	X
EQ									X		
TPP								X	X		X
TPPc							X		X	X	
NTPP											X
NTPPc										X	

TABLE III. Possible Obscuration Types for Each VRCC-3D+ Base Relation

Figure 2 shows three examples, further illustrating some of the possible combinations of the VRCC-3D+ partial overlap relation PO (as computed in 3D), and various types of obscuration (for a particular 2D projection *p*).

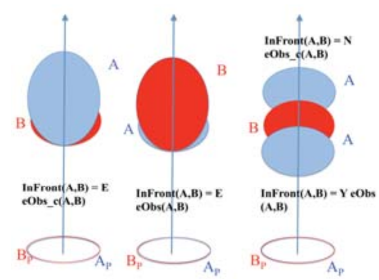


Figure 2. Examples of Partial Overlap (PO) in VRCC-3D+

C. *The Conceptual Neighborhood Graph*

As previously mentioned, a conceptual neighborhood graph (CNG) represents the possible transitions from one relation to another. Expansion of the set of relations for VRCC-3D+ necessitated the construction of a new CNG.

Table IV shows all the inter-relation distances, computed in the same manner as was discussed for the earlier model, VRCC-3D.

	DC	EC	PO	EQ	TPP	TPPc	NTPP	NTPPc
DC	0	1	4	6	5	5	4	4
EC	1	0	3	5	4	4	5	5
PO	4	3	0	6	3	3	4	4
EQ	6	5	6	0	3	3	4	4
TPP	5	4	3	3	0	6	1	7
TPPc	5	4	3	3	6	0	7	1
NTPP	4	5	4	4	1	7	0	6
NTPPc	4	5	4	4	7	1	6	0

TABLE IV. Inter-relation Distances for Base Relations

However, the calculation of intra-relation distance for VRCC-3D+ is different since it now includes the qualitative relative distance between a pair of obscuration types. An example of the computation is as follows. As displayed in Table V, since pObs is valid when IntExt is empty or non-empty, the IntExt contribution to the distance between pObs and cObs is 1 if they are different, or 0 if they are the same. In order to cover uncertainty (not knowing which occurs in actual instance), we use a distance of 0.5, the average of 0 and 1, for matching {∅, ¬∅} with {∅} or {¬∅}. For *InFront,* Y represents "closer", N represents "farther", and E ("equal") is midway between; thus, it was decided to use 1 for the difference between Y and N, and 0.5 for the difference between E and Y, or E and N. The conceptual distance for each pair of obscuration types is given in Table VI.

	IntInt	IntExt	ExtInt	InFront
pObs	¬∅	{∅,¬∅}	¬∅	Y
cObs	¬∅	¬∅	∅	Y
cObs_c	¬∅	∅	¬∅	N
nObs_e	∅	¬∅	¬∅	E

Table V. Predicates Used to Compute Intra-Relation Distance (e.g., distance between pObs and cObs is 1.5, cObs and cObs_c is 3, and cObs_c and nObs_e is 2.5)

	nObs	nObs_c	nObs_e	pObs	pObs_c	pObs_m	eObs	eObs_c	eObs_e	cObs	cObs_c
nObs	0	1	0.5	1.5	2.5	1.5	3	4	3.5	2	3
nObs_c	1	0	0.5	2.5	1.5	1.5	4	3	3.5	3	2
nObs_e	0.5	0.5	0	2	2	1	3.5	3.5	3	2.5	2.5
pObs	1.5	2.5	2	0	2	1	1.5	2.5	2	1.5	1.5
pObs_c	2.5	1.5	2	2	0	1	2.5	1.5	2	1.5	1.5
pObs_e	1.5	1.5	1	1	1	0	2.5	2.5	2	1.5	1.5
eObs	3	4	3.5	1.5	2.5	2.5	0	1	0.5	1	2
eObs_c	4	3	3.5	2.5	1.5	2.5	1	0	0.5	2	1
eObs_e	3.5	3.5	3	2	2	2	0.5	0.5	0	1.5	1.5
cObs	2	3	2.5	1.5	1.5	1.5	1	2	1.5	0	3
cObs_c	3	2	2.5	1.5	1.5	1.5	2	1	1.5	3	0

TABLE VI. Intra-Relation Distances for Obscuration Types

The row and column headings of Table VII correspond to a partial set of the VRCC-3D+ relations; the complete table is too large to display herein. The entries in the table are symmetric. Each relation is within three of the conceptual distance to its nearest neighbors. Six relations are within 0.5, eleven are within 1+0, and fifteen are within 0+1. Eight edges of length 0+1.5 and three edges of length 3+0 also were selected to yield a connected graph of conceptual neighbors for the set of VRCC-3D+ relations.

A graphic depiction of the CNG is shown in Figure 3. The nodes in the CNG are grouped vertically by their base relations, and horizontally by their possible obscuration relations. This arrangement facilitates the visual distinction between the inter-relation distance and the intra-relation distance. All edges in the graph are undirected.

It should be noted that there are different methods for constructing a CNG, and not all such methods will produce the same graph; the number of edges may vary depending upon the construction method utilized. Furthermore, depending upon the semantics of the particular set of relations being modeled, some construction methods will not produce a useful CNG. For example, the Snapshot Model approach [10] could not be utilized (without significant modifications) for VRCC-3D+. Application of that algorithm resulted in a collection of disconnected subgraphs, some of which contained nodes for relations that were more closely related based on inter-relation topological distance, while others contained nodes for relations that were more closely related based on intra-relation topological distance. As future work, other approaches for constructing the CNG will be considered and analyzed to determine any benefits over the All-Pairs-Shortest-Path (graph) approach that the authors did use.

		PO	PO	EQ	TPP	TPP	TPPc	TPPc	TPPc	NTPP	NTPPc
		pObs	eObs	eObs_e	eObs_c	eObs_e	eObs	eObs_e	cObs	cObs_c	cObs
PO	pObs	0	0+1.5	6+2	3+2.5	3+2	3+1.5	3+2	3+1.5	4+1.5	4+1.5
PO	eObs	0+1.5	0	6+0.5	3+1	3+0.5	3+0	3+1	3+1	4+2	4+1
EQ	eObs_e	6+2	6+0.5	0	3+0.5	3+0	3+0.5	3+0	3+1.5	4+1.5	4+1.5
TPP	eObs_c	3+2.5	3+1	3+0.5	0	0+0.5	6+1	6+0.5	6+2	1+1	7+2
TPP	eObs_e	3+2	3+0.5	3+0	0+0.5	0	6+0.5	6+0	6+1.5	1+1.5	7+1.5
TPPc	eObs	3+1.5	3+0	3+0.5	6+1	6+0.5	0	0+0.5	0+1	7+2	1+1
TPPc	eObs_e	3+2	3+0.5	3+0	6+0.5	6+0	0+0.5	0	0+1.5	7+1.5	1+1.5
TPPc	cObs	3+1.5	3+1	3+1.5	6+2	6+1.5	0+1	0+1.5	0	7+3	1+0
NTPP	cObs_c	4+1.5	4+2	4+1.5	1+1	1+1.5	7+2	7+1.5	7+3	0	6+3
NTPPc	cObs	4+1.5	4+1	4+1.5	7+2	7+1.5	1+1	1+1.5	1+0	6+3	0

TABLE VII. Partial Table of Minimum Distance Edges

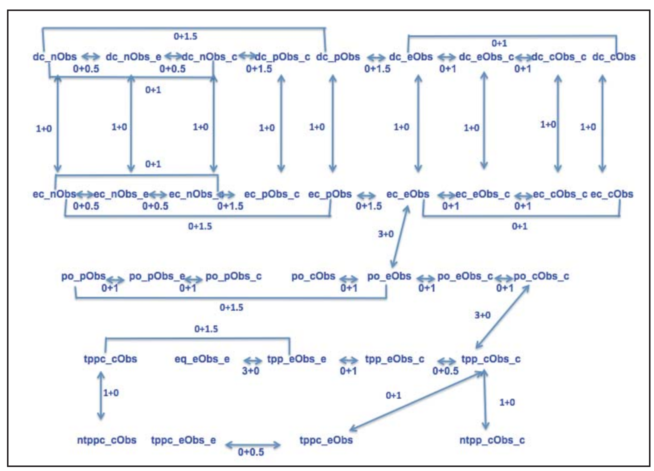


Figure 3. VRCC-3D+ Conceptual Neighborhood Graph

D. *The Composition Table*

Given the VRCC-3D+ relation R_1 between A and B, and the relation R_2 between B and C, we can determine R, the set of all "possible" composite (spatial and occlusion) relations for A and C. First, a composition table (CT) for only the base relations is computed, as is a CT for only the obscuration relations. These tables are then integrated based on which types of obscurations are possible for each spatial relation. Due to space limitations, the complete VRCC-3D+ composition table is not displayed here.

A simplified real-world example illustrates the potential usefulness of the CT, as well as the benefits of the more informative spatial knowledge that is considered by VRCC-3D+ (which is not present in other RCC-based models). Suppose that three airplanes, A, B, and C, of equal size are flying in the sky. If the air controller only knows that A and B are in disjoint regions, and that B and C are in disjoint regions, then the RCC-8 CT entry for A and C is universal (i.e., it could be any possible RCC-8 relation); hence, there is no guarantee that A and C cannot crash. As shown in Table VIII, even the corresponding entries in the VRCC-3D CT for this configuration cannot provide that assurance. However, additional depth information relative to the viewer, such as A is farther away from B, and B is farther away from C, is taken into consideration in VRCC-3D+. In this particular situation, the VRCC-3D+ CT entry for A and C is a single relation (i.e., A and C are disconnected, and A completely obscures C from the viewpoint of interest); thus, by consulting the VRCC-3D+ CT, we know that in the given configuration A and C will not crash.

