Studies in Computational Intelligence 226

Ernesto Damiani
Jechang Jeong
Robert J. Howlett
Lakhmi C. Jain (Eds.)

New Directions in Intelligent Interactive Multimedia Systems and Services - 2

Springer

Ernesto Damiani, Jechang Jeong, Robert J. Howlett, and Lakhmi C. Jain (Eds.)

New Directions in Intelligent Interactive Multimedia Systems and Services - 2

Studies in Computational Intelligence, Volume 226

Editor-in-Chief
Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 205. Ajith Abraham, Aboul-Ella Hassanien, and Václav Snášel (Eds.)
Foundations of Computational Intelligence Volume 5, 2009
ISBN 978-3-642-01535-9

Vol. 206. Ajith Abraham, Aboul-Ella Hassanien, André Ponce de Leon F. de Carvalho, and Václav Snášel (Eds.)
Foundations of Computational Intelligence Volume 6, 2009
ISBN 978-3-642-01090-3

Vol. 207. Santo Fortunato, Giuseppe Mangioni, Ronaldo Menezes, and Vincenzo Nicosia (Eds.)
Complex Networks, 2009
ISBN 978-3-642-01205-1

Vol. 208. Roger Lee, Gongzu Hu, and Huaikou Miao (Eds.)
Computer and Information Science 2009, 2009
ISBN 978-3-642-01208-2

Vol. 209. Roger Lee and Naohiro Ishii (Eds.)
Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2009
ISBN 978-3-642-01202-0

Vol. 210. Andrew Lewis, Sanaz Mostaghim, and Marcus Randall (Eds.)
Biologically-Inspired Optimisation Methods, 2009
ISBN 978-3-642-01261-7

Vol. 211. Godfrey C. Onwubolu (Ed.)
Hybrid Self-Organizing Modeling Systems, 2009
ISBN 978-3-642-01529-8

Vol. 212. Viktor M. Kureychik, Sergey P. Malyukov, Vladimir V. Kureychik, and Alexander S. Malyoukov
Genetic Algorithms for Applied CAD Problems, 2009
ISBN 978-3-540-85280-3

Vol. 213. Stefano Cagnoni (Ed.)
Evolutionary Image Analysis and Signal Processing, 2009
ISBN 978-3-642-01635-6

Vol. 214. Been-Chian Chien and Tzung-Pei Hong (Eds.)
Opportunities and Challenges for Next-Generation Applied Intelligence, 2009
ISBN 978-3-540-92813-3

Vol. 215. Habib M. Ammari
Opportunities and Challenges of Connected k-Covered Wireless Sensor Networks, 2009
ISBN 978-3-642-01876-3

Vol. 216. Matthew Taylor
Transfer in Reinforcement Learning Domains, 2009
ISBN 978-3-642-01881-7

Vol. 217. Horia-Nicolai Teodorescu, Junzo Watada, and Lakhmi C. Jain (Eds.)
Intelligent Systems and Technologies, 2009
ISBN 978-3-642-01884-8

Vol. 218. Maria do Carmo Nicoletti and Lakhmi C. Jain (Eds.)
Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control, 2009
ISBN 978-3-642-01887-9

Vol. 219. Maja Hadzic, Elizabeth Chang, Pornpit Wongthongtham, and Tharam Dillon
Ontology-Based Multi-Agent Systems, 2009
ISBN 978-3-642-01903-6

Vol. 220. Bettina Berendt, Dunja Mladenic, Marco de de Gemmis, Giovanni Semeraro, Myra Spiliopoulou, Gerd Stumme, Vojtech Svatek, and Filip Zelezny (Eds.)
Knowledge Discovery Enhanced with Semantic and Social Information, 2009
ISBN 978-3-642-01890-9

Vol. 221. Tassilo Pellegrini, Sören Auer, Klaus Tochtermann, and Sebastian Schaffert (Eds.)
Networked Knowledge - Networked Media, 2009
ISBN 978-3-642-02183-1

Vol. 222. Elisabeth Rakus-Andersson, Ronald R. Yager, Nikhil Ichalkaranje, and Lakhmi C. Jain (Eds.)
Recent Advances in Decision Making, 2009
ISBN 978-3-642-02186-2

Vol. 223. Zbigniew W. Ras and Agnieszka Dardzinska (Eds.)
Advances in Data Management, 2009
ISBN 978-3-642-02189-3

Vol. 224. Amandeep S. Sidhu and Tharam S. Dillon (Eds.)
Biomedical Data and Applications, 2009
ISBN 978-3-642-02192-3

Vol. 225. Danuta Zakrzewska, Ernestina Menasalvas, and Liliana Byczkowska-Lipinska (Eds.)
Methods and Supporting Technologies for Data Analysis, 2009
ISBN 978-3-642-02195-4

Vol. 226. Ernesto Damiani, Jechang Jeong, Robert J. Howlett, and Lakhmi C. Jain (Eds.)
New Directions in Intelligent Interactive Multimedia Systems and Services - 2, 2009
ISBN 978-3-642-02936-3

Ernesto Damiani, Jechang Jeong, Robert J. Howlett,
and Lakhmi C. Jain (Eds.)

New Directions in Intelligent Interactive Multimedia Systems and Services - 2

Springer

Prof. Ernesto Damiani
Università degli Studi di Milano
Dipto. Tecnologie dell'Informazione
Via Bramante, 65
26013 Crema
Italy
E-mail: damiani@dti.unimi.it

Prof. Robert Howlett
University Brighton
School of Engineering
Intelligent Signal Processing
Laboratories (ISP)
Moulsecoomb
Brighton, BN2 4GJ
UK
E-mail: r.j.howlett@brighton.ac.uk

Prof. Jechang Jeong
Department of Electronic and Computer Engineering
College of Engineering
Hanyang University
17 Haengdang-dong
Seongdong-gu, Seoul 133-791
Korea
E-mail: jjeong@ece.hanyang.ac.kr

Prof. Lakhmi C. Jain
School of Electrical and Information Engineering
University of South Australia
Adelaide, Mawson Lakes Campus
South Australia SA 5095
Australia
E-mail: Lakhmi.jain@unisa.edu.au

ISBN 978-3-642-02936-3 e-ISBN 978-3-642-02937-0

DOI 10.1007/978-3-642-02937-0

Studies in Computational Intelligence ISSN 1860-949X

Library of Congress Control Number: Applied for

Typeset & *Cover Design:* Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

This volume contains the Proceedings of the 2nd International Symposium on Intelligent Interactive Multimedia Systems and Services (KES-IIMSS 2009) This second edition of the KES-IIMSS Symposium was organized by the Department of Information Technologies of University of Milan, Italy in conjunction with Hanyang University, Korea and KES International. KES-IIMSS is a new series of international scientific symposia aimed at presenting novel research in the fields of intelligent multimedia systems relevant to the development of a new generation of interactive, user-centric services. The major theme underlying this year's symposium is the rapid integration of multimedia processing techniques within a new wave of user-centric services and processes. Indeed, pervasive computing has blurred the traditional distinction between conventional information technologies and multimedia processing, making multimedia an integral part of a new generation of IT-based interactive systems. The aim of the KES-IIMSS symposium, following the general structure of KES events, is to provide an internationally respected forum for publishing high-quality results of scientific research while allowing for timely dissemination of research breakthroughs and novel ideas via a number of autonomous special sessions and workshops on emerging issues and topics identified each year. IMSS 2009 co-located events include the International Workshop on Human-Computer Interaction in Knowledge-based Environments, and three invited sessions respectively on Intelligent Systems for Healthcare, Design of Intelligent Environments for Supporting Human Interaction and Multimedia Techniques for Device and Ambient Intelligence (MTDAI). The MTDAI session follows in the footsteps of a yearly by-invitation-only workshop traditionally organized by the University of Milan and held in the beautiful Villa Braida that this year has hosted the entire set of KES IIMSS events. KES IIMSS 2009 final program is composed of a General Track featuring 16 full papers by authors from four continents and 10 countries. These papers were selected via a careful refereeing process among 45 submissions, yielding a 35% acceptance rate. Satellite events include a total of session and workshop papers. We are very satisfied of the quality of the program and would like to thank the authors for choosing KES-IIMSS as the forum where to present their work. Also, we gratefully acknowledge the hard work of KES IIMSS international program committee and of the additional reviewers for selecting the accepted conference papers. We wish to thank the session and workshop organizers and authors for proposing a stimulating side program, hopefully capturing some the spirit and lively discussions of the original MTDAI event. We wish to express our gratitude to the KES Secretariat, for their help in the organizational issues. Thanks are also due to the local arrangement team, including Marco Anisetti, Valerio Bellandi and Olga Scotti. Finally, we would like to point out KES-IMSS 2009 impressive set of keynote speakers, including outstanding researchers like Borko Furht, Harald Kosh, Paolo Nesi and Vincent Oria. Their presence provided to KES-IMSS 2009 a unique opportunity for setting the research agenda of this exciting new field.

General Co-Chairs
Ernesto Damiani
Jechang Jeong

Programme Chairs
Gabriella Pasi
Roch Glitho

Executive Chair
Robert J. Howlett

KES Founder
Lakhmi C. Jain

Contents

Cost-Efficient Development of Virtual Sales Assistants

Markus Zanker, Marcel Bricman, and Markus Jessenitschnig

Abstract The metaphor of a *virtual sales assistant* symbolizes an interactive Web application that serves online shoppers just as their human counterparts in brick & mortar businesses would do. They help to create an enjoyable shopping experience for users and are intended to bring a social and emotional perspective to e-commerce environments. Although, researchers have investigated dozens of interesting pilot applications, few have actually been commercialized. There is a big gap between the technological state of the art and what pays off for e-commerce sites from an economical point of view. Based on our industrial experience we discuss success factors for the development of cost-efficient and persuasive virtual sales assistants and contribute a framework for conversational sales recommender systems that includes a design environment for the efficient creation of virtual characters. Furthermore, this paper reports on experiences gathered from deployed applications.

1 Introduction

The notion of *virtual sales assistants* has become commonplace on many online shops over the past decade, with the promise of Web applications serving online customers in the same way as the sales persons of a brick & mortar business. In reality, however, this promise is rarely kept. Many online shops misuse the term,

Markus Zanker and Markus Jessenitschnig
University Klagenfurt,
Intelligent Systems and Business Informatics Group,
9020 Klagenfurt, Austria
{markus.zanker,markus.jessenitschnig}@uni-klu.ac.at

Marcel Bricman
KCI-Group,
9500 Villach, Austria
mbricman@kci-group.com

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 1–11.
springerlink.com

offering only simple search tools or static query forms that do not provide personalized interactive behavior - let alone a social or emotional perspective.

Technically, a virtual sales assistant is a personified e-commerce shopping agent that converses with online users in a knowledgeable way and introduces them to items that might be of interest to them. Conversational interfaces facilitate multimodal interaction between the user and the online application [1]. In addition to question and answer style dialogues, different forms of user feedback can be observed to reason about users' preferences and their situational context. For instance, the browsing behavior or explicit ratings and critiques on viewed items provide information on what they have already seen, what they like or dislike and how knowledgeable they are in the specific domain. Furthermore, hypertext can be generated on the fly, multimedia content displayed and additional navigation opportunities introduced. Thus the agent's behavior and the users' shopping experience is determined by the adaptivity and responsiveness of the interaction mechanism and the reasoning engine's ability to compute recommendations.

The emotional perspective is typically related to the personification of such a conversational sales agent. The options for personifications range from artificial cartoon-like or anthropomorphic characters to depictions of real humans. Furthermore, the degree of animation ranges from simple picture files to full-fledged three dimensional (3D) characters being capable of mimics and gestures. Compare to Figure 1 for several examples like the legal advisor *METIS* [2], the lingubots *Miss Suzuki* and *Frank* created by [3] a major commercial provider for customer service solutions or a group of virtual shop and service assistants developed by the authors. The personified sales agent forms not only part of the communication strategy of a

Fig. 1: Examples of virtual characters

company but also is responsible for ensuring that users have an enjoyable shopping experience. However, building an intelligent conversational sales agent, providing him or her with a visually attractive appearance and ensuring that the system can be efficiently maintained and updated by the shop owner is not an easy task and quite costly. Currently, only large online retailers can afford such investments and deploy virtual sales agents for mainly their prestige. Hundreds of thousands of small and medium-sized online shops are economically excluded from offering a knowledgeable sales advisor to their clients. This paper therefore discusses practical considerations and success factors for deployment of sales agents in the field. Its contribution lies in giving details of a commercialized development framework for virtual sales agents that addresses the issues of cost-efficiency and short time to market.

The paper is structured as follows: first it discusses recommendation techniques and conversational sales assistants. Section 3 elaborates on factors that affect commercial roll-outs of recommendation technologies and virtual sales agents. Furthermore, we present a framework for the development of virtual characters and discuss its integration within a conversational recommendation system. Finally, the paper reports on commercial experiences and concludes by giving an outlook on future work.

2 Background

Previous research within the field of conversational online sales support systems has mainly focused on three different approaches: static and dynamic flow of predefined input forms [4, 5, 6] as well as natural language interaction [1]. Predefined input forms mimic the preference elicitation dialogues of real sales persons for specific tasks such as providing advice about a product category or finding an appropriate present. Systems with static interaction paths ask all users the same sequence of questions, while a dynamic interaction flow personalizes page content and flow [7]. Recently, Mahmood and Ricci [8] proposed an adaptive interaction strategy that employs reinforcement learning techniques to identify the optimal sequence of actions for each user. Systems that support natural language interaction allow users to pose free questions. However, the answering capabilities of such systems depend on the knowledge engineering effort that has been invested when setting up the system [1]. Comparable to the early ELIZA system [9] natural language based virtual assistants - termed lingubots - parse the user input for keywords and patterns or apply transformation rules to rephrase the user input. Based on their set of rules lingubots attempt to mimic ”intelligent” behavior by matching keywords from the user input to their rule base and reply for instance using a predefined answer or applying generic transformation rules to reuse the keywords.

Lingubots and natural language interaction are particularly successful in online complaint management systems [10] and in question answering applications. Leuski et al. [11] give an outlook on mining techniques to find appropriate answers for users' questions. In our applications we employ a dynamic form-based dialogue component as described in [6, 12].

The core recommendation capability, i.e. how product proposals are determined and argued to the customer, is another characteristic of virtual sales assistants. Knowledge-based technology and case-based recommender systems [4, 5, 13, 6] are widely used for conversational shopping agents, building on explicit preference elicitation from users. This is in contrast to the most prominent variants of recommender systems in the literature, namely collaborative and content-based filtering systems [14, 15], which in their purest forms are not conversational but follow instead a *one-shot* interaction paradigm. Hybrid approaches have recently been developed [16] that combine several of these techniques and may also possess conversational capabilities. The *ISeller* recommendation system [17] developed by our research group supports several recommendation strategies and is in addition capable of hybridizing these strategies when needed [18, 19].

Another issue related to our work is the online users' perception of the virtual agents. The personification of virtual sales assistants and their interactive and personalized behavior impacts the way users treat them. Reeves and Nass [20] worked on this topic and laid the foundations for new research fields, such as persuasive computing [21], a research area at the intersection of technology and social acting. Computing applications that influence people's attitudes or behavior are also increasingly in the center of ethical discussions [22]. Heckman and Wobbrock discuss the legal implications like who is liable for anthropomorphic agents that mislead consumers [23]. The focus of the work presented here lies in developing virtual characters that act as personified sales agents for the Web. Few empirical studies exist that research the impact of personified agents on Web stores. Komiak et al. [24] compared trust in virtual salespersons with trust in their human counterparts. One of the most interesting findings is that virtual salespersons were considered to be slightly more trustworthy than human ones. McBreen and Jack [25] compared three-dimensional human-like agents with two-dimensional cartoon-like agents. According to their user study three-dimensional realistic agents performed best.

In the following we will discuss practical considerations and requirements for the installation of an online shop assistant.

3 Requirements

On the one hand the development of virtual sales assistants faces ambitious requirements in terms of visual appearance and interaction design; but on the other hand technology providers face stringent budget and time restrictions, particularly when developing small and medium sized online merchants. Based on our industrial experience in the field, the following issues are of great importance:

- Integration
- Ramp-up efforts
- Deployment time

- Maintainability

Small online shops are characterized by a high degree of technical heterogeneity. The majority of Web shop systems only support proprietary product and customer data formats, hindering interoperability. However, the **integration** of a virtual shop assistant requires access to existing data sources. First of all, product data is needed and furthermore most recommendation techniques require a structured representation of items. Only pure collaborative filtering approaches do not require any information on the product domain except unique item identifiers. In addition, customer records and historic transaction data are often of interest for personalizing the interactions as well as the recommendations. This data has to be derived from the e-shop system itself or from a CRM database.

The **ramp-up efforts** are another important issue. The initial development cost of customizing and instantiating a virtual sales assistant for a concrete problem instance are crucial. They encompass the data integration effort, the construction of a recommendation strategy as well as graphical character design. However, preconfigured virtual characters are typically not successful at convincing online users as they do not match with the design and marketing strategy of the specific etailer. Therefore, an additional challenge is to provide unique virtual characters at a moderate cost.

Deployment time - i.e. the duration of the development process including the time required for the creation of a specific virtual character. Most web applications are expected to be released after a few months. Hence, rapid prototyping and early evaluation by potential users are necessary.

Maintenance is another crucial issue in rapidly changing environments. Product prices and portfolio change continuously, while product descriptions themselves may need to be updated, e.g. to provide details of new extensibility options or features. Deeper product knowledge like *'How many megapixels does a state-of-the-art digital camera have?'* must be evaluated on a periodical basis. With respect to the maintenance task itself, easy administration of the system - ideally by the shop owner is a key requirement.

Concluding, all aforementioned aspects relate to the *cost of ownership* of a virtual sales assistant. Thus, cost-efficiency is key for widespread application of the technology. However, the persuasiveness of the system - i.e. influencing users such that they are more likely to buy - does not come free of charge: designing an attractive appearance and intelligent system behavior are major cost drivers. In the next section we will present a development framework that focuses on balancing the persuasive traits of the visual appearance and cost-efficiency.

4 Implementation

The framework for developing virtual sales assistants follows a knowledge-based system development process. The reasoning knowledge for the sales advisory tasks as well as all system parameters are declaratively maintained using an integrated editing environment. The user-interfaces themselves are automatically generated using template and style files as described in [12]. Figure 2 presents the main system components such as *Knowledge Acquisition* and *GUI Generation* which collectively define conversation paths at design time. The *Advisor Suite Server* and the *Interaction & Personalization Agent* are, on the other hand, responsible for managing the user interaction and deriving personalized recommendations at run time. The *Advisor Suite* system also includes an integration and preprocessing layer to further ease setup and maintenance tasks. Furthermore, a design environment for the development of virtual characters was developed and is also presented in this section.

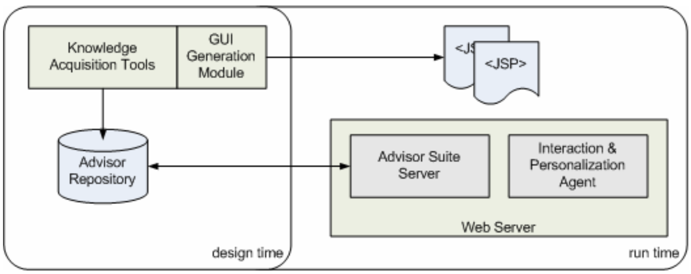


Fig. 2: Architecture and system setup

4.1 Integration layer

The integration layer is responsible for accessing preexisting data sources when the *Advisor Suite* system is released as part of an e-commerce shop, such as the product database or a collection of customer records. The maintenance environment is based on the Eclipse Rich-Client Platform[1] (RCP) which allows the system's functionality to be easily extended by introducing additional plug-ins. During installation, external data sources such as XML files or relational database tables can be chosen. Comparable to the knowledge discovery process in databases, the system operator selects from a set of available actions for data transformation and pre-processing, such as up/down-scaling of numbers, transforming strings to numbers or relationally joining and transposing the input data structures [17].

[1] See www.eclipse.org for details.

4.2 Recommendation strategy

The recommendation component of the *Advisor Suite* system operates on product data, questions and the dialog flow (stored in the *Advisor Repository*) and provides personalized advice and item propositions to users (*Interaction & Personalization Agent* in Figure 2). At design time the domain specific application logic is acquired by the *Knowledge Acquisition Tools* from a domain expert or system operator. The dynamic page flow is represented as a directed graph whose nodes are question pages and edges are associated with a series of conditions required for a transition to the next page. Based on a template mechanism, the *GUI Generation Module* generates a Java Server Page (JSP) for each node at design time. When the *Interaction & Personalization Agent* identifies at runtime that a specific question needs to be invoked, the respective JSP renders the HTML including user-specific content (i.e. questions, product presentations and explanations) and invokes the API of the virtual character framework.

4.3 Virtual characters

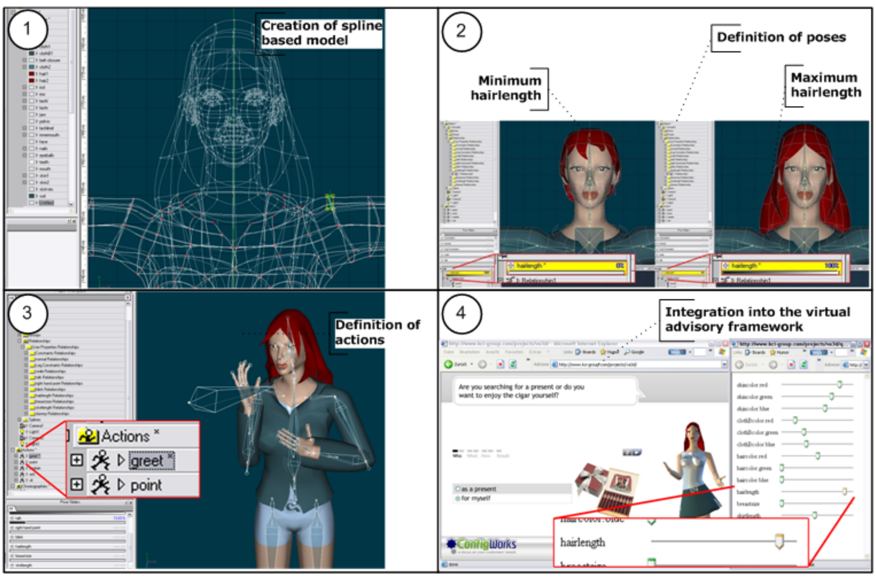


Fig. 3: Character development process

The personification of the recommendation system is implemented using an artificial character. The character development process is outlined in Figure 3. The character creation framework defines a pipeline of inexpensive but highly efficient and effective tools that facilitate reuse as the primary means of optimizing throughput. First, the characters are built as models in a spline-based 3D modeler (see Figure 3 - Nr. 1) and fitted with a bone structure. In Figure 3 - Nr. 2, the structure is then modified to create poses, for instance characters with different lengths of hair. The poses hold modification data for the model and can later be applied to the model to provide subtle alterations. Furthermore, the model is animated to create actions, such as greeting users when they start a dialogue as displayed in Figure 3 - Nr. 3.

For characters with a similar bone model, the same poses and actions are reusable. This collection of work assets (model, poses and actions) are then exported into a realtime 3D format for web integration.

The virtual character is fully implemented on the client-side, so the basis of the implementation is the web browser and the HTML output of the JSP for the particular page that is invoked by the server-side *Interaction & Personalization Agent*. This output includes API calls to the character framework which control the behavior of the virtual agent. For instance, before asking the first question to the user the character performs a *greet* action or after a period without user input has elapsed a randomly selected *idle* action is performed. This assignment of actions and poses to question pages is thus done at design time as part of the initial knowledge acquisition effort. The framework is open to different implementations, technologies, and even host applications. Conceptually, the user dialog is sequenced into questions, each of which is represented by a system-rendered HTML page. The framework API calls contained in these pages are scheduled as actions for the virtual character. After a specified waiting period these actions are passed up to the implementation layer, which actually makes the corresponding interface changes.

Figure 3 - Nr. 4 displays a question page with an embedded virtual character. The configuration window on the right side displays different sliders that allow the parametrization of the character's appearance. Using server-side API calls both, the character's actions and its appearance (different poses) can be dynamically adapted to match the users' requirements and contexts.

In our implementation the rendering of the advisor is achieved using Macromedia Flash technology. The implementation layer passes commands via Macromedia LiveConnect to the Flash container, effectively telling it which emotion to display. The division of the dialog into singular questions might seem restrictive for the implementation of recommender systems, but the resulting tunneling of the user dialog strengthens the complexity reduction principle. A very important aspect of the framework is compatibility with a wide range of target platforms. While the technology choices ensure flexibility, the framework implementation can further extend this concept by adding fallback measures. For example, if the client system does not support javascript, HTML files are still accessible.

5 Experiences from commercial projects

More than 10 different conversational recommendation applications have been successfully deployed online. The earliest of them was *Mortimer*, who advises its users on cigars and premium coffee blends. In [26] the question of whether the advice given by the virtual shop assistant affects the buying behavior of online shoppers was investigated. The study compared the sales records of a webshop for premium cigars in the period before the introduction of the conversational recommender system with those in the period afterwards. One interesting finding was that the lists of top ten ranked products in the two periods differed considerably. It turned out that the cigars that the system had recommended typically showed an increasing demand, i.e. the increase in sales figures was positively correlated with the recommendation frequency of the system.

Further evidence that virtual sales assistance helps to convert surfing online users into customers is provided by a case study on the female online assistant *vibe* in the tourism domain [27]. The investigation confirmed that the *looker-to-booker* rate of users that interacted with an interactive travel advisor is often more than twice that of those that didn't. In their work [28], Jannach and Bundgaard-Joergensen report on initial practical experiences with our framework from a rather unusual recommendation domain: advising on investor-ready business plans. The application features an optionally male or female animated character that guides users through a series of questions on their business plan and which have now been completed by more than one thousand users. The completion rate for the rather lengthy dialogue with over 30 questions lies around 50% which is remarkably high in the context of an online service.

6 Conclusions

Based on our industrial experience, this paper discussed the cost-efficient development of virtual sales assistants. It presented extensions to the *Advisor Suite* system which counteract the apparent contradiction between cost-driving system traits like attractive appearance or intelligent system behavior and stringent budget restrictions that promote the use of generic knowledge-based system development and automated code generation approaches. Specifically, the paper's contribution lies in targeting the issue of cost-efficiency in order to assure the widespread applicability of virtual sales advisory systems as well as extending an existing system as proof of concept.

References

1. Lucente, M.: Conversational interfaces for e-commerce applications. Communications of the ACM 43(9), 59–61 (2000)
2. Wize, C.: So kuenstlich dass es echt wirkt. Manz (2004)
3. Artifical Solutions, http://www.artificial-solutions.com
4. Burke, R.D., Hammond, K.J., Young, B.C.: The findme approach to assisted browsing. IEEE Expert, 32–40 (July/August 1997)
5. Ricci, F., Werthner, H.: Case base querying for travel planning recommendation. Information Technology and Tourism 3, 215–266 (2002)
6. Felfernig, A., Friedrich, G., Jannach, D., Zanker, M.: An integrated environment for the development of knowledge-based recommender applications. International Journal of Electronic Commerce 11(2), 11–34 (2006)
7. Jannach, D., Kreutler, G.: Personalized User Preference Elicitation for e-Services. In: Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE) (2005)
8. Mahmood, T., Ricci, F.: Adapting the interaction state model in conversational recommender systems. In: Proceedings of the 10th International Conference on Electronic Commerce, pp. 1–10 (2008)
9. Weizenbaum, J.: Computational linguistics. Communications of the ACM 9(1), 36–45 (1966)
10. Angeli, A.D., Johnson, G.I., Coventry, L.: The unfriendly user: exploring social reactions to chatterbots. In: Helander, Khalid, Tham (eds.) International Conference on Affective Human Factors Design. Asean Academic Press, London (2001)
11. Leuski, A., Patel, R., Traum, D., Kennedy, B.: Building effective question answering characters. In: 7^{th} SIGdial Workshop on Discourse and Dialogue, pp. 18–27. Springer, Heidelberg (2006)
12. Jannach, D., Kreutler, G.: Rapid development of knowledge-based conversational recommender applications with advisor suite. Journal of Web Engineering 6(2), 165–192 (2007)
13. McCarthy, K., Reilly, J., McGinty, L., Smyth, B.: Experiments in Dynamic Critiquing. In: Proceedings of the Intelligent User Interfaces Conference (IUI), pp. 175–182 (2005)
14. Balabanovic, M., Shoham, Y.: Fab: Content-based, collaborative recommendation. Communications of the ACM 40(3), 66–72 (1997)
15. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Analysis of recommendation algorithms for e-commerce. In: ACM Conference on e-Commerce (EC), pp. 158–167 (2000)
16. Burke, R.: Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction 12(4), 331–370 (2002)
17. Zanker, M., Jessenitschnig, M.: ISeller - A Generic and Hybrid Recommendation System for Interactive Selling Scenarios. In: 15^{th} European Conference on Information Systems (ECIS), Gallen, Switzerland, pp. 70–80. Springer, Heidelberg (2006)
18. Zanker, M., Jessenitschnig, M., Jannach, D., Gordea, S.: Comparing recommendation strategies in a commercial context. IEEE Intelligent Systems 22, 69–73 (2007)
19. Zanker, M., Jessenitschnig, M.: Case-studies on exploiting explicit customer requirements in recommender systems. In: Tuzhilin, A., Mobasher, B. (eds.) User Modeling and User-Adapted Interaction: The Journal of Personalization Research. Special issue on Data Mining for Personalization, vol. 19(1-2), pp. 133–166 (2009)
20. Reeves, B., Nass, C.: The Media Equation - How People treat Computers, Television, and New Media Like Real People and Places. CSLI Publications, Center for the Study of Language and Information, New York (1996)

21. Fogg, B.J.: Persuasive Technologies. Communications of the ACM 42(5), 27–29 (1999)
22. Berdichevsky, D., Neuenschwander, E.: Toward an Ethics of Persuasive Technologies. Communications of the ACM 42(5), 51–58 (1999)
23. Heckman, C.E., Wobbrock, J.O.: Put your best face forward: Anthropomorphic agents, e-commerce consumers and the law. In: Autonomous Agents and Multiagent Systems, pp. 435–442. Springer, Heidelberg (2000)
24. Komiak, S., Wang, W., Benbasat, I.: Comparing Customer Trust in Virtual Salespersons With Customer Trust in Human Salespersons. In: Proceedings of the 38^{th} Hawaii International Conference on System Sciences (HICSS) (2005)
25. Breen, H.M., Jack, M.A.: Evaluating humanoid synthetic agents in e-retail applications. IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans 31(5), 394–405 (2001)
26. Zanker, M., Bricman, M., Gordea, S., Jannach, D., Jessenitschnig, M.: Persuasive online-selling in quality & taste domains. In: Proc. 7th International Conference on Electronic Commerce and Web Technologies (EC-Web), Krakow, Poland, pp. 51–60. Springer, Heidelberg (2006)
27. Zanker, M., Fuchs, M., Höpken, W., Tuta, M., Müller, N.: Evaluating recommender systems in tourism - a case study from austria. In: Proceedings of the International Conference on Information and Communication Technologies in Tourism (ENTER), pp. 24–34 (2008)
28. Jannach, D., Bundgaard-Joergensen, U.: Sat: A web-based interactive advisor for investor-ready business plans. In: Proceedings of International Conference on e-Business, pp. 99–106 (2007)

Security of Linguistic Threshold Schemes in Multimedia Systems

Marek R. Ogiela and Urszula Ogiela

Abstract This publication will present a new approach how to extend well-known algorithms of secret sharing, towards another stage of information encoding with the use of the grammar formalism. Such an algorithm would be based on the appropriate sequential LALR grammars allowing shared bit sequences, and more generally blocks of several bits, to be changed into new representations, namely sequences of production numbers of the introduced grammar. This stage can be executed by a trusted third party or arbiter generating shadows of the secret. Such methods would form an additional stage improving the security of shared data.