VRCC-3D	DC_COBS(B,C)
DC_COBS(A,B)	**DC_COBS(A,C), EC_COBS(A,C), NTPPC_COBS(A,C), PO_COBS(A,C), TPPC_COBS(A,C)**
VRCC-3D+	DC_ cObs(B,C)
DC_ cObs(A,B)	**DC_cObs(A,C)**

TABLE VIII. Comparison of a Composition Table Entry for VRCC-3D and VRCC-3D+

IV. FUTURE WORK

In addition to the aforementioned investigation of additional methods for constructing the CNG, future work will include testing an implementation of the VRCC-3D+ reasoner on a variety of datasets from different domains (e.g., anatomy, mechanical design, etc.) to analyze its usefulness, accuracy, and scalability. We then plan to further enhance the model, considering additional dimensions of information such as time and physical properties (e.g., transparency, translucency, and repulsion). It is anticipated that such enhancements will require the definitions of additional predicates (similar to the *InFront* predicate discussed herein), and, more challengingly, the identification of valid combinations of all the various attributes.

V. CONCLUSIONS

In this paper we presented an extension of our previous work to combine quantitative and qualitative reasoning in 3D. The enhanced model, VRCC-3D+, is more expressive in that it now distinguishes several types of obscuration. Further, it differs from other region connection calculi in that it determines the composite relation between two objects in 3D, not just 2D. Given the greater wealth of information that can be obtained when additional dimensions are considered, and the enormous amount of 3D image data that are available, we are optimistic that this work will contribute to the implementation (and adoption) of software that more closely replicates human cognitive spatial reasoning.

VI. REFERENCES

[1] J. Albath, J. Leopold, C. Sabharwal, and A. Maglia, RCC-3D: Qualitative Spatial Reasoning in 3D, *Proceedings of the 23rd International Conference on Computer Applications in Industry and Engineering (CAINE 2010),* Las Vegas, NV, Nov. 8-10, pp. 74-79, 2010.

[2] D.A. Randell, Z. Cui, and A.G. Cohn, A Spatial Logic Based on Regions and Connection., *KR*, 92, pp. 165–176, 1992.

[3] 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.

[4] A.G. Cohn, B. Bennet, J. Dooday, and N.M. Gotts, Qualitative Spatial Representation and Reasoning with the Region Connection Calculus, *GeoInformatica* 1(1), pp 1-44, 1997.

[5] J.O. Yang, Q. Fu, and D. Liu, A Model for Representing Topological Relations Between Simple Concave Regions. *ICCS 2007, Part I, LNCS 4487*,Y. Shi et al., Editors, pp. 160-167, 2007.

[6] A.P. Galton, Lines of Sight, in *AISB Workshop on Spatial and Spatio-Temporal Reasoning*, 1994.

[7] 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.

[8] 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 *16th International Conference on Distributed Multimedia Systems*), Oak Brook, IL, Oct. 14-16, 2010, pp. 277-282.

[9] M. Egenhofer and K. Al-Taha, Reasoning about Gradual Changes of Topological Relationships, *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, A. Frank, L. Campari, and U. Formentini, Editors, Lecture Notes on Computer Science 639, pp. 196-219, 1992.

[10] M. Santos and A. Moreira, Conceptual Neighborhood Graphs for Topological Spatial Relations, *Proceedings of the 2009 World Congress on Engineering*.,Vol. I, London, UK, July 1-3, 2009.

Late Arriving Paper

Course-centric vs subject-centric vs community-centric approaches to ICT-enabled learning settings

Luigi Colazzo, Andrea Molinari
Department of Computer and Management Sciences, University of Trento
Via Inama, 5 – 38100 Trento (Italy)
Tel. +39 0461 88 {2144, 2344}
emails: {luigi.colazzo, andrea.molinari}@unitn.it

Nicola Villa
Laboratory of Maieutics, University of Trento
Via Inama, 5 – 38100 Trento (Italy)
Tel. +39 0461 882339
Fax. +39 0461 882327
email: nicola.villa@unitn.it

The paper presents a discussion on the different approaches that we have found in learning settings respect to ICT platforms that support the educational activities. Three are the main approaches: LMSs, where the course is put at the center of the system, social networks (like Facebook™, Twitter™, Flickr™ etc) where the subject with her network of social relationships is put in the center of the system, and virtual communities systems, where the core paradigm of the platform is the (virtual) community that offers specialized services for the purpose of the community to the enrolled members, and where the subject is just a participant that adheres to the rules of the community, with duties, rights, tasks to do and objecftives to achieve. We will discuss all these three approaches, the different levels of applicability in learning settings, and specifically the potential of the virtual communities-based approach that we adopted in the experimentations conducted in the last ten years.

Keywords: e-learning, virtual communities, social networks

I. Introduction

In this paper, we will present our experience as designers, developers and administrators of a virtual communities management system, called "Online Communities" that we are using in learning settings since 2002. Initially, the platform has been substantially used as a replacement of an e-Learning system (LMS) used by the Faculty of Economics of the University of Trento, as at the time we have no low-cost, widely used LMSs like today. We recently managed the evolution of the system towards the provision of what we called a "Personal Learning Space" to the user. This approach is, at first sight, quite similar to the web 2.0 and social networks interaction spaces. In this way, we passed through the three different possibilities that are used today by educational institutions to support educational settings: Learning Management Systems (LMS), Social networks and Virtual communities systems.

LMSs are software platforms where the "course" is put at the center of the system, and the user finds herself "included" in a box (the course) where educational activities are available depending on the decision of the teacher. These well-known software platforms (like Moodle™ or Blackboard™) put the "course" at the center of the system, adding services and tools for expanding the capabilities of the system to support learning and (sometimes) collaboration activities. Accordingly to this, these systems substantially have a "course-centric" approach to service provision, as their original main purpose, where the whole architecture of the system is grounded, is to support educational tasks,

Social networks (life Facebook™, Twitter™, Flickr™ etc) are web sites where the network of social relationships of the users is put in the center of the system (or better, at the center of the interest of the owners of the platform). The incredible success of these social network platforms, has convinced some educational institutions, teachers and tutors to use them for educational purposes. On the research and experimentation side, researchers of learning and teaching techniques based on new technologies have been naturally attracted by the many possibilities offerted by the evolution of the ICTs towards social settings. The main reason for this is probably the widespread convinction that the learning process is a social phenomenon. Normally, some shades of this idea exist, from strong positions like "learning is a social process" to lighter positions like "learning is also a social process". In general, there is an agreement about the social component in the learning processes, and therefore it is almost automatic to assist to the great debate around the so-call e-learning 2.0, as an evolution of the old-style e-learning 1.0 (first generation of e-learning) towards a greater social and participatory sense.

Here we would like to distinguish the power of social networks in the literal sense, as a social phenomenon in e-learning, from the implementation of this sociological concept that Web 2.0 has done. On one side, we have a network of people collaborating to a common objective, managing relationships and exchanging their experiences, and this is a long studied concept in sociology that in recent years has been updated thanks to ICT-mediated networks. Another perspective is the usage of these software platforms (with concepts like "friendship", "friends", "groups" as building blocks of the interaction) in educational settings. To us, it seems that some educational institution is using Facebook™ and similar platforms more due to their success among young persons, rather than a precise and well-designed usage for educational purposes. Of course, we are not denying the enormous potential of social interaction in learning settings: the Web 2.0 and e-learning 2.0 are building blocks of many educational activities, but we would like to distinguish the concept of social network, collaboration, and all web 2.0 services (blogs, wiki, user-generated contents etc.) from the implementation that has been done in social network platforms, and from the usage that (due to this implementation) the educational institutions are doing with them.

The third approach is the one that we conceptually adopted in 2002, in order to get rid of the rigid metaphor of "classroom" or "course" typically implemented in LMS, and to "anticipate" the social network phenomenon just for the small part of the usage of collaboration dynamics in learning settings. What attracted us was the collaboration metaphor in the idea of "virtual community", as originally presented by Rheingold in 1993[1]. So we decided to build from scratch a software platform that implemented "by design" (and not "by derivation" or solely due to hype) the concept of a "virtual community", as a set of people aggregated around a common interest that will use ICT-based tools to interact. Virtual communities systems are therefore those systems where the core paradigm of the platform is the (virtual) community that propose services to the enrolled members, specialized for the purpose of the community, and where the subject is just a participant subject to the rules of the community, with duties, rights, tasks to do and objecftives to achieve.

In this paper we will introduce the problems related to the integration of these different logics (learning spaces social networks and virtual communities) into a single software environment, and how to connect these different worlds into one single architecture. We will describe the developed solution, together with some implications of these choices onto the system, named On Line Communities. The system provides the idea of virtual communities as pillars of the interaction mechanisms provided by the platform to the users. This revealed, in our opinion, on one side the limitations of tradional LMS, and on the other side, the inappropriateness of social networks as pure platforms for supporting learning activities. On Line Communities was designed to allow a continuous exchange of experiences among its users; in that sense, a change of the core of the system towards a "social network" approach is, in our opinion, by far easier than doing the same changes in LMSs that adopt different metaphors (the "course" or "classroom" instead of the "community"). By having the metaphor of "community" as the pillar of our system, we are able to:

- easily integrate social network tools (blogs, wikis, etc) as they are based on the concept of "community" and not on a specialization of it, i.e., "classroom".
- provide additional services that improve the concept of "social network" and "community".

The Learning management systems (LMSs) are normally used by educational institutions to manage the training activities. These systems use the network to create different learning environments related to the learner needs (distance learning, blended training, back-end activities management related to training processes, etc.). As any other type of management system, these applications are also connected to the management model that is represented in the software. In the case of e-learning applications, the represented model is the way by which the institution conceives its learning / teaching processes. The simplest model is the one used in distance learning, when the system becomes a container of learning objects, designed to be effective in the students self-learning processes, the remote control of the current level of learning, the certification of the results achieved and the management of the organizational / financial relations with the training institution.

Much more complex are the systems oriented towards a blended approach. In this case, the LMS offers a virtual space corresponding to what is carried out in the real didactical institution. In this way, the student learns not only in the traditional courses (in the classroom) but also using the virtual space as a reinforcement to face-to-face lectures. This model is the most widely used by the academic institutions. More complex are the systems that tend to support innovative forms of learning such as learning by project, learning by problem and cooperative learning. In this case, the LMS must provide not just a virtual space associated with a course but also special virtual spaces able to work with other similar environments.

The LMS evolution from simply content containers to real cooperation spaces is now in a new phase of transition. The spread of Web 2.0 applications provides the possibility for these systems to evolve and support all those forms of learning excluded from the classical formal and institutional learning methodologies. It is trivial to note that the interrelationship between formal and informal learning includes new challenges to the educational institutions and also to the change of the LMSs' architectures. In this paper we present the solution that we have implemented, i.e., a technological integration between the e-learning processes and the current aspects of social networking into a robust, virtual community-centric platform.

The paper is organised as follows: in the first part we will describe the LMS developed for the University of Trento, named On Line Communities, and its evolution from a e-learning system based on the metaphor of course to a more complex virtual environment based on the metaphor of virtual community. In the second part we will introduce the current study of our research group directed to the integration of the social networking aspect into our environment. In particular we want to underline the risk of a direct adoption of a social network logics into an academic environment, and what could be the correct strategy for the integration of the two types of approaches into a "bridge" platform.

II. On-line communities

In the academic year 1999/2000 the Faculty of Economics of the University of Trento decided to have a software system able to enrich its traditional teaching as an extension on the Web. The first aim was to settle the increasing number of teachers' personal web pages into a single platform. To pursue this result it was necessary to have a Learning Management System (LMS), capable of supplying a virtual environment able to support the educational courses of the Faculty. The resulting system started to function from the second half of 1999 and during this period, the system counted approximately 1,200,000 accesses. Being a quite traditional LMS, in 2002 some observation convinced us to redesign the software. Conceptually, the "pillar" of the platform, i.e., the core component of the system for the idea of enrolling people to a "course". The original system was therefore based, like many LMS today, on the concept of "e-course". In this way, modeling teaching / learning (such as learning by problems, learning by projects, cooperative learning and their combinations) could hardly be connected to the *e-Course*, especially when the software directly represented the metaphor of traditional courses. Substantially, every course offered by the institution was associated to a virtual space available via web. In this virtual space some services were available to participants, in order to extend in the virtual the typical interaction student-teacher of the real classroom. The experimentation, as said, gave us a lot of indications regarding the limiting factors of this approach.

First of all, limiting the collaboration to two subjects (teacher and students) without involving the rest of the information system of the organization was a serious drawback. The needs for cooperation within the academic environments is extending to all the activities that constitute the context in which didactic takes place, not just to the specific "lecture" or "course". In an academic context, not everything concerns teaching: for example, the entire faculty is more than a container of degree courses and a degree course is more than a container of lessons. Another item of afterthought derived from a consideration regarding the didactics of an university, that are not built only as a set of studies and tests, but these activities are inevitably intertwined with the university's organization and its information system. Moreover, the organizational didactic scenario changed with new regulations made by academic institutions, and these changes inevitably reflected on the LMS functionalities. It is important to note that these types of changes are usually the result of a debate process in which both elements of cooperation and negotiation interact.

To answer these (and other) needs, another founding paradigm was needed, with at least four basic new elements:

1. a new definition of collaboration virtual space, generalized respect to what traditional LMS were offering for educational settings;
2. a new definition of the users' role in a community, based on the concept of duties and rights inside this virtual space
3. the virtual space thus created should be generalized, suitable to support not just strict educational activities (like a "lecture" or a "course"), but more extended and complex processes like cooperation and collaboration;
4. the capability of modeling and preserving organizational structure and roles of the educational institution, for example in hierarchical structures like university-faculty-degree-course hierarchy, or any other organizational, network-based structure (like social communities)

This new way of conceiving the collaboration platform was found in the concept of virtual community. The system that arose, called *On Line Communities* [2], was born in 2003 and was deployed for all users in February 2005. Nowadays, among the differents sites and public institutions where it is used, it counts approximately 20.000 users and a rate of 40.000 unique accesses per month.

The collaborative approach [3][4] is a very strong incentive for us to develop On Line Communities; what led us to rebuild the system to supersede LMSs limitations is to allow the exchange of users' experiences within a virtual environment, and within well-defined areas known as "communities". This approach is very different, for example, from the traditional ones available in other LMSs. Our work started before the boom of web 2.0 [5], that has now invaded and changed the way people think and build services on the net. From a technological point of view, trying to re-interpret the structure of our platform under the light of social networks, we could say that "Online Communities" can be seen as a social network specialized in the domain of learning activities. Nevertheless, due to the profound conceptual, design and technological starting points and approaches, some relevant differences exist between our platform and social networks on one side, and between our platform and LMSs on the other side.

The core of the application is composed by some abstract entities, i.e., VCs as aggregation of people to which some communication services are available in order to obtain certain objectives. With this approach, it could be possible to represent all the hierarchical relationships between different types of educational communities (such as Faculties, Didactic Paths, Master Degrees, Courses, etc.), as any other relationship among communities inside organizations.

In a nutshell, the main characteristics of a community could be summed up as follows:

- Each Community offers many services to registered users that have different roles/permissions inside the community
- The services are general applications that enable users to communicate both in synchronous and asynchronous way, to publish contents, to exchange files, to coordinate events, etc.
- Services offered by a community are activated by a manager of the community according to the needs, and the users of a community can use them with different rights and duties.
- Rights/duties in the community are different from rights/duties for the services
- Communities can be aggregated into larger communities with hierarchic mechanisms and infinite nesting levels. Communities can also be aggregated in an arbitrary way into larger communities disregarding the possible position of a hierarchical structure, in a sort of "transversal" link that overcomes the concept of "hierarchy" and follows the idea of "mesh". Thanks to these features, a complex but powerful mechanism of propagation of services/roles/permissions/ rights/duties can be set among communities of the same branch or of different branches.
- All users are recognized by the system and by the community: people external to the system can see public part of the community (services, material, contents etc.) only if managers allow this (ex. a blog of one community could be opened to external contributions)
- Services can take advantage of the "mesh" structure of Online Communities to provide some interesting though non-existing features, like "transversal wikis", or "merged blogs". One blog, in fact, can be the "fusion" of all the blogs of children communities, or a wiki can take the definition transversally from all wikis in related communities.
- Last but not least, a VC is the container for collaboration processes not limited to educational activities, but for any collaboration activity needed in an organization. Research teams, recreation groups, friends, meetings, conferences, secretariats, board of directors, colleagues, next social

dinner, anything could be an aggregation of people around a scope that can take advantage of the virtual spaces offered by the Virtual community.

III. SOCIAL NETWORK AND LEARNING SETTINGS

The collaborative approach described above [6][7] represented a very strong incentive for us on the development of new versions of "OnLine Communities". The philosophy that led us to rebuild the system was to allows the exchange of users' experiences into a virtual environment, and within well-defined areas known as "communities". This "community-centric" approach is very different, as said, from the traditional "course-centric" one, typical of other e-learning management systems. The boom of the web 2.0 that changed the way people think and build services on the net. The "social" approach adopted by Facebook™, LinkedIn™, YouTube™ or Twitter™ is very interesting, and according to statistics, it involves an increasing number of web users[8]. Such services facilitate the bi-directional communication and the exchange of experiences between users, but on the other side can hardly be applied *"tout court"* to e-learning experiences. For example, students or teachers are "forced" to use blogs inside the "classroom" most of the time due to the lack of more appropriated tools. Many teachers used blog platforms because they did not have any other place to create a "diary" of the lectures, with teaching material associated to each lecture. However, in principle, a blog is a personal diary, that cointains ordered-by-date comments on personal facts. The "experiential" component is relevant, respect to the usage of it as a repository of PDF files. Also students, that probably know better than us the role and scope for which blogs should be used, most of the time are not so comfortable in "participating" in a blog when used in educational settings, if not for curiosity, dictation of the teacher, trend or simply in order to have access to course information and materials. Of course, we are not denying the usage of blog as a very interesting, participative experiences (we built a blogging tools in our platform). We are simply saying that due to the eclecticism of web 2.0 tools, when the appropriate tools are lacking, people involved in educational tasks use what they find to be accessible, quick and easy to use. Thus, a blog (or a social network personal home page, or group) can be used as the classroom registry where annotating the lecture's topics. If we finally add to blog's posts the capability of associating uploaded files, we can have a very simple but efficient LMS with just one service. Quite strange, but we are not surprised to see teachers that use their one private blog or facebook page for coordinating and distributing educational material. So, forcing web 2.0 services to become e-learning services in an e-learning platform is a hazardous operation: the result could be a loss of quality in the learning process (like any usage or unappropriate tools for certain context), confusion, workarounds and possibly users' dissatisfaction. Of course, if you consider e-learning just as distributing PDF files to students, most of these tools are instead perfect, but this is not our perspective.