1 Introduction

Advanced information management systems use various techniques and methods to manage data, search for it semantically, classify it as secret, make meaning tables etc. They are designed to facilitate the access to and improve the effectiveness of finding important information with specific meaning. Cryptographic algorithms for threshold secret splitting and sharing are among such techniques. They can be used to split important information and to assign its components to people from an authorised group. Such authorised, selective access to information is used when it is necessary to safely manage strategic data. This information may be military, but also economic data or multimedia files. This last type of data is gaining increasing importance due to the fast development various multimedia platforms or modalities. Sometimes such data is classified and inaccessible to ordinary people. Within the structure of the specific organisation there are individuals at the appropriate management levels who have access rights to the data addressed to them. Such rights are exercised in hierarchic structures, usually connected with the office held. In practice,

AGH University of Science and Technology,
Al. Mickiewicza 30, PL-30-059 Krakow, Poland
e-mail: {mogiela,ogiela}@agh.edu.pl

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 13–20.
springerlink.com

this means that higher-placed individuals have access to more confidential data, and people at lower levels to less information. Consequently, the flow of information within such structures may require implementing hierarchical threshold schemes for secret and data splitting, which schemes assign the appropriate level of rights to individuals who want to receive authorised access to secret data at particular levels. Obviously, when talking of information management, we refer to data stored on digital media or in computer databases. For such data, there is a need to intelligently split it between the authorised individuals and then to reconstruct it in secret. Therefore, it is worth turning our attention to the other significant question related to intelligent information management. It is the question of the capacity to ensure secrecy and selective access to such data for the authorised persons. Such a potential of managing strategic information may be acquired thanks to the use of certain mathematical techniques, originating from the fields of cryptography and steganography. In our case, the task comes down to searching for the formulas that allow intelligent sharing of information in a way that would allow its reconstruction to appropriately authorised people. The only condition here is the possibility of splitting the data and later their reconstruction by a group of appropriately authorised people. In this work we make an attempt to enrich known algorithms for secret sharing, with an additional stage of splitting the linguistic representation, defining the split data in the binary form. To achieve this, a sequential grammar of LALR type is introduced to allow converting a sequence of bits into its linguistic representation. This representation will then be subject to sharing with the use of one of the known threshold schemes.

2 An Idea of Secret Sharing

Secret sharing and splitting algorithms are quite young branch of information technology and cryptography. In the most general case, their objective is to generate such parts for the data in question that could be shared by multiple authorised persons. What arises here is the problem of splitting information in a manner allowing its reconstruction by a certain n-person group interested in the reconstruction of the split information. Algorithm solutions developed to achieve this objective should at the same time make sure that none of the groups of participants in such a protocol, whose number is lesser than the required m persons, could read the split message. The algorithms for dividing information make it possible to split it into chunks known as shadows that are later distributed among the participants of the protocol so that the shares of certain subsets of users, when combined together, are capable of reconstructing the original information. There are two groups of algorithms for dividing information, namely, *secret splitting* and *secret sharing*. In the first technique, information is distributed among the participants of the protocol, and all the participants are required to put together their parts to have it reconstructed. A more universal method of splitting information is the latter method, i.e. *secret sharing*. In this case, the message is also distributed among the participants of the protocol, yet to have it reconstructed it is enough to have a certain number of constituent shares defined while building the scheme. Such algorithms are also known as threshold schemes, and were proposed independently by A. Shamir [13] and G. Blakley [5] [6], and were thoroughly analysed by G. Simmons [14]. The next section describes a method of extending classical threshold schemes for secret sharing to include an additional linguistic stage at which binary representations of the shared secret are coded into new sequences representing the rules of a formal grammar introduced. Such stage will introduce additional security against the unauthorised reconstruction of the information and can be executed in two independent versions of protocols for assigning created shadows to protocol participants. The first one is the version involving a trusted arbiter to mediate in the assignment and reconstruction of information. The second is the version without the arbiter, but with the assignment of the introduced grammar as a new, additional part of the secret.

3 Protocol for Linguistic Threshold Schemes

Executing the secret sharing protocol with the use of sequential grammars will lead to generating one additional shadow for the shared information. As already mentioned above, depending on the function of this information element, you can execute an arbitration protocol with a trusted arbiter holding the linguistic component necessary to reconstruct the secret, or a simple protocol without an arbiter. However, in the second case, the person holding the linguistic information (the grammar rule set) will be privileged, as his/her information will always be necessary to reconstruct the original secret. This situation will depend on the scheme executed and will be independent of the selected threshold secret sharing algorithm. This version can be beneficial when executing hierarchical threshold schemes, i.e. schemes with privileged shares. However, if it is necessary to create a fair, equal threshold scheme, the generation rule set of the grammar can be made public and then all shadows will have exactly equal rights. Further, this work proposes an algorithm for expanding the operation of such schemes and generation of a single additional shadow in the form of linguistic information necessary for the reconstruction of the entirely secret. Main steps of using the grammatical approach to the expansion of threshold systems are presented in Fig. 1.

4 Grammar for Coding Bit Positions

Expansion of the threshold scheme by an additional stage of converting the secret recorded in the form of a bit sequence is performed thanks to the application of context-free grammar in the following formula:

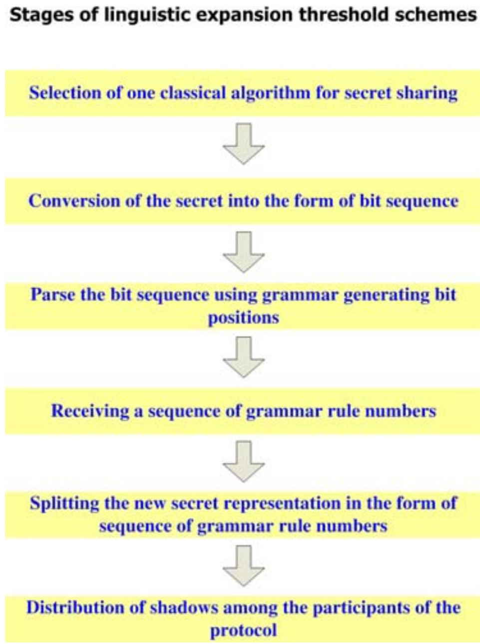


Fig. 1: Methodology of linguistic expansion threshold schemes.

$$G_{SECRET} = (N, T, P, STS), \quad (1)$$

where:
$N-$ non-terminal symbols, $T-$ terminal symbols, $\varepsilon-$ an empty symbol, $STS-$ grammar start symbol, $P-$ is a production set. Depending on the production set such grammar can change the bit sequences in the form of zeros and ones into a sequence of grammar production numbers that allow the generation of the original bit sequence.
The conversion of representation is ensured through syntax analyser that changes the bit sequence into numbers of linguistic rules of the grammar in square time. The graphic representation of using the grammar expansion in classical threshold schemes is presented in Fig. 2.
After performing such a transformation, any scheme of secret sharing can be applied to distribute the constituents among any number of n participants of the protocol. This means that at this stage, any classical (m, n)-threshold algorithm for secret sharing can be run. However, the secret being split is not a pure bit sequence, but a sequence composed of numbers of syntactic rules of the introduced grammar. Depending on its structure and type, it can contain values of two or more bits. So you can imagine a situation in which the grammar conversion will not consist in transforming single bits but also transforming pairs or greater numbers of bits at the same time (i.e. values of two, three, four and more bits will be considered). In that case,

the structure of the grammar will be similar, but the sequence of generation rule numbers obtained will have a greater range of values (i.e. the number of generation rules of the grammar defined for the conversion will increase).

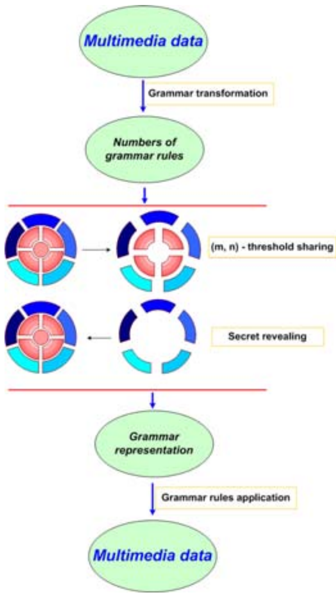


Fig. 2: Shadow generation scheme in the expanded threshold algorithm. The expansion concerns the use of grammar at the stage of converting the bit representation into sequences of numbers of linguistic rules in grammar.

To illustrate the idea of such linguistic coding, an example of a grammar that converts 3-bit clusters is presented. Such a grammar can be defined as follows:

$$G_{3\ BIT} = (N, T, P, STS), \tag{2}$$

where:
$N = \{SECRET, LN, 3B\}-$ non-terminal symbols,
$T = \{000, 001, 010, 011, 100, 101, 110, 111, \varepsilon\}-$ terminal symbols which define each 3-bit value,
$\varepsilon-$ an empty symbol.
$STS =$ SECRET - grammar start symbol.

A production set P is defined in following way:

1. $SECRET \rightarrow LN$

2. $LN \rightarrow$ 3B LN

3. $LN \rightarrow \varepsilon$

4. $3B \rightarrow 000|001|010|011$

5. $3B \rightarrow 100|101|110|111$

A grammar introduced in this way can support a quicker and briefer recoding of the input representation of the secret to be shared. Versions for longer bit blocks can be used in the same way. However, this will require introducing a greater number of linguistic rules. An obvious benefit of grouping bits into larger blocks is that during the following steps of the secret sharing protocol we get shorter representations for the split data. This is particularly visible when executing procedures that use excessive bit representations, i.e. when single-bit or several-bit values are saved and interpreted using codes in 8 or 16-bit representations. As said previously executing the introduced algorithms provides an additional stage for recoding the shared secret into a new representation using grammatical rules. The grammar itself can be kept secret or made available to the participants of the entire protocol. If the allocation of grammatical rules is to remain secret, as mentioned earlier, what we deal with is an arbitration protocol, which - to reconstruct the secret for the authorised group of shadow owners - requires the participation of a trusted arbiter, equipped with information about grammar rules. Should the grammar be disclosed, the reconstruction of the secret is possible without the participation of the trusted person and only on the basis of the constituent parts of the secret kept by the authorised group of participants in the algorithm of information sharing.

5 Conclusion

In this work we present the potential way of expanding clasic threshold secret sharing schemes towards the linguistic descriptions that allow obtaining additional representations that improve the security of the information being split. Linguistic representations were achieved as a result of using sequential grammars that allow conversion of bit representation of the shared secret to the form of a series of numbers of grammatical rules. Such a conversion to the linguistic form is possible thanks to the use of an analyser of polynominal complexity. The possibility of establishing new types of arbitration protocols is the result of introducing linguistic descriptions to the schemes used. The arbitration protocol operates when the rules of the introduced grammar remain secret and are stored with a trusted arbiter. In this case, however, what is necessary to reconstruct the secret is the participation of the arbiter, who will have to disclose his share (being the rules of grammar). Another solution is developing an extended scheme in the case when the grammar defined is public. In such a case, the secret split has the form of a series of grammar production numbers.

Such a presentation is shared by all the participants of the protocol with the same authorisation. The authorised subset of generated shadows allows for the composition of the secret, and the knowledge of the grammatical rules allows for converting this secret into the form of a bit, and later numerical or text, sequence.
The research conducted in this field by the author is focused on the definition of methodology and effective means of using threshold techniques for information sharing for multilevel, intelligent management of strategic or multimedia data stored in digital form. The implementation of the method described allows using mathematical techniques for tasks from the realm of intelligent information management in the case of information assigned to large groups of users or employers of institutions. As application of such methods allows sharing information in any institution, the subsequent step in our research will be an attempt to define the model structure or flow and assignment of constituent information to individual groups of interested and authorised persons. Such a model may later be implemented practically for sharing special multimedia information in the form of multiply digital signatures or multimedia watermarking.

Acknowledgments

This work has been supported by the AGH University of Science and Technology under Grant No. 10.10.120.783

References

1. Asmuth, C.A., Bloom, J.: A modular approach to key safeguarding. IEEE Transactions on Information Theory 29, 208–210 (1983)
2. Ateniese, G., Blundo, C., De Santis, A., Stinson, D.R.: Constructions and bounds for visual cryptography. In: Meyer auf der Heide, F., Monien, B. (eds.) ICALP 1996. LNCS, vol. 1099, pp. 416–428. Springer, Heidelberg (1996)
3. Beguin, P., Cresti, A.: General short computational secret sharing schemes. In: Guillou, L.C., Quisquater, J.-J. (eds.) EUROCRYPT 1995. LNCS, vol. 921, pp. 194–208. Springer, Heidelberg (1995)
4. Beimel, A., Chor, B.: Universally ideal secret sharing schemes. IEEE Transactions on Information Theory 40, 786–794 (1994)
5. Blakley, G.R.: Safeguarding Cryptographic Keys. In: Proceedings of the National Computer Conference, pp. 313–317 (1979)
6. Blakley, G.R.: One-time pads are key safeguarding schemes, not cryptosystems: fast key safeguarding schemes (threshold schemes) exist. In: Proceedings of the 1980 Symposium on Security and Privacy, pp. 108–113. IEEE Press, Los Alamitos (1980)
7. Blundo, C., De Santis, A.: Lower bounds for robust secret sharing schemes. Inform. Process. Lett. 63, 317–321 (1997)
8. Charnes, C., Pieprzyk, J.: Generalised cumulative arrays and their application to secret sharing schemes. Australian Computer Science Communications 17, 61–65 (1995)
9. Desmedt, Y., Frankel, Y.: Threshold Cryptosystems. In: Brassard, G. (ed.) CRYPTO 1989. LNCS, vol. 435, pp. 307–315. Springer, Heidelberg (1990)

10. Hang, N., Zhao, W.: Privacy-preserving data mining Systems. Computer 40, 52–58 (2007)
11. Jackson, W.-A., Martin, K.M., O'Keefe, C.M.: Ideal secret sharing schemes with multiple secrets. Journal of Cryptology 9, 233–250 (1996)
12. Ogiela, M.R., Ogiela, U.: Linguistic Cryptographic Threshold Schemes. International Journal of Future Generation Communication and Networking 1(2), 33–40 (2009)
13. Shamir, A.: How to Share a Secret. Communications of the ACM, 612–613 (1979)
14. Simmons, G.J.: An Introduction to Shared Secret and/or Shared Control Schemes and Their Application in Contemporary Cryptology. The Science of Information Integrity, pp. 441–497. IEEE Press, Los Alamitos (1992)
15. Tang, S.: Simple Secret Sharing and Threshold RSA Signature Schemes. Journal of Information and Computational Science 1, 259–262 (2004)
16. Wu, T.-C., He, W.-H.: A geometric approach for sharing secrets. Computers and Security 14, 135–146 (1995)

Surveillance Camera Tracking of Geo positioned Objects

Piotr Dalka, Andrzej Ciarkowski, Piotr Szczuko,
Grzegorz Szwoch, and Andrzej Czyżewski

Abstract A system for tracking moving objects with a known GPS position using a set of PTZ cameras with limited areas of coverage is presented. The basic idea of the system and its possible applications are discussed. The proposed camera calibration technique is used to transform the GPS position to camera settings. The current position of the tracked object is predicted in order to compensate the transmission and processing delays. The distributed client-server system using mobile terminals and its application to tracking objects movement, with additional functionality such as showing the position of the objects on the map, is presented. The results of the tests performed in real-life conditions, as well as perspectives of a future development of the system, are discussed.

1 Introduction

Modern video surveillance systems are comprised of multiple cameras, covering a wide range of the observed area, and a central monitoring system for realtime detection of events occurring in the monitored area and for recording of the camera images for the future use as an evidence. The task of constant monitoring of video streams from a large number of cameras and reviewing the recordings in order to find a specified event requires a considerable amount of time and effort from the system operators and it is prone to errors. A solution to this problem is an automatic system for constant analysis of camera images being able to raise an alarm if a predefined event is detected. Methods for analysis of video images from a single camera are now well established (Chen et al. 2005). Most of them process the camera images in several stages including detection and tracking of the moving objects, with classification of the objects using classes like humans, vehicles, objects,

Gdansk University of Technology, Multimedia Systems Department,
Narutowicza 11/12, 80 233 Gdansk, Poland
e-mail: indect@multimed.org

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 21–30.
springerlink.com

etc. (Czyzewski and Dalka 2008). Next, using a system of rules, a detection of predefined events is performed. These events may range from the simple ones (like a car entering the observed area) to complex events such as loitering, theft, fights, road collisions, etc. (Szwoch et al. 2008). However, in real video surveillance systems, multiple cameras are used for monitoring different areas that may be separated or may overlap. Video monitoring systems use stationary wideangle cameras (with constant filed of view) and pan-tilt-zoom (PTZ) cameras (where field of view may be moved and zoomed). The main problem here is that the results of video analysis obtained from multiple cameras have to be combined in order to achieve an accurate event detection. For example, if any object leaves a field of view of one camera and, after some time, it enters the field of view of another camera, the system has to recognize that it is the same object.

In the automatic multi-camera monitoring system there are two main problems that need to be solved. First, a proper configuration of the cameras has to be chosen. Wideangle cameras may be used for efficient object detection and tracking but they do not have zoom capabilities. On the other hand, PTZ cameras provide a detailed fullscreen view of the object, however a detection and tracking of other moving objects is not possible at the same time. Therefore, a solution to this problem can be a combined camera system. The wide-angle cameras are used for object detection and tracking, as well as for event detection. In case an event is detected, PTZ cameras are used to provide a detailed view of the selected moving object and to track its movement. Therefore, a second problem to be solved is related to positioning and zooming the view of PTZ cameras in order to track a particular object. Moreover, if the object leaves the field of view of a camera, another camera has to capture this object and should continue its tracking.

The problem of object tracking using multiple cameras is still a topic of research and development, with various approaches to this task. Cai and Aggarwal used statistical models to find the most likely matches of human subjects between consecutive frames taken by cameras mounted in various locations (Cai and Aggarwal 1996). Lee et al. tracked objects using multiple cameras by applying planar geometric constraints to the objects and by matching and fitting them to a planar model in order to align the scene's ground (Lee et al. 2000). Calderara et al. utilized the entry edges of field of view to solve the labeling problem between partially overlapped views, without any manual calibration of the cameras (Calderara et al. 2005). Zhou and Aggarwal used feature fusion to integrate spatial position, shape and color information to track objects in multiple cameras (Zhou and Aggarwal 2005). Numerous alternative methods may be found in the literature.

This paper presents a novel approach to the described problem, utilizing modern technologies. The tracked object is equipped with a GPS receiver and a mobile terminal running a client service that continuously transmits the position of the object to the server. The task of the server is to convert geographical coordinates of the object to parameters of PTZ cameras, to select the cameras that are able to track the object at the moment and to adjust the field of view of these cameras appropriately. In order to implement this functionality, two additional operations have to be performed: cameras have to be calibrated so that the conversion of coordinates is possible, and prediction of the object's movement is needed to compensate for the delay related to signal transmission and processing. Although the proposed method is not universal, as it requires that the tracked object uses the necessary equipment, it may be useful in some real-life scenarios such as a police car chasing an offending vehicle, a secret agent following the suspect, etc. The paper is organized as follows. Section 2 presents the algorithm for converting the position of the object to the camera parameters and applying the needed compensation. In Section 3, the architecture of the client-server system, data transmission, controlling the cameras and visualization of objects' positions are presented. Section 4 includes the results of testing of the described system in real life scenario and discussion of these results. The papers ends with conclusions and presentation of possible applications of the proposed system and possibilities of its future improvements.

2 Object tracking with PTZ cameras

Efficient object tracking with PTZ cameras requires two steps to be performed. First of all, an algorithm for conversion of object's GPS position to appropriate setting of a PTZ camera needs to be implemented. Furthermore, each camera must be calibrated in order to be able to lock on a moving object with enough high accuracy. And finally, a method for tracking the same object with multiple cameras, considering their limited areas of coverage, needs to be developed. Such a system, fulfilling the aforementioned assumptions and demands is presented in this section.

2.1 Conversion of object's GPS position to PTZ camera settings

This section describes the algorithm used to fix a PTZ camera on a nearby object based on its known GPS position. The aim of the algorithm is to calculate pan, tilt and zoom parameters for a dome camera to guarantee that an object of known longitude and latitude will be present in a video stream. Because of a very low accuracy of altitude data from a GPS receiver, it is assumed that all objects move on the same altitude equal to the ground level.
The algorithm (Fig. 1) has four settings regarding the camera: its longitude lon_C, latitude lat_C (both in degrees), height above the ground level h_C and pan offset p_{off}, which is defined as a clockwise distance (in degrees) between the north and a camera zero pan position. Input data for the algorithm include 4 parameters measured by a GPS receiver moving with an object being tracked: longitude lon and latitude lat of the object (in degrees), its speed v and bearing br (in degrees). In the first stage of the algorithm, local, Cartesian 3D coordinate system is formed with the camera in its center. Y axis of this system is headed North and x axis is headed East. Camera position is therefore denoted as $(0,0,h_C)$ and object position is described as $(x,y,0)$. Object's lat and lon coordinates are translated to x, y coordinates as follows:

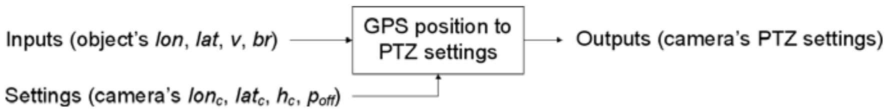


Fig. 1: Block scheme of object's GPS position to PTZ camera settings algorithm.

$$x = (lon - lon_c) \cdot D_{lon}(lat_c), \quad y = (lat - lat_c) \cdot D_{lat}(lat_c) \tag{1}$$

where D_{lon} and D_{lat} denote surface distance (in meters) per 1° change in longitude and latitude for a given latitude. Because of a relatively small distance between the camera and the object, it is assumed that D_{lon} and D_{lat} parameters for latitudes of the camera and the object are the same and are equal to the ones of the camera. They are calculated according to the WGS84 ellipsoid used by all GPS devices (Tyler and Blewitt 2006). D_{lat} value changes from 110574 m at 0° latitude to 111694 m at 90° latitude and is taken from a precalculated table. D_{lon} value for a given latitude is obtained from the following equation:

$$D_{lon}(lat) = \frac{\pi}{180} \cdot cos(lat) \cdot \sqrt{\frac{a^4 cos(lat)^2 + b^4 sin(lat)^2}{(a\,cos(lat))^2 + (b\,sin(lat))^2}} \tag{2}$$

where $a = 6378137$ m and $b = 6356752.3$ m are Earth's equatorial and polar radii, respectively. There is a significant delay in the system of tracking video objects with video cameras, caused by the GPS receiver, data transmission to the camera's node station (e.g. with GPRS protocol), computing time and executing PTZ command by the camera. This delay must be compensated in order to assure that a fast-moving object is always present in a video frame (preferably: in its center).
System delay compensation is performed by setting the PTZ camera to the predicted position of the object. Prediction time should be equal to the total delay in the system. In the current implementation, a linear predictor is used that estimates object position based on its speed v and bearing br reported by a GPS device:

$$\hat{x} = x + d \cdot v \cdot sin(br), \quad \hat{y} = y + d \cdot v \cdot cos(br) \tag{3}$$

where d is the delay of the system and was set to 2 seconds based on experimental results.
The pan parameter p for the camera is calculated as follows:

$$p = 90 - atan2(\hat{y}, \hat{x}) - p_{off} \tag{4}$$

and scaled to the range required by the PTZ camera.
The tilt parameter t is given with the equation:

$$t = -atan\left(\frac{h_c}{\sqrt{\hat{x}^2 + \hat{y}^2}}\right) \text{ if } \hat{x} \neq 0 \vee \hat{y} \neq 0, \quad t = -90 \text{ otherwise} \tag{5}$$

The last camera parameter setting, the zoom, is set based on the object's distance. The closer the object is to the camera, the smaller is the zoom value. This approach assures that object dimensions in video remain more or less constant.

2.2 System calibration

In order to achieve satisfactory accuracy of moving object tracking, 4 camera parameters (lon_C, lat_C, h_C, p_{off}) must be defined with a very high precision. Initial experiments proved that simple measurement of camera position with a GPS receiver does not provide the required precision. Also, determining camera's altitude above the ground might be problematic. Theoretically, the easiest parameter to obtain is the camera pan offset, however the precision required (especially when tracking distant objects) makes any direct measurement very difficult.
Therefore, a one-time optimization approach was chosen to estimate camera parameters with a sufficient precision. Initial approximations of four parameters are obtained with direct measurements. All parameters are then further tuned during two stage, non linear optimization process, which iteratively minimizes cost functions describing the difference between pan and tilt values calculated by the algorithm and the ones measured directly. Localizations of N points scattered equally in a camera vicinity were measured with a GPS receiver for this purpose,. In the same time, pan and tilt values were obtained from a camera pointing at each spot. The camera zoom does not influence pointing accuracy, thus this parameter is omitted from the optimization process.
The number of calibration points should not be too small because reference data gathered with GPS receiver are often inaccurate and their precision is unknown. During experiments, 10 calibration points for each camera were used.
In the first stage, 3 parameters responsible for object tracking in the horizontal plane (lon_C, lat_C, p_{off}) are estimated. The cost function for this stage is defined as: follows:

$$E_1(lon_C, lat_C, p_{off}) = \sum_{i=1}^{N} w_i \cdot (p_i^c - p_i^r)^2 \tag{6}$$

where p_i^c is a pan value calculated according to Eq. (4), p_i^r is a real pan value obtained from the camera for the i^{th} calibration point and w_i is the weight of the calibration point. The weight of a point is directly proportional to its distance from the camera. This assures that the farther the point is from the camera, the greater is its influence on the results of optimization.
During the optimization procedure, utilizing conjugate gradient method (Avriel 2003), lon_C, lat_C and p_{off} values are tuned up in order to minimize the value of E_1 function. At the second stage of calibration process, the h_C value responsible for object tracking in the vertical plane is tuned, based on calibration data and on 3 parameters found during the first stage. The cost function for this stage is given as:

$$E_2(h_C) = \sum_{i=1}^{N} w_i \cdot (t_i^c - t_i^r)^2 \tag{7}$$

where t_i^c is a tilt value calculated according to the Eq. (5) and t_i^r is a real tilt value obtained from the camera for the i^{th} calibration point. The optimization procedure allows to alter h_C value in order to minimize value of E_2 function.

2.3 Multi-camera tracking

The system supports tracking of an object by many cameras at the same time, provided that the object may be "seen" by a camera. Area of coverage (AOC) for each camera is defined by an operator. This area denotes regions in the camera vicinity that can be observed by a camera and are not obscured, e.g. by buildings. AOC for each camera is defined as a polygon (or set of polygons) with vertices described as geographical coordinates. This facilitates creating AOC for each camera and allows an operator to use any common mapping software. Whenever an object being tracked enters AOC of any camera i.e. its GPS position is contained inside any of polygons forming AOC of the camera, the camera is automatically pointed at the moving object and tracks the object as long, as it stays in its AOC.

3 Surveillance system architecture

3.1 Transport and service layers

This section describes the architecture and operation of a distributed surveillance system that is used as a platform for camera tracking services. The presented solution was designed as an open and extensible system for hosting surveillance components and camera-tracking application. The system is based on the foundation of Extensible Messaging and Presence Protocol (Saint-Andre 2004 RFC3920, RFC3921), which was primarily designed for creating instant messaging applications, however its deployment practice proves its usefulness also as an excellent tool for creating service-oriented architectures. This is due to exceptional extensibility of the protocol stack and a number of already-developed extensions, which are standardized within well-defined process. It is worth emphasizing that the use of open and standardized framework like XMPP makes it possible to exploit its numerous features including, but not limited to transport-layer security, authentication, addressing, service discovery without any additional cost related to the design and implementation of them inhouse. Moreover, large base of XMPP deployments guarantees a reliability of employed security mechanisms, which are thoroughly reviewed by a number of experienced developers, and which are crucial to ensure confidentiality of data transmitted in surveillance systems. Other important aspect of using XMPP is the availability of server software, often with a very advanced functionality substantially lowering initial development costs.