Following these considerations, another component of the Web 2.0 phenomenon has attracted people operating in educational settings, i.e., using social networks as teaching / learning platforms. Our consideration is more or less identical of the above consideration regarding the usage of blogs, but due to different perspectives. According to some recent statistics [8], the majority of users who use the so called "social networks services" are concentrating on the well known *"peoplesurfing"*: navigate into the friends' profiles, see pictures, personal information, etc. Voyeurism and curiosity behind the success of these applications? This study seems to suggest this scenario, but what interests us is that they are building a private community of people around their own interest: exactly the definition of "virtual community", and not the definition of "classroom".

We are aware of the clear phenomenon that is emerging from friends' social network [9]; it is true that the action of adding a person to the friends' list requires an approval, but it is also true that a user can see at any moment the people connected to his/her friends, and there is no approval mechanism on others' friends. This opportunity on one hand could be positive, but on the other can be critical within a learning context (within a university, but also within business contexts). Once again, our virtual community approach solves (in our opinion) this problem with this "self reputation" capability of moderated virtual communities. The person that enters a virtual community in our system is authorized by the community administrator, and from that moment the person is in contact automatically with the people inside the community. This is the pillar of the virtual community: I'm in the community because I share the objectives of the community. So I don't have to declare, accept, or manage my contacts inside that community, and I'll be never connected to a friend of a friend of a friend. Of course, Online Communities allows the users to manage friends' lists, but this is different from managing community members. The differences between "friends" and "community members" is very clear, and the platform with its security mechanisms based on roles / permissions / rights / services allows the user to manage these two different concepts.

Given that the increase of the social interactions is not a negative aspect, the risks coming from the direct use of Facebook's approach into an environment with different aims (something like "I'm a friend of a friend who was the friend of my friend….) are complex to be evaluated. A critical consequence is to become implicitly a friend of my contacts' friends, thus starting a sort of recursion in the friends' list of friends.

We have therefore considered the best part of social networks (the services) but starting from re-thinking our system with a consolidated view on the virtual communities with some "social" extensions, changing our community system to a sort of "community 2.0" system: we like to define it as a "*Private community Environment*" (PCE).

As seen above, our approach is not the pure "classroom-based" approach of LMSs like Moodle™, nor the generalist approach of social networks like Facebook™, halfway between the traditional learning environments and the social network applications. A PCE does not accept the simple subscription; it requires you to enter into a community in order to interact with others. A PCE is not usable just for e-learning related activities, it allows you to create a community (for example):

- to manage your next scientific conference with your research team and any other person involved in that conference;
- to collect opinions around the next meeting of the Board of Director, where the meeting IS the virtual community, the virtual space inside which the participants to the meeting will have, with different roles ranging from the secretariat to the President, the various services of the community at their disposal;
- to aggregate in a virtual space those people interested in supporting an organization of volunteers for healthcare assistance, to discuss with doctors the problem of applying pain therapy to patients.

By the way, such a virtual space could respond simultaneously to all of these instances, but also to "simpler" issues related to educational scopes; the didactical needs and the improving of the interaction level necessary for the development of a participative environment. Online communities is therefore an alternative to both systems, better positioned (in our opinion and experience) to solve collaboration problems of people that want to be supported by ICTs.

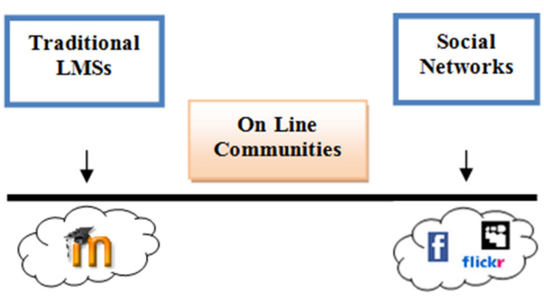


Figure 1: On Line Communites and the other systems

The community metaphor helps us to follow this type of approach; in fact, it is very similar to the metaphors used by the most popular social networking services, but it is also different from what is used in the traditional LMSs. Generally, the increase of the social interactions in learning is never a bad thing, so using a tool that makes social interaction so easy is surely a good experimentation. The integration between a learning space and a social space could also be positive because it reinforces the idea that learning doesn't just happen in privileged spaces, but learning can happen anywhere and without the intervention of an "official teacher". This allows us to integrate the learning activities in a broader context that students are interested in, because the border between "learning" and "life" isn't so distant.

IV. LEARNING AND SOCIAL NETWORKS: TWO STRATEGIES INTO A POSSIBLE ARCHITECTURE

If we look at the whole range of application fields where we are using On Line Communities, the platform clearly evidences its nature of a collaborative environment that wants to stimulate the participation and put to value users' cooperative work. Today, with the advent of new communication and collaboration paradigms, On line Communities has become an example of a computer support cooperative work system (CSCW) devoted to teaching/learning. In recent years, we extended our system to functionalities and services typical of Web 2.0. However, some relevant differences exist between the approaches used by web 2.0 applications and the ones used in On Line Communities. To overcome these differences, a changing of the rules used in the virtual space is required, and these changes have a direct influence on the entire architecture of the system.

The cooperative virtual space of On Line Communities is actually a closed environment. The users participate to the activities inside the system directly with their real identity. In fact, a person who enters a virtual community of our system is authorized firstly by the platform administrator (for certifying user's credentials), and after by each community administrator for each community the user wants to enrol with. Once the user is accepted inside the community, from that moment he/she is automatically in contact with all the people inside the community. This is the pillar of the virtual community: as said, I'm in the community because I share its scope, and all the people of the community have more or less the same interests / objectives / tasks. Following this logic, the user is not obliged to declare, accept, or manage his/her contacts inside that community: s/he will never have to face the "domino" effect of friendship typical of most social networks.

The circumstances that we consider favourable in our system (lack of anonymity and control of the external accesses) have origin in two explicit requirements of our Faculty of Economics. The exclusion of anonymity is the result of a belief, that normally indicates that the anonymity into virtual learning environment should be banned, so that the actors cannot shirk from their responsibilities. The second circumstance (access control) stems from the will of a substantial number of teachers to block the publication on the network of their own courses' Learning Objects. These choices made the system impermeable to the users' social dynamics, or to the communities existing in the social networks.

To overcome these limits without affecting our constraints requires a radical change of the system architecture, that sees the person as a member of one or many communities. On the other hand, in the web 2.0 applications, the participants exist as individuals who, for example, can create themselves a specific community. The rethinking of the system with these ideas, could change our community system to a sort of "community 2.0" system: we like to define it as a "*Private community Environment*" (PCE). The difference between the two approaches is that the communities in our systems are created as an extension in the virtual space of real didactics. On the contrary, in web 2.0 social networks, virtual communities emerge from the interaction among users' own networks.

Following this line, we studied how to modify the architecture of our system, as we wanted to implement the good part (from our perspective) of the incredible revolution introduced by social networks. We wanted to transform our virtual communities platform into a sort of bridge system between the classical methodology followed into the most famous LMSs (like for example Moodle [10]) and the new web 2.0 and social networks applications (like for example Facebook, MySpace and Flickr), without losing the focus on the learning processes.

Figure 2. On Line Communities as a bridge between two different approaches

The architecture that we developed has two fundamental goals: a) to make our system more permeable to all experiences that take place inside the web, including applications for social networking and Web 2.0; b) keep control, up to a certain level, of the actions taken by users of our system. In fact, our context is connected to learning environments / academic settings, and not directly to leisure time.

Following these approaches, many drastic changes have been introduced into the platform, moving the focus from "community" to "user". As an example, when the user connects to the system, the user' personal home page and its services are presented, trying to create a real Personal Learning Space (PLS). We are imagining the new users' Personal Learning Space as an aggregation of two distinct environments. The user will be free to decide what part of his/her relations and contents to import (into On Line Communities) or export (to social networks applications). From a technological and management point of view, this approach presents more problems than solutions. This solution also required a strong review of many parts of On Line Communities, and in particular the management of users' roles and permissions.

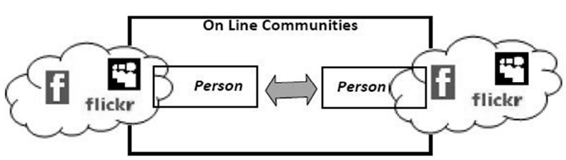


Figure 3. The representation of the new architecture

This approach has different values, in particular regarding the teaching strategies; in fact exporting the contents outside an e-learning platform could accentuate the social role of the educational institution as a source of knowledge and of better didactic practices.

On one side, this solution give to the user more freedom than into a classic LMS, but on the other side, it is more difficult for didactic institutions to be implemented. In fact, while the institutions are becoming a knowledge centre through the participation of its members, at the same time they are being exposed to the risk of the complexity and the personal relationships of its members.

According to this solution, it is important also to develop a new interface of the Personal Learning Space of each user; the metaphor of community makes possible to implement some interesting features, directly connected to the user and his/her list of contacts. In other words, this gives the possibility to the users to manage their own learning spaces: in some way to enable the users to create a "*Personal Learning Space*" for their needs. Each user will have the opportunity to access to his/her personal page, which will contain personalized services. As a result, some interesting new services can be provided, for example:

1. access to communities where the user was registered;
2. view the most used services by each user;
3. access to contextual services for each community;
4. access to the personalized services;
5. add some services into the personal learning area.