The described solution uses XMPP primarily as an abstraction of "bus" which allows for transmitting of arbitrary kind of data, however the design of the protocol makes it suboptimal with regard to transfer of real-time multimedia data. Therefore, in order to optimize the architecture towards a better multimedia streaming, being a substantial functionality for a surveillance system, the design has been augmented with the use of one of XMPP extensions - the Jingle protocol (XSF 2009 Jingle). Its main area of usage is the implementation of Voice and Video over IP applications (VoIP), however it is possible to build arbitrarily complex media streaming architectures based on it, which is efficiently exploited in presented system. Jingle is essentially a session control (signaling) protocol, however in this paper this term will be referred to a signaling coupled to the controlled media transport, which is based on SRTP (secure real-time transport protocol).

The described application constitutes a service layer built on top of XMPP deployment. The agents which form the core of surveillance functionality are in fact specialized XMPP clients which connect to XMPP server in order to exploit its aforementioned features. Therefore, from the perspective of the presented system, agents may play different roles: some of them host the higher-level services (they serve some surveillance functionality), the others are the clients to these services. This distinction is in fact quite flexible, as the same agent may act as a server to some services and client to the others. This situation is even reflected on the implementation level, where the "agent" executable is solely a container providing communication facilities to the hosted services "plug-ins", it may be a camera control module, GPS-related service or any other. Summarizing, XMPP protocol and its extensions allow to create a multi-agent system capable of transfer of arbitrary data, together with advanced media streaming features.

3.2 Agents

The presented system utilizes several configurations of agents, which are described below.

- **Node station** is an agent which runs on an embedded computer system with attached numerous kinds of sensors. In the aspect of described system, these sensors are primarily the cameras, however the same systems have been used for other applications, like microphone arrays, noise measurement microphones, weather stations etc. The hardware configuration determines which agent "plug-ins" are enabled. The node station typically is installed in some remote location and uses wireless communication (WiFi, GPRS, 3G, etc.) to connect to the system. Node station also includes services for remote management and remote software update.

- **Surveillance server** is an agent, which hosts services used for gathering and processing of data from node stations and other terminals. In the presented scenario, the most important service is the “live map” bridge, which translates GPS and presence data obtained from mobile terminals into commands for live map application. Other than that, the surveillance server typically includes plug-ins whose task is to process (e.g. log, combine into higher-level description) security events from other agents. The system architecture allows for using multiple surveillance servers on one system, which makes it easy to replicate crucial algorithms and to further distribute the calculation-intensive processes.
- **Mobile termina**l is a special agent which runs on a mobile wireless device. In the described scenario, handheld computers running Windows Mobile OS have been used. The mobile terminal agent presents its user a GUI, which, apart from access to surveillance-related functionality, allows to use standard XMPP features like instant messaging and presence management. These features, together with GPS position reporting, make it a particularly well-suited tool for building more complex security systems with it. It is worth emphasizing that the GPS reporting component utilizes another XMPP extension (XSF 2009 User Location), which makes it possible to fully implement this functionality within XMPP standards.

Apart from the agents described above, the system includes also the “live map” application which allows positioning the system terminals on a map. During the experiments, a simple Adobe Flash application was used for this purpose, which provided real-time updates of terminal localization and presence information. In the future development this application will be replaced by an XMPP agent, which will implement UI allowing for interaction with other agents.

4 Test Results

A test system has been set up, containing two cameras, one node station agent for each camera, a surveillance server agent for system management and a PDA - mobile terminal with GPS and HSDPA modem. The PDA reads its position from a GPS receiver and transfers it to the server. The server feeds GPS data to the appropriate node stations according to the areas of coverage of their cameras, where it is translated to PTZ settings of the cameras. The user interface of the system presents live video feeds with additional GPS information displayed, and “live map” for terminal positioning (Fig. 2a). Three test drives were performed, two involving a typical civil car and one involving a police car. A driver was equipped with the tracked mobile terminal. The experimental setup, including camera locations and route plan is presented in Fig. 2b. It adopts turning left at the crossroad 1, then left at the crossroad 2, next U-turn at the parking marked as “P”, turning right on crossroads 4 and 5. Maximal speed achieved during test drives was 50km/h (between crossroads 1-2 and 4-5 on a 250 m section). The drives were observed by two cameras positioned on a roof of the building, marked as circles with “c” inside. Areas of coverage of these two cameras are marked by dashed lines. Video streams from two cameras were

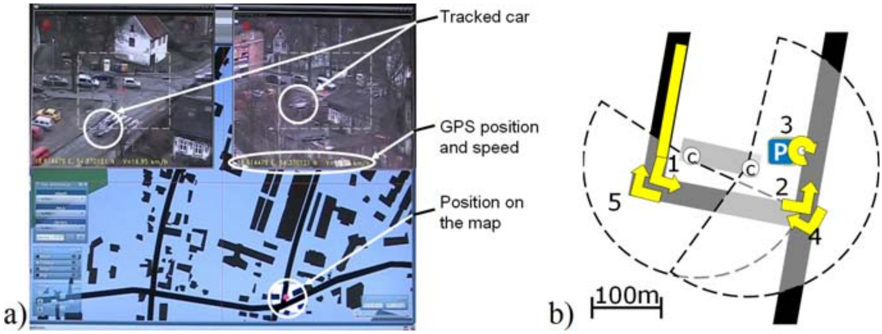


Fig. 2: a) User interface of the system. Live video feeds and a map are provided, b) Test drive route plan, AOCs of cameras are marked as dashed lines.

recorded during the test drives, together with the GPS data, for further analyzing. For over 90% of time of every test drive a position was calculated correctly, even for high speed driving, therefore the cameras were tracking the object effectively, i.e. object was kept in the proper position in the video frame, defined as a centered rectangle of 25% of the frame area. Minor errors occurred for about 10% of time, originating in GPS imprecision, however the object was still visible on the screen because for the zoom ratio used in the experiment, the screen width corresponded with at least 15m on the road.

5 Conclusions

The system developed proved to be efficient and suitable for video tracking of objects with known position. Its open architecture is prepared for extension with additional cameras, covering practically any area. In the future, localization data can be provided by other means, not only by GPS, for example RFID systems with a capability of locating tags passing near transceivers can be used. Currently, a new version of an application for tracking of moving objects is developed, based on the image analysis, only. In this case, fixed cameras with calibrated fields of view, adopting a correlation of the pixel location on the screen to a geographical position, are utilized. Any moving object might be detected with the background subtraction methods, whereas its position on the screen is translated to the geographical position, utilized to move PTZ cameras.

Acknowledgements

Research is subsidized by the Polish Ministry of Science and Higher Education within Grant No. R00 O0005/3 and by the European Commission within FP7 project "INDECT" (Grant Agreement No. 218086).

References

Chen et al. 2005. Chen, T.P., Haussecker, H., et al.: Computer Vision Workload Analysis: Case Study of Video Surveillance Systems. Intel Technology Journal 9, 109–118 (2005)

Czyzewski and Dalka 2008. Czyzewski, A., Dalka, P.: Moving Object Detection and Tracking for the Purpose of Multi-modal Surveillance System in Urban Areas. In: New Directions in Intell. Interactive Multim. (1st Int. Symp. on Intell. Interactive Multim. Syst. and Services), pp. 75–84. Springer, Heidelberg (2008)

Szwoch et al. 2008. Szwoch, G., Dalka, P., Czyzewski, A.: Objects classification based on their physical sizes for detection of events in camera images. In: Proc. NTAV/SPA 2008 Signal Processing and New Trends in Audio/Video, Poznan, Poland, pp. 15–20 (2008)

Cai and Aggarwal 1996. Cai, Q., Aggarwal, J.K.: Tracking human motion using multiple cameras. In: Proc. 13th Int. Conf. on Pattern Recognition, vol. 3, pp. 25–29 (1996)

Lee et al. 2000. Lee, L., Romano, R., Stein, G.: Monitoring activities from multiple video streams: establishing a common coordinate frame. Proc. IEEE Trans. Patt. Anal. and Mach. Intell. 22, 758–767 (2000)

Calderara et al. 2005. Calderara, S., Vezzani, R., Prati, A., Cucchiara, R.: Entry edge of field of view for multi-camera tracking in distributed video surveillance. In: Proc. IEEE Conf. Advanced Video and Signal Based Surveillance, pp. 93–98 (2005)

Zhou and Aggarwal 2005. Zhou, Q., Aggarwal, J.K.: Object tracking in an outdoor environment using fusion of features and cameras. Image and Vision Computing 24, 1244–1255 (2006)

Tyler and Blewitt 2006. Tyler, G., Blewitt, G.: Intelligent Positioning: GIS-GPS Unification. Wiley, Chichester (2006)

Avriel 2003. Avriel, M.: Nonlinear Programming: Analysis and Methods. Dover Publishing, New York (2003)

Saint-Andre 2004 RFC3920. Saint-Andre, P., et al.: RFC 3920: Extensible Messaging and Presence Protocol (XMPP): Core (2004), `http://www.ietf.org/rfc/rfc3920.txt` (accessed January 12, 2009)

Saint-Andre 2004 RFC3921. Saint-Andre, P., et al.: RFC 3921: Extensible Messaging and Presence Protocol (XMPP): Instant Messaging and Presence (2004), `http://www.ietf.org/rfc/rfc3921.txt` (accessed January 12, 2009)

XSF 2009 Jingle. XMPP Standards Foundation, XEP-0166: Jingle (2009), `http://xmpp.org/extensions/xep-0166.html` (accessed January 12, 2009)

XSF 2009 User Location. XMPP Standards Foundation, XEP-0080: User Location (2009), `http://xmpp.org/extensions/xep-0080.html` (accessed January 12, 2009)

Distributed knowledge creation, recording and improvement in collaborative design

Piero Mussio, Stefano Valtolina, Barbara Rita Barricelli, Monica Bordegoni, Francesco Ferrise, and Marco Ambrogio

Abstract The design of complex industrial products requires more knowledge than the one that a single person can possess. A successful industrial design entails a collaborative activity between designers and engineers during which they share their specialized knowledge accepting and interpreting their points of view through a multi-disciplinary and collaborative methodology. In particular, this paper focuses on issues related to the knowledge creation, management and enrichment for industrial collaborative design in virtual environment. The idea is to endow the virtual environment, used to create the 3D models and to carry out the analysis, with a knowledge management (KM) module able to gather information about different versions of the product and simulation analysis data. Moreover, this KM module includes an annotation tool used to support annotation-based collaboration between designers and engineers through which they can exchange impressions, ideas, and comments in order to achieve a shared solution about an industrial design issue.

1 Introduction

The diffusion of information technology favors new forms of collaborative design in the industrial field which require that experts in different disciplines share their specialized knowledge, skills and work across different geographic and time zones. In these situations one of the major problems to face and acknowledge is that different stakeholders belonging to different technical and scientific communities have different views on reality, different ways of reasoning, learning and teaching [1]

Piero Mussio · Stefano Valtolina · Barbara Rita Barricelli
Department of Computer Science and Communication, University of Milano, via Comelico 39, Milano, 20135, Italy, e-mail: {mussio,valtolin,barricelli}@dico.unimi.it

Monica Bordegoni · Francesco Ferrise · Marco Ambrogio
Dipartimento di Meccanica, Politecnico di Milano, via La Masa 1, 20156, Milano, Italy, e-mail: {monica.bordegoni,francesco.ferrise,marco.ambrogiomail}@polimi.it

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 31–41.
springerlink.com

and therefore work, read documents and communicate following different ways of reasoning [2]. Each stakeholder performs his/her activity, reads, creates documents, takes his/her decisions and works using implicit information, articulatory skills and tacit knowledge which rise from his/her experience. Actually, these factors are among those ones determining the so-called invisible work [3]. Implicit information - e.g. information embedded in grouping, spatial displacement physical appearance of the text and graphical elements in a document - is often significant only to users who posses the knowledge to interpret it. Most of this knowledge is not made explicit and codified but is tacit knowledge, namely it is knowledge that users possess and currently use to carry out tasks and to solve problems, but that they are unable to express in verbal terms and that they may even be unaware of. It is a common experience that stakeholders exploit mainly their tacit knowledge, since they are often more able to do rather than to explain what they do [4]. Communicational and reasoning gaps arise among stakeholders in the team due to their different cultural backgrounds.

This paper reports the results of a design experience developed in the project PUODARSI, whose aim is to create a prototype of co-laboratory [5] for the design and analysis of industrial products. The co-laboratory supports the cooperative activities of structural analysts, fluid-dynamics analysts and style designers, who can access the Virtual Reality (VR), Augmented Reality (AR) and Simulation tools of the PUODARSI co-laboratory. In order to overcome the communicational and reasoning gaps, each stakeholder can represent the results of his/her work as a 3D model enriched by multimedia annotations (see figure 2). In this way each stakeholder externalizes the “why” and “how” of her/his activity, embedding implicit and explicit information in the model and his explanations in the annotations. The 3D model becomes a boundary object [6] which all the other stakeholders can see, touch and interpret according to their culture, thus complementing the multimedia explanations in the annotations. In turn, each stakeholder can answer to a proposal by updating old or producing new 3D models externalizing his/her ideas and/or updating old or adding new annotations. The system supporting the co-laboratory organizes the different versions of 3D models and annotations organizing them in a knowledge base, that each stakeholder can navigate and through that she/he can reconstruct the story of the design, explore alternative designs and also re-use results from earlier experiences. Thus the knowledge produced in the design activity is captured, organized and made available to the different communities of stakeholders.

The rest of the paper is organized as follows: the section 2 presents some background information and works related to the communication issues between different actors involved in an industrial design process. The section 3 describes the knowledge model at the base of the PUODARSI system underling the importance of the annotation tool for triggering the collaboration between style designers and mechanical engineers in order to bring to a successful conclusion of the industrial design process. Then in the section 4 more technical details about the PUODARSI system are presented and finally the section 5 outlines some conclusions.

1.1 A scenario

Carlo is a style designer working in the Style Lab of an automotive company, collaborating to the design of a front fairing of a motorbike with experts in Fluid Dynamics, working in the remote Computational Fluid Dynamic (CFD) Lab and mechanical engineers expert in mechanical structural studies of Finite Element Method (FEM) Lab of an allied company. The members of the team collaborate asynchronously by accessing the simulation programs, VR and annotation tools of the PUODARSI co-laboratory, which allow them to create, visualize, analyze, annotate and modify different versions of the front fairing until they reach a fully accepted model of it. Figure 1, shows Carlo in the first step of the industrial design process, diving in an immersive environment, to create a first 3D model of the front fairing exploiting VR tools and haptic devices, the Sensable Phantom tool. He designs the front fairing mainly from the aesthetic point of view.

Fig. 1: VR environment in which a style designer creates or modifies the front fairing.

He then makes the model available to the other stakeholders, using the PUODARSI co-lab facilities. The model on the screen, being a 3D model, can be rotated and observed by each stakeholder from different points of view, and can be explored through simulation or feature extraction programs as well as modified. The experts in Fluid Dynamics, working in CFD Lab, analyze the front fairing by simulating its behavior at different speeds. Mechanical engineers expert in mechanical structural studies simulate its behaviour under different conditions of stress in the FEM lab. Then, exchanging annotations, they can propose some changes to make in the model of the front fairing.

Carlo uses the PUODARSI co-lab facilities to access the model to modify it through the use of the phantom device according to advices provided by other members of the team. For example, he can dislike the changes proposed by the engineer, because of some un-aesthetic reflections of the light on the surface. He can now propose a modification by changing the model and explains in his language the why, by creating the following annotation: “in this point the curve of the profile is emphasized to make it more pleasant and because in this way the light runs better”. (figure 2).

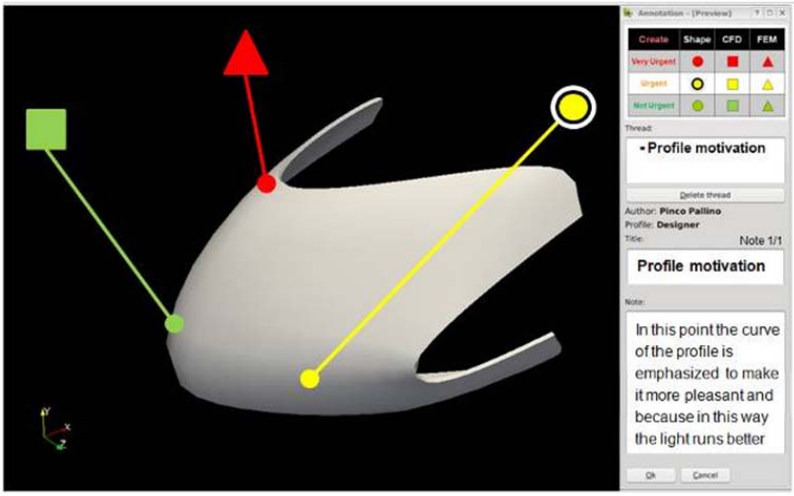


Fig. 2: The screen in the VR environment in which the style designer sees the model proposed by collaborators, modifies it and creates an annotation in order to describe the reasons of his modification The visual link associated to the displayed annotation has a thick border.

An annotation is formed by two components: a visual link and a thread: The visual link is a graphical element, pointing to the voxel (3d pixel) being annotated. Its shape resumes its meaning according to the rules shown in the palette on the upper right corner of the image. The thread contains multimedia texts (called notes) related to the point in discussion together with some data characterizing each notation. In figure 2 three different visual links are present. On the right of the image the palette and the thread appear. The thread displays the note written by Carlo to explain his point.

The Carlo's note is written in the style designer language, and may seem vague to the CFD expert. He can use the tools of CFD lab to explore the 3D model and to evaluate the Carlo's annotation and work. For example, a CFD analysis can demonstrate that the 3D model has low air penetration features. For this reason the CFD engineer responds to Carlo's annotation adding his own observation to the annotation thread: "in this point the flow creates some whirlwind because the second derivative of the surface changes abruptly". Being aware of the communicational gap, he adds to the shape model the fluid models, which represents the air traces in the space around the front fairing as green lines (figure 3).

Looking at the screen and rotating the image to see the air behavior from different points of observation, Carlo can understand the why the CFD expert is unhappy and going on in the bargain to reach an aesthetically satisfactory and yet efficient solution.

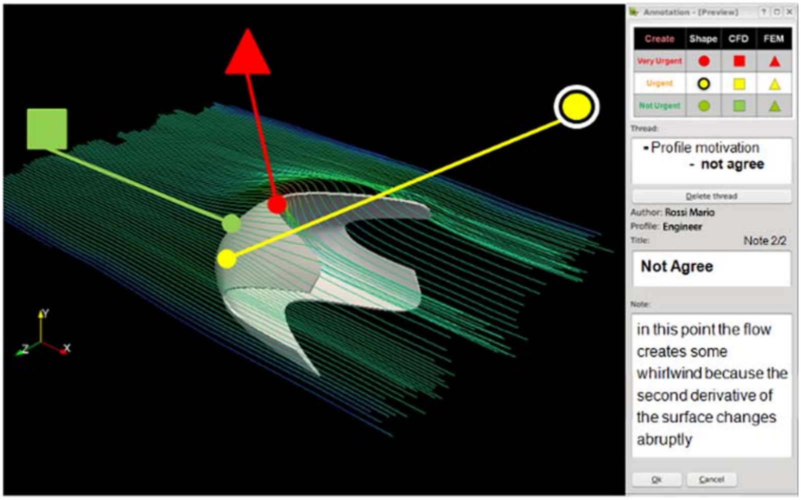


Fig. 3: Reasoning on the same model a CFD engineer, after a CFD analysis, replies to the annotation written by the style designer.

The PUODARSI co-lab endows the VR environments, in which the different stakeholders operate, with a Knowledge Management module (KM), which gathers and organizes the knowledge accumulated during the design process - different versions of the model, simulation analysis data under each shape's hypothesis and messages, comments and ideas exchanged between style designers and engineers during the whole industrial process. Each stakeholder can reason on the current situation, taking into account the whole history so far. Thanks to these features, the different stakeholders reach a final solution to their design problem, making faster the whole industrial design process and documenting the history of the modeling and engineering analysis which brings to the definition of the final product. All this knowledge is made available for future reuse.

2 Backgrounds and related

Several research activities are performed in the context of the collaborative design. These researches do not consider that the design of complex artefacts such as a front fairing requires more knowledge than the one that a single person can possess [7]. For example [8] presents a 3D collaborative virtual environment in which different users have the opportunity to exchange data to achieve a common solution about the design of an industrial product.
However, the scenario presented in the previous section shows as the design of aesthetic and functional artifacts requires the combination of different competences and

knowledge related to the work of experts in style design and CFD/FEM analysis that operate in specialized labs, and cooperate through the net. To achieve these results it is necessary to underline as during the industrial design process different experts can acquire knowledge from each other, accepting and interpreting their points of view through a multi-disciplinary and collaborative methodologies [1,2]. Therefore, different actors belonging to different cultural spheres and thus having different points of view and competences, sometimes in contrast, cooperate to create a common awareness in order to achieve a successful design of an industrial product [7].
In the scenario presented in the first section, each stakeholder (designer or engineer) can examine comments placed in the virtual environment by the other members of the team using an annotation tool. The use of the annotation tool as communicational medium for exchanging ideas and impressions is used in different situations [9].The concept of annotation has been defined by different authors [9,10,11]. An annotation is a note, added by way of comment or explanation [10] to a document or to a part of a document. The entity being annotated is called the base of the annotation. The base of the annotation is often made evident by a visual identifier. In some case, the human annotating a base makes explicit the link between annotation and base using a visual link.
The idea at the base of this paper is to exploit an annotation tool to allow different experts (designers and engineers) to explain implicit knowledge associated to the models that is the motivations that have brought to the definition of a 3D geometry or at the base of the analysis performed by the engineers.

3 Knowledge Model

As said before, PUODARSI collaborative environment is a virtual environment in which designers and engineers work in a collaborative way during the various stages of the industrial design process [12]. Within the framework of this collaboration it is necessary to design a KM system able to manage:

- information regarding the history connected to the definition of the shape. The shape is first defined by the style designer that haptically deforms a surface until she/he reaches the desired shape. Then as in a closed loop the shape is analyzed by the engineers and, if the results from the analyses are bad, some changes are required. This process ends when the shape has been optimized from both the aesthetic and the functional points of view. The KM system has to be able to record completely this iterative process and makes it possible to replay it again at the end of it all. Moreover, the KM system has to be able to record information about the numerical simulations at different steps and all these data are useful for the analysts in order to understand, once the process of the definition of the shape is concluded, the relations among different shapes and different behaviors. In other words, the KM system has to store 3D models, CFD analysis and FEM analysis in order to trace the history of the project and its evolution

- information needful to formalize the intentions of designers or engineers in order to represent what they wish to associate to the product. In other words, information related to the annotations supporting the collaborative activities between designers and engineers that externalize their reasoning using the annotation tool

The KM module adopted in the PUODARSI project is based on the use of an annotation tool able to trigger a collaboration between style designers and mechanical engineers in order to bring to a successful conclusion of the industrial design process.

3.1 Annotation tool

In the PUODARSI system the annotation is thought of as author-attributable content placed within a 3D scene in association with (or reference to) a particular element of that scene. Using scene annotation capabilities, the user is able to add commentary in association with specific aspect (or aspects) of the artifact. This new content persists in the environment and is made available to other users as a form of annotation attributed to the author. Moreover, this new content is also searchable, so that other users could locate all commentary on a specific artifact or commentary created by a specific author. A PUODARSI annotation is always presented in association to the 3D model that it refers to and to its visual link. It is always put in context with regard to annotated entity.
A visual link is formed by a stalk and a shape (see figure 2). The stalk in turn is a digital line connecting the base (the voxel being annotated) to the shape. The annotation is organized as a thread recording a set of multimedia notes produced by different authors. An annotation is accessed by selecting its visual link. The selected visual link shape is highlighted by making its border thick, so that the user can associate it to the open annotation.
The geometrical form and color of vl shapes are used to convey knowledge from the annotation's author to the other stakeholders. The members of the design team defined a simple code, to describe the importance of the note (how urgent it is to respond to it) and its argument (does it regards shape, CFD/FEM behavior?) (see Table 1).

In this way the external features of a vl are used as vehicle for knowledge transfer from the annotation's author to the other stakeholders. According to these rules, a user can understand at a first glance on the vl the kind of theme discussed in a thread associated and its urgency. The vl is a visual identifier associated to an annotation which body is a text, audio or video added by the author in order to express a comment or explanation. This body can be articulated as a set of notes composing a thread of discussion. Through a thread, like one presented in figure 3, different users can trigger a discussion process around the ideas exposed by the primary annotation's author. In this way, the annotation becomes a communication medium using which different users can exchange impressions, ideas and comments, that is the knowledge needful to achieve a shared solution about an industrial design issue.

Table 1: Code of interpretation of the vl shapes and colours.

	Shape	CFD	FEM
Very Urgent	●	■	▲
Urgent	●	■	▲
Not Urgent	●	■	▲

4 Architecture of the Virtual Environment

The figure 4 presents the final architecture developed in the context of the PUODARSI project. The final system is composed by mainly three modules: a VTK component, an annotation QT component and Km module.

The VTK component is used to visualize 3D models and CFD/FEM analysis. Exploiting this component a designer can work on the aesthetic aspect of the 3D model through the use of a haptic device. When the modelling phase is finished, the system saves an X3D file of the model named verN.x3d, where “N” stands for the number of times the file has saved. The VTK component is based on the VTK visual rendering library including a VTKrenderer function that has been specifically implemented for rendering 3D VTK objects (such as the output of the CFD analysis and annotations) and X3D file of the model. The visualization output is shown through the active stereoscopic window of the SenseGraphics system. The module for CFD analysis is implemented using the OpenFOAM library. The GMSH software library has been used to import the geometry and generate the mesh. Once the analysis is complete, the results are sent to the VTK component that renders these data. By importing the dataset as a VTK unstructured grid it is possible to choose different modalities for representing a fluid flow field, such as stream tubes, stream ribbons, glyphs and so on, according to the preferences of the analyst that is performing the CFD analysis [13]. Regarding the FEM analysis the results are generally represented with a deformed model that takes into accounts the displacements of the nodes of a discrete model, or with a map of colours showing the distribution of tensions. The representation can so be superimposed on the model and as a result of this the geometry becomes coloured. Therefore both the models, the CFD and FEM analysis are rendered in a VTK environment. Instead the annotation tool is not integrated in VTK (except the visual links) but in a QT component placed beside the visualization module (see figure 4).

In our case the annotation component has been developed in C++ using QT libraries. The idea is to create a VTK widget able to visualize only the visual links related to the annotations connected to the models, or the CFD/FEM analysis. Instead the annotation component is implemented by means of a QT dialog box using which the user can create or manage annotations. In this way it is possibility to exploit the

QT library for defining the text editing module and for integrating components to manage the threads of notes. Moreover QT supports easy integration of different multimedia files like audio, video or pictures related to the annotation.
Both the VTK component and QT component access archives storing 3D models, analysis and annotations through a KM module developed to guarantee a transparent information retrieval with respect to technical details of the databases. This KM module is composed by a set of classes able to recover information without writing any line of SQL code or without taking care of the database connections. This solution it is usable to easy retrieve explicit information about 3D models, analysis or annotations stored in different databases (DBs).

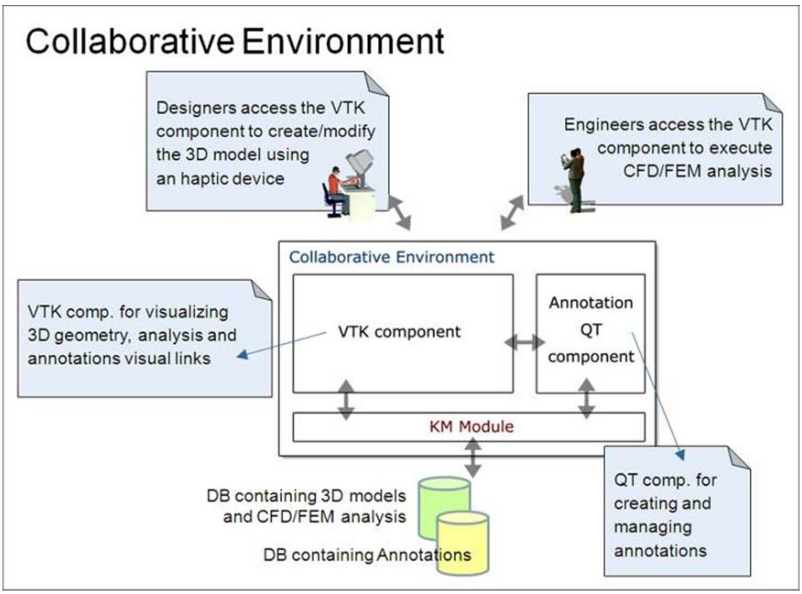


Fig. 4: The PUODARSI architecture.

The collaborative environment presented in figure 4 embeds the implicit knowledge reflected into the rules applied to materialize the 3D models and the annotations on the screen. These rules are defined experimentally according to usability principles and then checked and improved through usability analysis.

5 Conclusions

The paper presented a co-laboratory supporting the distributed creation, recording and improvement of knowledge in collaborative design of industrial products, through the cooperative activity of structural analysts, fluid-dynamics analysts and style designers. This laboratory makes use of VR/AR technologies in order to support the designer during the shape definition phase, and then to represent the results of engineering analysis on the products in the analysis phase. A KM module based on annotations is used to gather information about different versions of the product, simulation analysis data under each shape's hypothesis, comments and ideas exchanged between designers and engineers during the whole industrial process. These data are used to exchange information between the two different worlds, and to reduce the existing conceptual gap between them. For examples, engineers through the use of numerical simulations can explain the results of their analysis in a very precise and formal way but these numerical simulations are uncompressible for designers. Instead a solution less precise but more efficient is a communication based on the exchange of 3D models and annotations that become the boundary objects around which designers and engineers can achieve a common understanding about an industrial design issue. The example presented in the scenario has proved the feasibility of this solution, and also highlighted some issues to deal with for the future developments. These issues regard to the exchange of data between different users involved during the industrial design process and some problems about the visual representation of the annotations. The proposed solution is based on the use of an annotation tool designed to support the communication process between designers and engineers. The annotations and the 3D models are stored in different databases that which, in the next steps of the PUODARSI project, will be enriched with some algorithms for an intelligent organization and reuse of them that will make the co-laboratory easier to manage and use for both designers and engineers.