The user can access to the list of communities where s/he is enrolled in, because this is the primary scope of connecting to this system. But together with this, the user finds a set of services that are typically connected to his/her own person, a sort of personal space within the system. The services are "general", so in this condition the user will see services that are at "personal" level. This can be repeated and nested when the user enters inside a community: he will find (more or less) the same services, but this time these will be the services of *that* community, with different permissions, roles, list of contacts etc. A typical example is the Blog service: when I'm inside my PLS, the Blog is *my blog*, when I'm inside the community "workgroup XWZ", the service *Blog* refers to the blog of *that community*: same service, totally different context and contents, totally different the role of the user could be.Finally, thanks to the inheritance mechanism among communities provided by the platform, the blog of that community can be merged with the blogs of parent community/ies, or with the child communities, or with sister communities (children of the same parent community).

V. EASE OF USE

In this paper we presented a summary of our research activity in learning settings, where we faced with three different approaches to provide support to learning and teaching activities. The first possible approach is the one based on a "course-centric" vision of the learning processes, where every activity done by users is located around the "course". This is what we saw implemented as a main metaphor in Learning management systems. A second possible approach is a derivation of tools and services made available by social network platform, where we read a subject-centric vision of the learning processes. This approach is found inside those learning experiences where social networks are used as virtual spaces for implementing the interaction between teachers and students. A third approach is the one used in our implementation, called “Online Communities”, i.e., a community-centric approach. The system provides a virtual space where collaboration, communication and learning-specific tools and services are available for the participant to the community. The community is the center, the subject enrolls in the community depending on her interests and needs. The community therefore could be a course in educational settings, but could also be somethings not necessarely or specifically related with educational activities, or can be equipped with services that implement different collaboration mechanisms respect to traditional learning settings (download material, read a forum, do a questionnaire etc.)

We experimented the “long and winding road” of architectural choices, needed for taking full advantages from all these new approaches and suggestions coming from our experience on the field and from what the we 2.0 was introducing. Moreover, by coupling this web 2.0 tools with a virtual communities approach, rather than the traditional “course” metaphor, we obtained many advantages in the possible services provided to end users. In particular, it was evident that social applications are profoundly different from what is provided by traditional e-learning applications. Our system, originally followed a logic of blended learning, was also focused on the metaphor of the course. The evolution to a different metaphor (the community), has opened new perspectives, different from anything that can be seen as formal learning. Our focus on integrating web 2.0 and social network services, increasingly common in the worldwide web, seemed to be quite naturally.

VI. References

[1] Rheingold H., (1993) *The Virtual Community*, 1993 (itlian translation Comunità virtuali, Sperling & Kupfer, Milano, 1994).

[2] Colazzo L., Conte F., Molinari A., Villa N. (2007), *Real communities vs. Virtual communities: structural adaptation of a Learning Management System*. In: *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2007*, Chesapeake (VA): AACE, 2007. p. 2416-2423. "E-Learn 2007 Conference", Quebec City, October 2007

[3] Anderson C. (2006), *The long Tail: how endless choice is creating unlimited demand*. Random House Business Books, N.Y.

[4] Chambers S.A. (2005), *Working on the democratic imagination and the limits of deliberative democracy*, Pol Res Q 58(4):619-623

[5] O'Reilly, T. (2005). *Web 2.0: Compact Definition?* Retrieved from http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html

[6] Chambers S.A. (2005), *Working on the democratic imagination and the limits of deliberative democracy*, Pol Res Q 58(4):619-623

[7] Anderson C. (2006), The long Tail: how endless choice is creating unlimited demand. Random House Business Books, N.Y.

[8] Compete.com (2007), Facebook Activity Breakdown, August 2007. http://blog.compete.com/2007/09/14/facebook-activity-breakdown-application/. October 2008

[9] Boyd D. (2006) *Friends, friendsters and mySpace Top 8: writing community into being on social network sites*. First Monday 11(2), December. http://www.firstmonday.org/issues/issue11_12/boyd. Accessed October 2008

[10] Moodle. http://www.moodle.org, 2009.

Proceedings
Multimedia Arts
&
DMS/VLC Demo

Editor

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

Il Giocoliere

The Juggler

Max Coppeta[1], Marianna Carbone[2], Francesco Colace[3], Massimo De Santo[3], Gennaro Vallifuoco[1]

[1]MaxCoppeta Design Corporation
[2]Accademia di Belle Arti di Napoli
[3]DIEII - Department of Electronic and Information Engineering, University of Salerno, Italy
info@maxcoppeta.it, carmardue@hotmail.it, {fcolace, desanto}@unisa.it, info@gennarovallifuoco.com

Abstract

The Juggler is a performance that aims to investigate the relationship between the theatre and the ICT. The aim is to realize a real and at the same time imaginary place where actors can interact with imaginary characters, objects or situation. The plot of the performance is simple: the Juggler will meet two characters: Trick and the Shaman. These two characters will guide the Juggler inside a fantastic world where he will able to connect the terrestrial world and the afterlife searching for the memory, values, essence. Therefore, the Juggler will express the need to act with the imagination to tell a world of memories relived through signs, symbols and atmospheres suspended for walking away from a reality too much unnecessary concrete. The core of this experience is the relationship between video and body actors. They talk, returning one dimension to another. The video, in this experience, it is not intended as a form of sealed and assembled in support of a language (both theatrical and performative), but is conceived and developed as a flexible material at the disposal of actors / performers / dancers in a state of inter-connection . The actions of one determine the reaction of the other. In order to achieve this goal, a multimedia product was developed in Flash. This media allows the actors on stage to interact in unreal situations or create new and non-repeatable. In this way the director can be an active part of the show, changing the course of history in real time. The use of Flash, in fact, allows the director to choose which part of the storyboard to use and the situation going to build. In the same way the public can interact with the actors and scenes. It 'important to note that the media has the duty to act as a backdrop to the show and the costumes have been designed taking into account the need to interact with a virtual performance space. As previously mentioned the core of the show is the multimedia content developed in Flash. The artistic purpose, that we wanted to achieve with this product, was to build a show in which traditional methods and approaches of the theatre have merged with those of ICT. In this way an innovative way of understanding the visual representation has been built. So technology is not a simple tool to support the artistic performance but it is a protagonist of the action and able to modify and interpret the course of history. The multimedia content has been built making use of original texts and music. The show has been featured in various Italian opera houses, creating a performance much appreciated by audiences and critics. A version for the web of the product has been created and can be find at the following link: http://www.maxcoppeta.it/giocoliere/

Voci che nessuno ascolta

Voices that no one hears

Max Coppeta[1], Marianna Carbone[2], Francesco Colace[3], Massimo De Santo[3], Gennaro Vallifuoco[1]

[1]MaxCoppeta Design Corporation
[2]Accademia di Belle Arti di Napoli
[3]DIEII - Department of Electronic and Information Engineering, University of Salerno, Italy
info@maxcoppeta.it, carmardue@hotmail.it, {fcolace, desanto}@unisa.it, info@gennarovallifuoco.com

Abstract

Voices that no one hears is a short film realized by the use of the Stop Motion techniques. Stop motion is an animation technique to make a physically manipulated object appear to move on its own. The object is moved in small increments between individually photographed frames, creating the illusion of movement when the series of frames is played as a continuous sequence. The issue addressed in this work is the war as a universal struggle of men against men, focusing on women and children as symbols of innocence and victims of violence. In this project, the theme is bloody but lived in a dream, symbolic representation where natural events and animals (the storm, the seagull, the butterfly) become characters in a fantasy world in an ongoing exchange between dream and reality, absurd and paradoxical. This aim has been accomplished by the use of ICT. From a technical point of view the challenge was to realize a professional work by the use of "normal" technology. So with normal desktop, a digital camera, some software as 3D studio max, Adobe Premiere and fantasy a professional movie has been realized. A professional sculptor has constructed the scene and were made costumes and choreography. The actors are some Ikea puppets. By the use of an apple G5 server and a Canon EOS 5 Camera the movie has been realized in a week. The multimedia content has been built making use of original texts and music. The movie has been featured in various Italian opera houses, creating a performance much appreciated by audiences and critics. A version for the web of the product has been created and can be find at the following link:
http://www.premioceleste.it/opera/ido:74921/

Color-Based Recognition of Gesture-traced 2D Symbols

Danilo Avola, Paolo Bottoni, Alexandru Dafinei, Anna Labella
Department of Computer Science, Sapienza University of Rome
Via Salaria 113, 00198 Rome, Italy
{avola, bottoni, labella}@di.uniroma1.it, a_dafinei@yahoo.it

Abstract

Gesture-based interfaces provide an intuitive and natural way to convey 2D graphical command symbols for interacting with applications and services.

Users can prefer performing gestures "in the air" or through suitable tools, e.g. tablets, depending on the context of usage and the available tools. A collection of device-independent algorithms and tools for analysing gestures would facilitate such a plasticity in the interaction process.

This paper describes how a set of 2D graphical symbols, traced by moving a led pen in front of a single-camera, can be recognised in a gesture recognition system, using the same algorithms employed for recognising these symbols when used on a tablet.

1 Introduction

The main goal in the design of multimodal interfaces is to increase the naturality of human-machine communication process, exploiting typical communicational behaviours of human beings. Gesture- and speech-based modalities are considered among the most promising ways for human-machine communication, as they are most common in human-human communication. For the case of visual concepts (e.g. graphical symbols), these are better expressed through gesture modality, when it is possible to achieve an immediate abstract visualization of temporal, geometrical and spatial characteristics.

Users can perform gestures finalized to command a device using a specific tool (e.g. movement sensors, tablet PC) or simply by tracing signs with their hands captured by some optical device. Depending on several factors, e.g. application environments (e.g. kind of application/service, kind of device) or technical contexts (e.g. indoor, outdoor, presence of light), different interaction supports may be preferred. The type of support, as well as the choice to focus on the hand movements, heavily influence the approaches used in the recognition process and the involved algorithms.

A possible classification of gesture recognition algorithms is based on the way by which they extract movement information from low level data. In particular, algorithms can be classified as *tracker based* or *vision based* [3]. To the first class algorithms belong which base the recognition process on haptic devices (e.g. data gloves, body suits) by which to capture the whole gesture informative content, while in the second class we find algorithms that obtain information by analyzing a sequential set of images representing a human-computer interaction scene.

This paper describes a simple gesture recognition process in to the second class, where gesture movements, performed by the user through the management of a led pen (or some other simple gadget coloured with a uniform tonality), is detected by a color tracking algorithm that analyzes a sequential set of images captured by a single-camera. Once the spatial and temporal information of the tracked color has been obtained, a sketch recognition engine is adopted to interpret the related graphical symbol.

The basic idea of the system is that the user can define, through a sketch-based interface (i.e. based on the stroke identification), a set of recognizable gestures. In fact, the core system is based on a sketch recognition algorithm built to face the main aspects in this field *multi-domain definition*, *real-time recognition* and *tracing style interpretation* [1]. For this reason, our approach identifies the trajectory (i.e. spatial and temporal information) of the tracked color as a set of sequential points through which to define a stroke representing an abstraction of the user's gesture.

In the following, Section 2 presents the system architecture and shows some preliminary experimental results and Section 3 concludes the paper.

2 The System Architecture

The main aspects of the system architecture are discussed with reference to Figure 1.

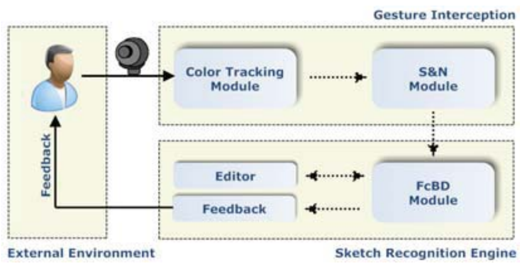


Figure 1. System Architecture

The architecture relies on the definition of a set of sketched graphical symbols to be recognised. An editor is available, which allows skilled user to define unistroke symbols. With each symbol, an agent is associated able to recognise instances of the symbol based on an analysis of mathematical and geometrical features extracted from the strokes drawn with the gestures.

The system is structured into three modules. Given a set of predefined symbols, previously defined with a sketch editor, and characterised in terms of strokes, the modules process a set of video frames captured by the camera to extract the points of the fixed color trajectory and identify a traced stroke, to be matched with the strokes defining the symbols.

The first module (*Color Tracking*) adopts a color tracking technique [2] to detect, within an image, all the regions that have a fixed target color. The algorithm follows, frame by frame, the spatial movement of these regions and produces, for each of them, a bounding box surrounding it. The set of bounding boxes defines the trajectory of the gestures performed by moving the led pen in front to the single-camera.

The second module (*S&N*, *Skeleton and Normalization*) receives the set of bounding boxes and reduces each of them to a single spatial point by calculating their barycenters. It also interpolates these points to deal with the possible lack of spatial contiguity due to the speed of the gesture in comparison with the *fps* (frames per second) capacity of the camera. At the end of the process the module will have produced a representation of the user's gesture as a stroke.

The third module (*FcBD, Feature calculation Bid Decision*) is an agent-based sketch recognition engine. Given a sketch/gesture library created with the editor, the FcBD module extracts features from the strokes produced by the S&N module and matches them against the features characterising the symbols in the library. Figure 2 shows a test library.

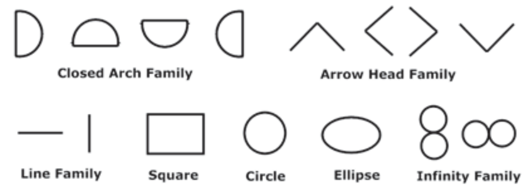


Figure 2. Basic Test Library

In particular, each agent checks the conformance between the evaluated features and the description of the symbol for which it is responsible, and decides whether the stroke can be recognized as a specific graphical symbol. When a conflict occurs, i.e. two or more agents claim recognition for a same stroke, a mediator agent solves them through some heuristic strategy.

The FcBD engine has been tested, with good results using traditional sketching tools (mouse and tablet). Experiments are now starting with the camera-based gesture-recognition. The first results are promising, but they have been obtained in a controlled indoor environment and with regular, not too fast gestures.

3 Conclusion

We have presented an agent-based sketch recognition engine for definition of 2D graphical symbols and their recognition through gesture color tracking. We need to improve the current color tracking algorithm to obtain independency from both the application environments and technical contexts. Moreover, we are working on a freehand gesture recognition system.

References

[1] C. Alvarado and R. Davis. Sketchread: a multi-domain sketch recognition engine. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, UIST '04, pages 23–32. ACM, 2004.

[2] C. R. Brice and C. L. Fennema. Scene analysis using regions. *Artificial Intelligence*, 1(3-4):205 – 226, 1970.

[3] M. Turk. Gesture recognition. *Microsoft Research Academic Press*, 1(10):1–20, 1998.

Music via Motion: A distributed framework for interactive multimedia performance

James Leonard
Ecole Polytechnique de l'Université de Grenoble1
Université Joseph Fourier, Grenoble, France, and
ICSRiM – University of Leeds, Leeds LS2 9JT, U.K.
james.leonard@icsrim.org.uk

Kia Ng
ICSRiM – University of Leeds
School of Computing and School of Music
Leeds LS2 9JT, U.K.
kia@icsrim.org.uk, k.c.ng@leeds.ac.uk

Abstract— Music performance is closely associated with body movements in many levels for both the instrument player and the listener. Recent studies show the existence of a strong link between bodily movement and conception and perception of music. The MvM (Music via Motion) framework was proposed over 10 years ago now and has been designed for several different scenarios including interactive dance performances and interactive installations. This paper presents a recent revision to the MvM framework and core system with a range of enhancements and optimizations. It discusses related background and details the new design together with a case study on conducting, using the MvM framework to offer visual and audio feedback on the beat position in order to study the perception of conducting gesture communication and different interpretations.

Keywords- motion; multimedia; distributed; framework; mapping; gesture; conducting

I. INTRODUCTION

Recent years have seen a multitude of developments and growth in motion-based systems for live interactions including gaming and performance. In musical contexts, such systems offer a more expressive interface for the performer, and the possibility to create music without the physical constrains of an analog instrument.

The original MvM (Music via Motion) framework [1] was designed to operate a trans-domain mapping between physical movement and multimedia events in creative domains with visualization and sonification. The current revision aims to optimize and enhance the MvM framework, offering a generic motion analysis and mapping device.

Starting with a brief description of related background, this paper will discuss the design and architecture of the new framework, new features and functionalities, and a case study on a conducting analysis application will show the results yielded by the system.

II. BACKGROUND

Various systems have been developed to use gestural control for many forms of interactive performance [2, 3], as was the original MvM [1]. Many of such systems allow for very flexible output, enabling control over any set of relevant features. Although certain body movements tend to relate to associated shaping of multimedia output, the effect of gestural control is often specific to one user application, hence the need for generic motion analysis and mapping.

Computer Vision software has evolved over the last decade, with more processing power and powerful new techniques. The rise in popularity of motion controlled video gaming systems (e.g. Wii, Playstation Move, Microsoft Kinect) and mobile devices with sensors (typically accelerometers) has also lead to affordable and robust hardware systems for motion tracking, which are relevant tools for augmented multimedia performance [4, 5].

The MvM revision aims to take advantage of these developments, accepting input data from multiple devices. Features are extracted from the data and mapped into generic information, which can then be translated into specific multimedia content. These optimizations require a new framework design, which handles distributed modules, multithreaded processing of multiple inputs, extendable generic analysis and mapping modules.

III. MUSIC VIA MOTION DESIGN

The new MvM framework revolves around a channel-based architecture. A channel is created for each input. Each channel is configured to process the input data through a number of analysis modules. The results of each of these modules are then sent to a set of configured mapping modules, which send output data.

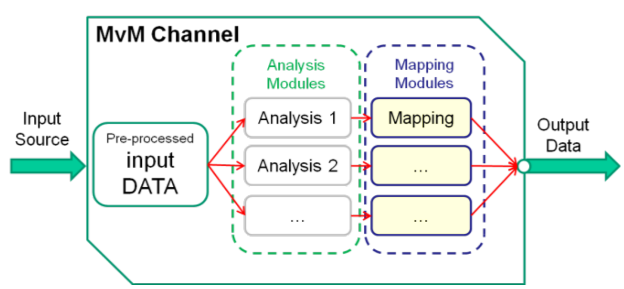


Figure 1. MvM channel design.

This design effectively allows for any number of channels, receiving from various input systems, to be processed by any number of channel-specific analysis and mapping modules, which generate output data relative to the associated channel. Currently integrated input channels include: raw video stream, kinect, mobile devices (via UDP).