Acknowledgments

Current researches related to the PUODARSI (Product User Oriented Development based on Augmented Reality and interactive SImulation) system, are partially funded by the Italian Ministry of University and Research. Special thanks are due to Vincenzo Pupillo for his contribution in implementing the annotation QT component and the data access interface.

References

1. Snow, C.P.: The Two Cultures. C. P. Cambridge University Press, New York (1993)
2. Valtolina, S., Mussio, P., Mazzoleni, P., Franzoni, S., Bagnasco, G., Geroli, M., Ridi, C.: Media for Knowledge Creation and Dissemination. In: Semantic Model and Narrations for a New Accessibility to Cultural Heritage, 6th Creativity & Cognition Conference, Washington DC, USA, June 13-15 (2007)
3. Nardi, B., Engeström, Y.A. (guest eds.): Web on the Wind: The Structure of Invisible Work. Computer-Supported Cooperative Work 8(1-2), 1–8 (1999)
4. Schŏn, D.: Designing as reflective conversation with the materials of a design situation. Knowl.-Based Syst. J. 5(1), 3–14 (1992)

5. Kouzes, R.T., Myers, J.D., Wulf, W.A.: Co-laboratories: Doing Science on the Internet. In: Computer IEEE, pp. 40–46 (August 1996)
6. Arias, E., Fischer, G.: Boundary Objects: Their Role in Articulating the Task at Hand and Making Information relevant to It. In: Proc. Int. ICSC Symposium on Interactive and Collaborative Computing (ICC 2000) (December 2000)
7. Fischer, G.: Symmetry of ignorance, social creativity, and meta-design. Knowledge-Based Systems 13(7-8), 527–537 (2000)
8. Aubry, S., Thouvenin, I., Lenne, D., Olive, J.: A knowledge model to read 3D annotations on a virtual mock-up for collaborative design. In: Shen, W., Yong, J., Yang, Y., Barthès, J.-P.A., Luo, J. (eds.) CSCWD 2007. LNCS, vol. 5236, pp. 669–674. Springer, Heidelberg (2008)
9. Marshall, C.C., Bernheim Brush, A.J.: Exploring the Relationship between Personal and Public Annotations. In: Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2004) (2004)
10. Marcante, A., Mussio, P.: Electronic Interactive Documents and Knowledge Enhancing: a Semiotic Approach, the Document Academy, UC Berkeley, Berkeley, CA, USA, October 13-15 (2006)
11. Costabile, M.F., Fogli, D., Marcante, A., Mussio, P., Parasiliti Provenza, L., Piccinno, A.: Designing Customized and Tailorable Visual Interactive Systems. International Journal of Software Engineering and Knowledge Engineering (IJSEKE), 18(3), 305–325 (2008)
12. Bordegoni, M., Ferrise, F., Ambrogio, M., Bruno, F., Caruso, F.: A Multi-layered Modelling Architecture for Virtual Design. In: Proceedings of IDMME - Virtual Concept 2008, Beijing, China, October 8-10 (2008)
13. Ambrogio, M., Bruno, F., Caruso, F., Muzzupappa, M., Bordegoni, M.: Interactive cfd simulation in virtual reality. In: XIX ADM Ingegraf, June 4-6 (2008)

Effectiveness of Business Alliances Partnerships

Sylvia Encheva and Sharil Tumin

Abstract This paper focuses on many valued logic applied for filtering out the most profitable collaboration alternatives. The final recommendation is based on methods from the Analytic Hierarchy Process.

1 Introduction

The majority of small and medium size enterprises feel challenged by the ever changing market demands that should be satisfied with limited resources and means. A large number of such companies considerably increase their revenues by forming alliances, where partnerships vary from an informal business relationship to a joint venture agreement.

One of the many open questions related to alliances is how to speed up the process of selecting the most suitable alternatives. In an attempt to facilitate this process we propose use of a system based on many valued logic and the Analytic Hierarchy Process (AHP), [19].

Many valued logic is applied for filtering out the most profitable collaboration alternatives. The final recommendation employs Analytic Hierarchy Process (AHP). The AHP combines judgments and measurements while obtaining ratio scale priorities considering various the factors and groups of factors affecting a decision. It is a comprehensive framework that is designed to cope with the intuitive, the rational, and the irrational when making multi-objective, multi-criteria and multi-actor decisions [18].

Sylvia Encheva
Stord/Haugesund University College, Bjørnsonsg. 45, 5528 Haugesund, Norway
e-mail: sbe@hsh.no

Sharil Tumin
University of Bergen, IT-Dept., P. O. Box 7800, 5020 Bergen, Norway
e-mail: edpst@it.uib.no

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 43–52.
springerlink.com

The rest of the paper is organized as follows. Related work and supporting theory may be found in Section 2. Collaboration options are discussed in Section 3. The AHP is proposed to facilitate a choice among projects that seem to be equally profitable, Section 4. A prototype system architecture is described in Section 5. The paper ends with a conclusion in Section 6.

2 Related Work

Market considerations imply that early entry into large, growing markets is more likely to lead to success [21].

Business decisions on projects with potential positive rate of return are discussed in [4] and the human side of a decision making process is considered in [13]. Theory and application of cost-benefit analysis are presented in [14] and [15].

An object oriented model of a simulation based decision support for supply chain logistics is presented in [8]. A decision support system for administration of shipping enterprises is presented in [11]. Both cases involve statistical methods while our approach is based on classification methods from formal concept analysis and lattice theory.

Let L be a non-empty ordered set. If $sup\{x,y\}$ and $inf\{x,y\}$ exist for all $x,y \in L$, then L is called a *lattice*, [9], [10].

The semantic characterization of a four-valued logic for expressing practical deductive processes is presented in [1] and [2]. In most information systems the management of databases is not considered to include neither explicit nor hidden inconsistencies. In real life situation information often come from different contradicting sources. Thus different sources can provide inconsistent data while deductive reasoning may result in hidden inconsistencies. The idea in Belnap's approach is to develop a logic that is not that dependable of inconsistencies. The Belnap's logic has four truth values 'T, F, Both, None'. The meaning of these values can be described as follows:

- an atomic sentence is stated to be true only (T),
- an atomic sentence is stated to be false only (F),
- an atomic sentence is stated to be both true and false, for instance, by different sources, or in different points of time (Both), and
- an atomic sentences status is unknown. That is, neither true, nor false (None).

The four truth values can be arranged in a lattice, called logical lattice, Fig. 1 where logical conjunction is identified with the meet operation and logical disjunction with the join operation.

Extensions of Belnap's logic are discussed in [12].

AHP is widely used as a stand-alone application and as a tool in combination with other techniques, [18]. Recent applications are presented in [20]. The AHP usually begins with making a control hierarchy or network of objectives and criteria that control the interactions in a particular system. Next stage involves development

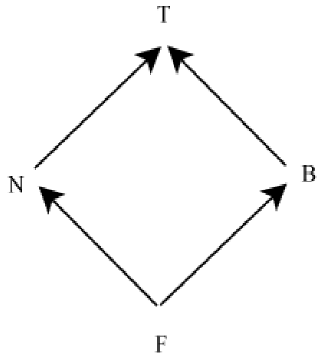


Fig. 1: Logical lattice

of sub-networks of influences among the elements and clusters of the problem, one for each control criterion.

A system contains components (levels or clusters) where the elements (attributes) of each component interact or have an influence on some or all of the elements of another component with respect to a property governing the interactions of the entire system. The influence of a given set of elements in a component on any element in the system is illustrated by a priority vector derived from the AHP paired comparison approach. Decision makers' judgements are consistent if

$$a_{ij}a_{jk} = a_{ik}, \quad i, j, k = 1, 2, ..., n.$$

In this content consistency means that if a basic amount of row data is available than all other data can be logically deduced from it. Application of eigen vectors leads to a very useful consistency measure called consistency index *CI*, [19]. The *CI* of a comparison matrix is defined as

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

where n is the order of the comparison matrix and λ_{max} is its maximum eigen value. *CI* measures the transitivity of a preference that is a part of the pairwise comparisons.

A random index *RI* is the mean *CI* value of random generated matrices of size n, [19], see Table 1. A consistency ratio *CR* [19] is defined as

$$CR = \frac{CI}{RI}$$

and is a measure of comparison between a given matrix and a random generated matrix in terms of consistency indexes. The upper bound for an acceptable *CR* is 0.1. A revision of judgements is required if larger values are obtained.

An alternative way to measure consistency is proposed in [16].

Table 1: Random indexces for matrices of size n

n	1	2	3	4	5	6	7
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32

AHP is an aggregation method that suffers rank reversal, [3] and [6]. An even stronger shortcoming consists in the fact that the addition of indifferent criteria (for which all alternatives perform equally) causes a significant alteration of the aggregated priorities of alternatives, with important consequences, [17].

3 Alternatives Evaluation

We propose employment of a decision support system for selecting the most desirable collaboration options. All collaboration options for a particular firm are evaluated by a predetermined number (say n) of experts, where they all agree on the number and type of evaluation alternatives. Once received all n-touples of responses are arranged in a lattice structure according to their degrees of certainty. The lattice type is determined by the number of experts and the number of evaluation alternatives. The most profitable collaboration options are the ones placed in the top nodes. Belnap's logic is subsequently applied for a finer grading, [5].

Further cost-benefit calculation for deciding how many partners to include in an alliance is needed [7].

4 Managing Seemingly Profitable Projects

In case several projects are found equally profitable we propose application of AHP for obtaining the last recommendation. The use of AHP is reserved for the last stage since it requires additional experts' opinions related to pairwise comparisons of homogeneous elements. It is illustrated by Example 0.1. Some details are omitted since we operate with real data and the data provider prefers to be anonymous.

Example 0.1. The hierarchical structure is presented in Fig. 2. Level 1 is related to the goal, Level 2 contains the applied criteria, and Level 3 contains the alternatives that are to be considered.

The goal of this evaluation model is to determine the most suitable collaboration arrangement for a particular firm. Three alternatives -

- collaborate with one firm A1,
- separate collaborations with two firms A2, and
- joint ventures with several firms A3

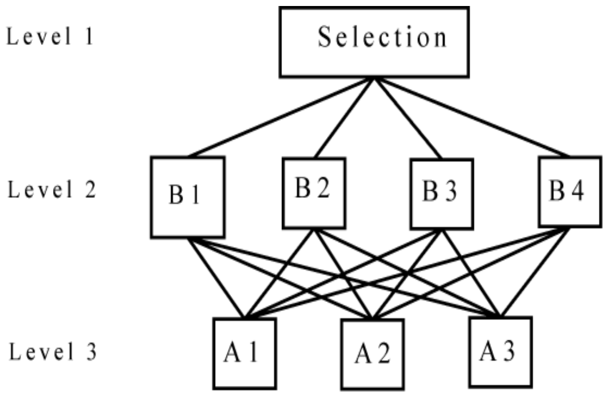


Fig. 2: Selection of the most suitable collaboration arrangement

are to be considered with respect to four criteria -

- benefits B1,
- costs B2,
- opportunities B3,
- risks B4.

The four criteria comparison has consistency index $CI = 0.08$.

	B_1	B_2	B_3	B_4
B_1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{3}$
B_2	2	1	$\frac{1}{2}$	5
B_3	4	2	1	4
B_4	3	$\frac{1}{5}$	$\frac{1}{4}$	1

General criteria distribution is shown in Fig. 3

A comparison of alternatives A_1,A_2 and A_3 with respect to criterion B_1 has consistency index $CI = 0.02$.

B_1	A_1	A_2	A_3
A_1	1	3	4
A_2	$\frac{1}{3}$	1	2
A_3	$\frac{1}{4}$	$\frac{1}{2}$	1

A comparison of alternatives A_1,A_2 and A_3 with respect to criterion B_2 has consistency index $CI = 0.07$.

B_2	A_1	A_2	A_3
A_1	1	2	5
A_2	$\frac{1}{2}$	1	6
A_3	$\frac{1}{5}$	$\frac{1}{6}$	1

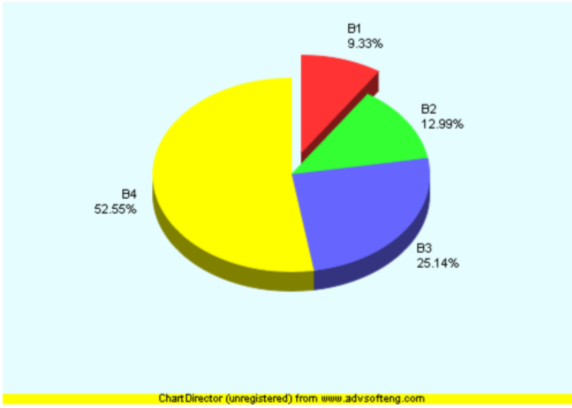


Fig. 3: Criteria distribution

A comparison of alternatives A_1,A_2 and A_3 with respect to criterion B_3 has consistency index $CI = 0.07$.

B_3	A_1	A_2	A_3
A_1	1	4	5
A_2	$\frac{1}{4}$	1	3
A_3	$\frac{1}{5}$	$\frac{1}{3}$	1

A comparison of alternatives A_1,A_2 and A_3 with respect to criterion B_4 has consistency index $CI = 0.09$.

B_4	A_1	A_2	A_3
A_1	1	$\frac{1}{2}$	$\frac{1}{3}$
A_2	2	1	$\frac{1}{4}$
A_3	3	4	1

The obtained normalized vector

	Normalized vector
A_1	0.378
A_2	0.242
A_3	0.380

indicates that alternative A_3 is better than alternative A_1 while alternative A_2 seems to be leas promising.

A Head-to-Head Analysis of alternatives A_1 and A_3 is enclosed in order to provide more detailed information about where the advantages of A_3 are, (see Fig. 4). It indicates that

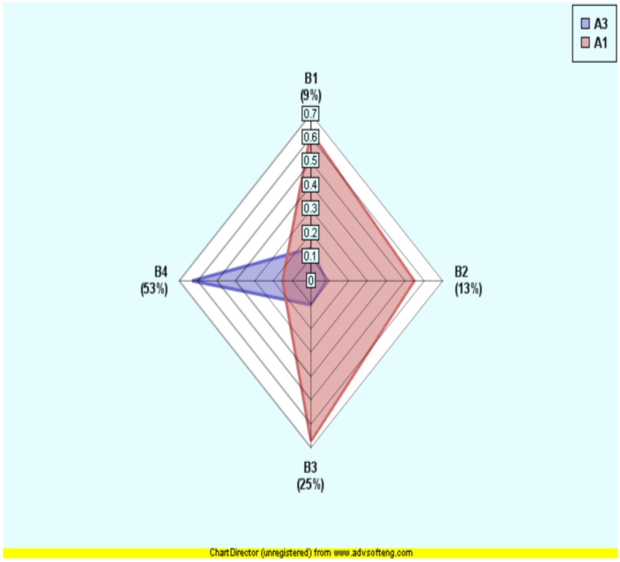


Fig. 4: Head-to-Head Analysis of alternatives A_1 and A_3.

- alternative A_1 dominates over alternative A_3 considerably with respect to criteria B1, B2 and B3, while
- alternative A_3 dominates over alternative A_1 considerably with respect to criteria B4 only.