IV. New Features and Capabilities

A. *Flexible analysis and processing:*

Depending on the type of input data, pre-processing may be required. For example, a raw video stream input needs pre-processing, while position or accelerometer data from a mobile sensor can be sent directly to the MvM input channel.

The framework offers a range of pre-processing functionalities including the use of OpenCV, with various forms of motion tracking. Currently, skeleton tracking of Kinect input is obtained with OpenNI [6].

B. *Distributed system:*

Input pre-processing and multimedia feedback require considerable CPU power. The new MvM design is a distributed system, which allows for the input data pre-processing, analysis and mapping, and multimedia output to be distributed across multiple systems. Communication is ensured via UDP.

C. *Multithreaded :*

All MvM channels run in parallel threads. Moreover, all analysis and mapping is multithreaded, each input is analyzed and mapped in parallel by a number of module instances, specific to the channel.

D. *Extendable pool of analysis and mapping modules:*

Gesture analysis and mapping modules are designed to accept several input sources, making them interchangeable and reusable for various applications. The existing base can be extended with very little change to the system. In addition to the generic gesture analysis functions implemented under the MvM framework, OpenCV modules can also be used.

V. Conducting Case Study

Analyzing and understanding conducting gesture proves to be an interesting potential application for augmented performances [4] and in depth analysis for learning tools. Several projects in ICSRiM currently aim to analyze and track beats and time signatures from a conductor.

The present case study is based on the use of the Kinect and OpenNI for body tracking. A beat detection analysis module finds beats, through an algorithm based on hand movement orientation, angle change between successive frames and speed. Detected beats are then mapped to trigger an event of visual or audio nature.

The new MvM design allows for sub-modules to analyze data from an existing analysis module. Implementations of this feature in the case study include tempo calculation, using a sliding window of the last beats, and dynamic time warping (DTW) for time signature recognition, by matching the normalized spatial location of a sequence of beats to a known reference.

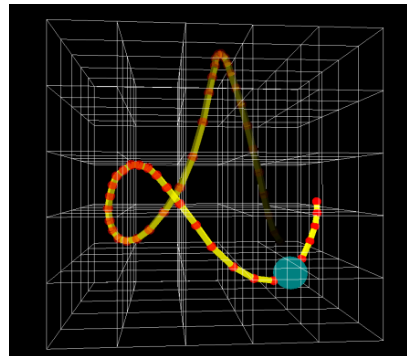


Figure 2. Conducting beat-detection visualization.

VI. Results

The preliminary results obtained for the conducting case study show robust real-time beat tracking, which was mapped into visual output in Fig. 2, with accurate tempo calculation and time signature matching. These results prove the efficiency of the new design. Future directions for conducting analysis could include meaningful visual feedback for augmented performance, and the design of a virtual orchestra, where conducting information is used in Max/MSP to conduct and control a musical piece.

VII. Conclusions

This paper has presented features and enhancements which have been made to the original MvM system. It describes a case study on conducting, built with the new framework, and demonstrates the possibilities of MvM for gesture controlled multimedia feedback for live performance.

This work is ongoing. Future directions include extensions with the new Kinect for Windows SDK and further case studies that involve shape-based analysis and multiple player interactivity modeling.

VIII. References

[1] K. C. Ng, Music via Motion: Transdomain Mapping of Motion and Sound for Interactive Performances, Proceedings of the IEEE, 92(4), April 2004.

[2] A. Camuri, B. Mazzarino, and G. Volpe, Expressive Gestural Control of Sound and Visual Output in Multimodal Interactive Systems, Sound and Music Computing '04, Jim/Cim'04, 2004, Paris, France.

[3] G. Qian, F. Guo, T. Ingalls, L. Olson, J. James, and T. Rikakis, A Gesture-Driven Multimodal Interactive Dance System, IEEE International Conference on Multimedia and Expo (ICME), 2004.

[4] E. Wong, W. Yuen, and C. Choy, Designing Wii Controller - A Powerful Musical Instrument in An Interactive Music Performance System, 6th International Conference on Advances in Mobile Computing and Multimedia (MoMM), 2008.

[5] M. Benatan, I. Symonds and K. Ng, Mobile Motion: multimodal device augmentation for musical applications, in Dunn, Bowen & Ng (eds.) Proceedings of EVA2011: the International Conference of Electronic Visualisation and the Arts, British Computer Society, London WC2E 7HA, UK, 2011.

[6] OpenNI, www.openni.org (last accessed 28/6/2011).

MyStoryPlayer, Collections, Playlists: Content aggregator tools in ECLAP Project

P. Bellini, I. Bruno, D. Cenni, A. Fuzier, P. Nesi, M. Paolucci, M. Serena

Distributed Systems and Internet Technology Lab, Department of Systems and Informatics

University of Florence, Via S.Marta 3, Florence, 50139 Italy

Tel: +39-055-4796523, fax: +39-055-4796523, http://www.disit.dsi.unifi.it

pbellini@dsi.unifi.it, nesi@dsi.unifi.it, paolucci@dsi.unifi.it, serena@dsi.unifi.it

Abstract—**The exploitation of content in the area of cultural heritage is a critical issue. There is a strong push on defining and collecting content metadata from cultural institutions (such as museums, university, archives, library, foundations, etc.) to see them accessible on Europeana for a larger audience, the so called European digital library. This action is going to reduce fragmentation and to integrate valuable collections in a unique place. Europeana foundation and portal that do not collect content items files, but only classification information (i.e., metadata), including the original URLs to the content. These URLs refer to the original content owner and/or to the content aggregator. Europeana could have a strong validity for educational purposes. In the demo, the main tools in terms of aggregations are presented. The proposed aggregation model attempts to satisfy the above mentioned requirements with a semantic model and tools providing support for playlists, collections and media synchronization through MyStoryPlayer annotation tool. Finally, the aggregations are also provided as semantic concepts to Europeana. The results have been produced in the ICT PSP project ECLAP founded by the European Commission, www.eclap.eu .**

I. CONTENT AGGREGATOR TOOLS IN ECLAP

A. Model

ECLAP e-library for performing art portal and service has been created for ECLAP ICT PsP project (European Collected Library of Artistic Performance) of the European Commission

ECLAP is a best practice network for collecting and aggregating content for Europeana.

In ECLAP, several different kinds of concepts and data are modeled ranging from content to users and their relationships.

The different kinds of content, and also the users are associated with a thematic taxonomy which describes the content in terms of genre, performing art type, historical period, subjects, management aspects, dissemination aspects, etc. Also the user profile includes such a classification to allow users to express their preferences about content theme.

Content in ECLAP can be: image, document, video, audio, archive, animations, etc., but also playlists, collections, web pages, annotations, comments, etc. This demo is focused on describing aggregator contents like playlists, collections and annotations according to the semantic model in ECLAP.

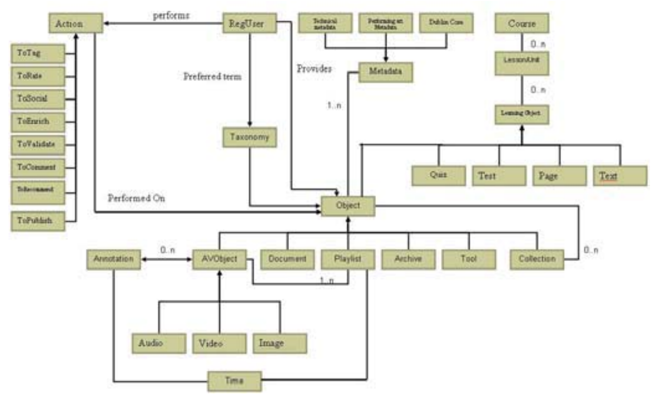


Figure 1. ECLAP semantic model

B. Playlists

According to the model of Figure 1, a playlist can be a set of audiovisual contents. Playlists are a way of playing a series of multimedia contents in a specific order, defined during the authoring phase. The ECLAP play list is not the classical play list of many social network, since in ECLAP the user may chose to put in the play list not an entire video or audio, but a segment of video, audio or a an image for a number of seconds. This can be done through a proper player (fig.2) that will be loaded during the editing phase in which user can select media, add them to playlists box, choose an order and choose a segment for each media. Even images has a time related to their permanence as annotations.

Once the edit phase is complete, user can save the playlist assigning full metadata to relate, in order to save it as an object that will be stored in database, will be available for the playback session, and will be shared with other users.

Thanks to the metadata, this object will be indexed properly, to keep easier the search on the portal.

Figure 2. Player loaded in order to choose the segment of media to add to playlist

C. *Collections*

A collection is a set of contents: audio, video, images, documents, pdf, playlists, zip files, etc., that are grouped together according to the personal purpose of the user. They may be thematic collection as well as used as the first step to collect content for preparing a lesson for e-learning environment, LMS. The teachers use the Collections as sources of contents that they can be grouped together following a topic or by choosing a connection amongst them.

The ECLAP Collections are directly created on the Portal and may be published or kept private by the user. Unpublished collection will be visible only to the creator in a draft form; Published collection has been uploaded as new ECLAP content and it will be visible to all registered users. In both cases, they are automatically exported, on the LMS side of ECLAP to be used by the teachers to create a lesson.

Figure 3. block where organize collections

D. *Mystoryplayer annotation tool*

MyStoryPlayer is a tool that allows users to be the central part in fruition of multimedia objects annotating them, and offers new solutions for educational and infotainment purposes.

The innovative part of MyStoryPlayer lies in the fact that no difference between media and the user's annotation exists, because both categories are referred to multimedia objects and they are temporally connected. Annotations are related to one or two media, and each media can be associated to many annotations. Therefore, according to the model, two audiovisual objects can be synchronized each other through an annotation containing temporal elements that define this kind of relation.

An annotation involves one or two media and is composed by a start time, an end time, an identifier of the media annotated, an identifier of the media to relate with the first one, and a textual description of the annotated segment. This information are codified by RDF triples and saved to a Sesame database, external to ECLAP one, interfaced with the Flash MyStoryPlayer that interprets the information and play the contents in synchronized way according to their temporal lines.

As for the other features like playlist and collection, no alteration is done on the original file, when two media are synchronized, because the system provide to generate new code RDF that identify this relation, and save on internal database the information about the annotation in order to allow the user to access to his annotations from his profile.

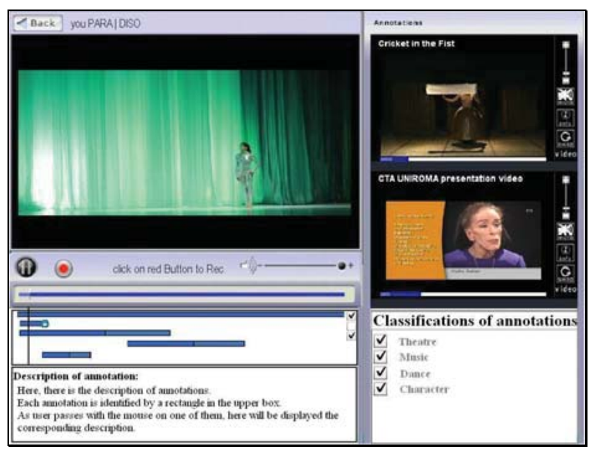


Figure 4. MyStoryPlayer interface

Acknowledgments

The authors would like to thanks the ECLAP partners, and to the European Commission. ECLAP is co-funded by the European Commission ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme, Theme CIP-ICT-PSP.2009.2.2.

Note for the demo: all the necessary equipment to present the demo will be brought by us. We will be present in our workstation at the conference, available for any question and/or explanation.

Reviewer's Index

M
Paolo Maresca
Andrea Molinari

N
Paolo Nesi
Kia Ng

O
Vincent Oria

P
Sethuraman Panchanathan
Ignazio Passero
Joseph J. Pfeiffer
Antonio Piccinno
Fabio Pittarello
Beryl Plimmer
Giuseppe Polese

R
Michele Risi
Teresa Roselli
Veronica Rossano

S
Nicoletta Sala
Faramarz Samavati
Giuseppe Santucci
Giuseppe Scanniello
Monica Sebillo
Buntarou Shizuki
Tom Stahovich
Peter Stanchev
Gem Stapleton
Mahbub Syed

T
Christian Timmerer
Genny Tortora

V
Stefano Valtolina
Athanasios Vasilakos
Giuliana Vitiello

W
Ellen Walker

X
Haiping Xu

Y
Kazuo Yana
Chi-Lu Yang
Chunsheng Yang
Maria Yang
Ersin Yumer

Z
Kang Zhang
Roberto Zicari

Author's Index

Poster/Demo Presenter's Index

DMS 2012 Call For Papers

The Eighteenth International Conference on Distributed Multimedia Systems

www.ksi.edu/seke/seke12.html

Eden Roc Renaissance Miami Beach, USA
August 9 - August 11, 2012

Organized by

Knowledge Systems Institute Graduate School

With today's proliferation of multimedia data (e.g., images, animations, video, and sound), comes the challenge of using such information to facilitate data analysis, modeling, presentation, interaction and programming, particularly for end-users who are domain experts, but not IT professionals. The main theme of the 18th International Conference on Distributed Multimedia Systems (DMS'2012) is multimedia inspired computing. The conference organizers seek contributions of high quality papers, panels or tutorials, addressing any novel aspect of computing (e.g., programming language or environment, data analysis, scientific visualization, etc.) that significantly benefits from the incorporation/integration of multimedia data (e.g., visual, audio, pen, voice, image, etc.), for presentation at the conference and publication in the proceedings. Both research and case study papers or demonstrations describing results in research area as well as industrial development cases and experiences are solicited. The use of prototypes and demonstration video for presentations is encouraged.

TOPICS

Distributed Multimedia Technology
- media coding, acquisition and standards
- QoS and Quality of Experience control
- digital rights management and conditional access solutions
- privacy and security issues
- mobile devices and wireless networks
- mobile intelligent applications
- sensor networks, environment control and management

Distributed Multimedia Models and Systems
- human-computer interaction
- languages for distributed multimedia
- multimedia software engineering issues
- semantic computing and processing
- media grid computing, cloud and virtualization
- web services and multi-agent systems
- multimedia databases and information systems
- multimedia indexing and retrieval systems
- multimedia and cross media authoring

Applications of Distributed Multimedia Systems
- collaborative and social multimedia systems and solutions
- humanities and cultural heritage applications, management and fruition
- multimedia preservation
- cultural heritage preservation, management and fruition
- distance and lifelong learning
- emergency and safety management
- e-commerce and e-government applications
- health care management and disability assistance
- intelligent multimedia computing
- internet multimedia computing
- virtual, mixed and augmented reality
- user profiling, reasoning and recommendations

Workshops and Special Sessions

The following workshops and special sessions are being planned:
International Workshop on Distance Education Technology (DET)
International Workshop on Visual Languages and Computing (VLC)
Multimedia Arts Workshop
Please contact the conference program co-chairs if you are interested in organizing a workshop or a special session. A one-page proposal with the following items is required:

Title of the workshop/special session
Name of the organizer(s)
A list of program committee members (reviewers)
E-mail address of the lead organizer
A brief description of the theme

Special issue of Journals

The DMS conference is closely coordinated with the International Journal of Software Engineering and Knowledge Engineering, the Journal of Visual Languages and Computing and the International Journal of Distance Education Technologies. Usually the best ranked papers from the conference, after rigorous reviews, extensive revisions and further enhancements, will appear in one or more special issues of the above journals. Papers suggested for the special issue(s) will be reviewed by external reviewers following the standard procedure of review stipulated by the respective journal. Welcome.

CONFERENCE SITE (HOTEL INFORMATION)

The DMS 2012 Conference will be held at the Eden Roc Renaissance Miami Beach, USA. The hotel is only USD$117 for DMS 2012 attendees for single/double, not including sales tax only these limited dates **8/6 - 8/13/2012**. Please join us in August 2012 at this most prestigious conference.

About The Eden Roc Renaissance Miami Beach Hotel

The Eden Roc Hotel is 300 yards away from the sea, The Eden Roc is more commonly known as 'seaside paradise'. To view the official website for the Eden Roc Hotel, please follow the provided link: http://www.marriott.com/hotels/travel/miasr-eden-roc-renaissance-miami-beach/. Also, please see the provided link which include exclusive pictures and brief details about the Eden Roc not available on the official Eden Roc website: http://www.ksi.edu/dms/Hotel_Brochure.pdf. The hotel is located on Miami's famous Collins Street in front of the breathtaking blue waters of Miami Beach's pristine glittering sea. Miami Beach is one of the 40 world's most famous sites to see. For a list of more attractions in Miami please follow this link: http://www.ksi.edu/dms/MiamiAttractions.pdf.

INFORMATION FOR AUTHORS

Papers must be written in English. An electronic version (Postscript, PDF, or MS Word format) of the full paper should be submitted using the following URL: http://conf.ksi.edu/dms2012/submit/SubmitPaper.php (submission site will be open Jan 1, 2012). Manuscript must include a 200-word abstract and no more than 6 pages of 2-column text (include figures and references). All submissions must not be published or under consideration for publication in a journal or in a conference with proceedings. Papers will be evaluated based on originality, significance, technical soundness and clarify of exposition. Depending upon the results of evaluation a paper may be accepted as regular paper (6 pages), short paper (4 pages) or poster (poster presentation only).

INFORMATION FOR REVIEWERS

Papers submitted to DMS'12 will be reviewed electronically. The users (webmaster, program chair, reviewers...) can login using the following URL: http://conf.ksi.edu/dms2012/review/pass.php
If you have any questions or run into problems, please send e-mail to:
E-mail: dms12@ksi.edu.

Conference Contact Information

DMS 2012 Conference Secretariat
Knowledge Systems Institute
3420 Main Street
Skokie, IL 60076 USA
Tel: 847-679-3135
Fax: 847-679-3166
E-mail: seke12@ksi.edu

IMPORTANT DATES

March 31, 2012	*Paper submission due*
May 20, 2012	*Notification of acceptance*
June 10, 2012	*Early registration deadline and Camera-ready copy*

DMS 2011

Proceedings of the Seventeenth International Conference on Distributed Multimedia Systems

Printed by
Knowledge Systems Institute
3420 Main Street
Skokie, Illinois 60076
(847) 679-3135
info@ksi.edu | www.ksi.edu
Printed in USA, 2011
ISBN 1-891706-30-6 (Paper)

Sponsored by:
Knowledge Systems Institute, USA

Co-Sponsored by:
Eco Controllo SpA, Italy
ICSRiM, University of Leeds, United Kingdom
Distributed Systems and Internet Technology Lab, University of Florence, Italy
Visual Design and Engineering Lab, Mechanical Engineering Department, Carnegie Mellon University, USA