Serious discussions among decision makers and stakeholders are in place before a decision model is implemented. Software supporting the AHP is presented in [22].

5 Prototype system

A prototype system for automated evaluation and selection of collaboration alternatives will be implemented as a Web-based application server using; Apache HTTP server, Python programmable runtime support using mod_python, and a self-contained, serverless, zero-configuration, transactional SQlite database engine.

System's users, i.e. customers, experts, and administrators (see Fig. 5), interact with the system using Web-based interface. The system will support multiple database architecture. System users' information is saved in USERS database. System's projects information is saved in PROJECTS database. Each separate project will have its own database, for example PR108VDR. The project database will be used to store project's related data.

Initial data structure for a project, say PR108VDR, that represents a particular lattice structure will be constructed and saved in the PR108VDR database. The status of experts' evaluation of collaboration options for project PR108VDR will also be stored in the PR108VDR database.

The system supports three interfaces: Project interface, Admin interface, and Evaluation interface. These interfaces provide different functionalities to different user types. Before any users can access the system, they need to be authenticated through their respective interfaces.

Project's interface provides access point for the customers to submit projects, view report and diagram of customers' projects collaboration options and reads experts' evaluation reports of projects. The project databases belonging to a customer are viewable by that particular customer at any time. The customer can give view access right to other customers or to the public on any of the databases belonging to the customer.

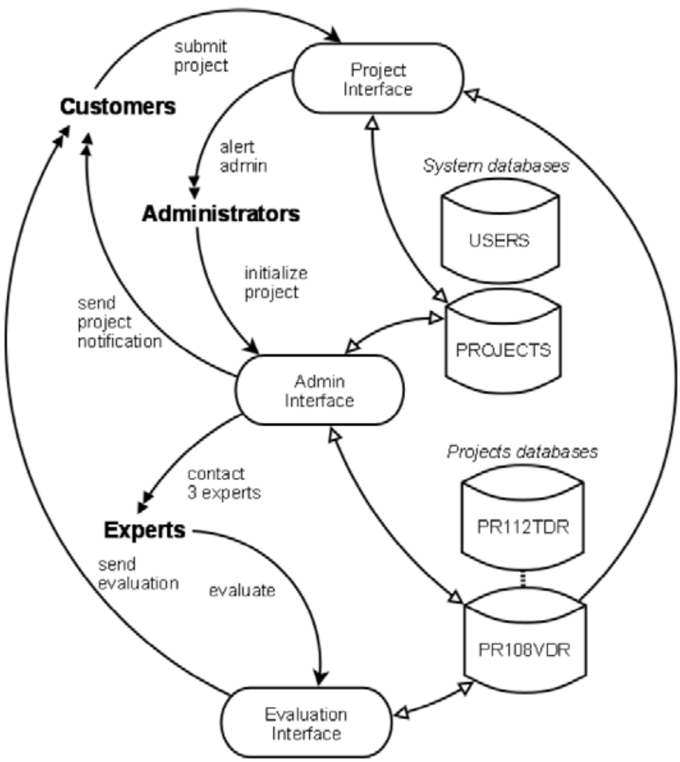


Fig. 5: System workflow for evaluations of profitable projects

Admin interface provides administrative access point to system administrator. A project is initiated by a customer. The administrators then initialize the project by creating the project database and constructing the project collaboration options lattice structure. The administrators then contact three best available exports for this particular project. Once all the experts are assigned to the project, the customer will be notified and be given all particulars concerning the project for example project's URL.

Evaluation interface provides a specific project access point for the evaluation experts. The evaluations for each of the collaboration options for being profitable in relation to a specific criterion are done using dynamically created Web-forms. Each expert has her own private evaluation space for the project. Once all the experts are done with the evaluation process and the system completed its calculation, the customer will be notified that the project is accomplished.

Finally, the final project's diagrams and reports will be written into the system PROJECTS database.

6 Conclusion

AHP proves to be very useful in allocating resources according to their ratio-scale priorities, since it works with deriving ratio scale measurements. AHP could have been used from the start but it would take much more resources.

References

1. Belnap, N.J.: How a computer should think. In: Proceedings of the Oxford International Symposia on Contemporary Aspects of Philosophy, Oxford, GB, pp. 30–56 (1975)
2. Belnap, N.J.: A useful four valued logic. In: Dunn, J.M., Epstain, G. (eds.) Modern uses of multiple-valued logic, pp. 8–37. D. Reidel Publishing Co., Dordrecht (1977)
3. Belton, V., Gear, T.: On a Short-coming of Saaty's Method of Analytic Hierarchies. Omega 11(3), 228–230 (1983)
4. Bierman, H., Smidt, S.: The Capital Budgeting Decision. Routledge, New York (2007)
5. Coste-Marquis, S., Marquis, P.: Complexity Results for Paraconsistent Inference Relations. In: Fensel, D., Giunchiglia, F., McGuinness, D.L., Williams, M.-A. (eds.) Proceedings of the Eights International Conference on Principles and Knowledge Representation and Reasoning (KR 2002), Toulouse, France, pp. 61–72. Morgan Kaufmann, San Francisco (2002)
6. Dyer, J.S.: Remarks on the Analytic Hierarchy Process. Management Science 36(3), 249–258 (1990)
7. Dyer, J.H., Kale, P., Singh, H.: How to Make Strategic Alliances Work. Sloan Management Review 42(4), 37–43 (2001)
8. Ganapathy, S., Narayanan, S., Srinivasan, K.: Logistics: simulation based decision support for supply chain logistics. In: Proceedings of the 35th conference on Winter simulation: driving innovation, New Orleans, Louisiana, pp. 1013–1020 (2003)
9. Ganter, B.: Two Basic Algorithms in Concept Analysis, FB4-Preprint 831, TH Darmstadt (1984)
10. Ganter, B., Stumme, G., Wille, R. (eds.): Formal Concept Analysis. LNCS, vol. 3626. Springer, Heidelberg (2005)
11. Jing, L., Jiafu, W.: Administration decision support system for shipping enterprises. In: Proceedings of 1999 IEEE International Conference on Systems, Man and Cybernetics 3, pp. 1048–1053 (1999)

12. Kaluzhny, Y., Muravitsky, A.Y.: A knowledge representation based on the Belnap's four valued logic. Journal of Applied Non-Classical Logics 3, 189–203 (1993)
13. Klein, G.A.: Recognition-Primed Decision Making. In: Sources of power: How people make decisions, pp. 15–30. MIT Press, Cambridge (1998)
14. Layard, R., Glaister, S.: Cost-Benefit Analysis, 2nd edn. Cambridge University Press, Cambridge (1994)
15. Nas, T.F.: Cost-benefit Analysis, Theory and Application. Sage Publications, Thousand Oaks (1996)
16. Nishizawa, K.: A consistency improving method in binary AHP. Journal of the operations research Society of Japan 38(1) (1995)
17. Perez, J., Jimeno, J.L., Mokotoff, E.: Another Potential Shortcoming of AHP. Business and Economics 14(1), 99–111 (2006)
18. Promentilla, M.A.B., Furuichia, T., Ishiia, K., Tanikawa, N.: Evaluation of remedial countermeasures using the analytic network process. Waste Management 26(12), 1410–1421 (2006)
19. Saaty, T.L.: The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. McGraw-Hill, New York (1980)
20. Vaidya, O., Kumar, S.: Analytic Hierarchy Process: an overview of applications. European Journal of Operational Research 169, 1–29 (2006)
21. Zirger, B.J., Maidique, M.: A model of new product development: an empirical test. Management Science 36, 867–883 (1990)
22. http://www.superdecisions.com/

Towards Semantic-based RSS Merging

F. Getahun, J. Tekli, M. Viviani, R. Chbeir, and K. Yetongnon

Abstract Merging information can be of key importance in several XML-based applications. For instance, merging the RSS news from different sources and providers can be beneficial for end-users (journalists, economists, etc.) in various scenarios. In this work, we address this issue and mainly explore the relatedness relationships between RSS entities/elements. To validate our approach, we also provide a set of experimental tests showing satisfactory results.

1 Introduction

Really Simple Syndication (RSS) [17] is an XML-based family of web feed formats, proposed to facilitate the aggregation of information from multiple web sources. This way, clients can simultaneously access content originating from different providers rather than roaming a set of news providers, but they often have to read related (and even identical) news more than once as the existing RSS engines[1] do not provide facilities for merging related items.

Merging XML-based documents stands for (*i*) identifying semantically related elements between two documents, and (*ii*) generating a merged document that collapses these related elements preserving remaining source elements. In this work, we address the first problem, and particularly focus on measuring the *semantic*

Laboratoire Electronique, Informatique et Image (LE2I) - UMR-CNRS
Universit de Bourgogne - Sciences et Techniques
Mirande, Aile de l'Ingénieur, 9 av. Savary - 21078 Dijon Cedex, France
e-mail: {fekade-getahun.taddesse, joe.tekli, marco.viviani, rchbeir, kokou}@u-bourgogne.fr

[1] AmphetaDesk, MetaDot, Meerkat, Portal Software, PullRss, Radio UserLand, SlashCode/Slashdot, Weblog 2.0 aggregate, search, filter or display news in RSS format

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 53–64.
springerlink.com

relatedness[2] [2] between RSS elements/items (labels and contents) and consecutively element semantic relationships w.r.t. the meaning of terms and not only their syntactic properties, as a necessary prerequisite to performing efficient RSS merging. To motivate our work, let us consider Figure 1 and Figure 2 showing a list of news (showing only their title and description) extracted from CNN and BBC's RSS feeds. Identifying (and merging) related news would enable the user to more easily and efficiently acquire information. XML news feeds (e.g., RSS items) can be related in different manners:

- The content of an element might be totally included in another *(inclusion)*.
  Example 1. The title content of *CNN1* "Hong Kong cheers Olympic torch" includes the title content of BBC1 "Torch cheered through Hong Kong" [3].
- Two news may refer to similar and related concepts *(intersection)*.
  Example 2. The title content of *CNN2* "Bush wants $700 million more emergency food aid" and title content of BBC2 "US president offers $700m for food crisis" are related and very similar, they share some words/expressions ('$700m', 'food') and semantically related concepts ('emergency' and 'crisis', 'US President' and 'Bush').

```
<CNN_RSS>
    <item>
        <title>Hong Kong cheers Olympic torch</title>
        <description>Hong Kong stages the Olympic torch relay, the first time the event is held on the soil of the host of this year's      CNN1
            Summer Games, with thousands lining the route in support of the event and a peppering of protests.</description>
    </item>
    <item>
        <title>Bush wants $770 million more emergency food aid</title>
        <description>U.S. President George W. Bush urges Congress to approve $770 million in new global food aid to be made      CNN2
            available beginning in October. The sum would be in addition to$200 million in emergency food aid announced two
            weeks ago.</description>
    </item>
</CNN_RSS>
```

Fig. 1: RSS news extracted from CNN.

Hence, the main objective of this study is to put forward a specialized XML relatedness measure, dedicated to the comparison of RSS items, able to (*i*) identify RSS items that are related enough to be merged and (*ii*) identify the relationships that can occur between two RSS items (i.e., *disjointness, intersection, inclusion,* and *equality*), to be exploited in the merging phase. Identifying common/different parts in the items to be merged would help decide on the merging rules to be executed in different application scenarios. Note that the merging phase itself (merging rules, merging process, ...) is not developed in this paper. The remainder of this paper is organized as follows. In Section 2, we discuss background and related work.

[2] Semantic relatedness is a more general concept than similarity. Dissimilar entities may also be semantically related by lexical relations such as meronymy and antonymy, or just by any kind of functional relation or frequent association

[3] After text pre-processing such as stop-word removal, stemming, and semantic analysis

```
<BBC_RSS>
    <item>
        <title>Torch cheered through Hong Kong</title>
        <description>Cheering crowds and a few protesters turn out in Hong Kong to watch the Olympic torch parade.</description>      BBC1
    </item>
    <item>
        <title>US president offers $770m for food crisis</title>
        <description>George W Bush offers $770m (£390m) in new international food aid to help ease the effects of surging food      BBC2
            prices.</description>
    </item>
</BBC_RSS>
```

Fig. 2: RSS news extracted from BBC.

Section 3 defines basic concepts to be used in our measure. Section 4 details our RSS relatedness measure. Section 5 presents experimental results. Finally, Section 6 concludes this study and draws future research directions.

2 Related Work

Identifying correspondence or matching nodes is a known precondition in schema matching [4] and merging XML document [9]. In schema matching, corresponding nodes or elements are identified using the match operator. A lot of research has been done to determine similarity and are categorized into structure-based, semantic-based and hybrid-based approaches. It is to be noted that most of the proposed approaches in XML comparison are based on structural similarity using tree edit distance [1]. Chawathe [3], Nireman and Jagadish [12] consider the minimum number of edit operations: insert, delete and/or move to transform one XML tree to another. Also, the use of Fast Fourier Transform [5] has been proposed to compute similarity between XML documents.

The semantic similarity between concepts is estimated either by the distance between nodes [19] or the content of the most specific common ancestor of those nodes involved in the comparison [15][10] and defined according to some predefined knowledge base(s). Knowledge bases [14][16](thesauri, taxonomies and/or ontologies) provide a framework for organizing words (expressions) into a semantic space. In Information Retrieval (IR) [11], the content of a document is commonly modeled with set/ bag of words where each concept (and subsumed word(s)) is given a weight computed with Term Frequency (TF), Document Frequency (DT), Inverse Document Frequency (IDF), and the combination TF-IDF. In [7], the authors used a Vector Space having TF-IDF as weight factor in XML retrieval.

More recently, there are hybrid-based approaches that attempted to address XML comparison. In a recent work [18], the authors combined an IR semantic similarity technique with a structural-based algorithm based on edit distance. However, the semantic similarity is limited only to tag name. In [8], *xSim*, a structure and content aware XML comparison framework is presented. *xSim* computes the matching

between XML documents as an average of matched list similarity values. The similarity value is computed as average of content, tag name and path similarity values without considering semantics.
However and to the best of our knowledge, none of the current techniques or measures identifies the semantic relationship between documents and semantic relatedness on content in general or items in particular and none of the approaches is RSS-focused.

3 Preliminaries

In the following, we define the basic concepts used in our approach and particularly detail RSS data model and hierarchal neighbourhood of a concept.

3.1 RSS data model

An RSS document comes down to a well-formed XML document (represented as a rooted ordered labeled tree following the Document Object Model (DOM) [20]) w.r.t. an RSS schema [17]. Note that different RSS schemas exist, corresponding to the different versions of RSS[4] available on the web. Nonetheless, analyzing different versions of RSS, we can see that RSS items consistently follow the same overall structure, adding or removing certain elements depending on the version at hand.

Definition 1 (Rooted Ordered Labeled Tree) It is a rooted tree in which the nodes are labeled and ordered. We denote by R(T) the root of T.

Definition 2 (Element) Each node of the rooted labeled tree T is called an *element* of T. Each element e is a pair $e = \langle \eta, \zeta \rangle$ where $e.\eta$ refers to the element name and $e.\zeta$ to its content. $e.\eta$ generally assumes an atomic text value (i.e., a single word/expression) whereas $e.\zeta$ may assume either an atomic text value, a composite text value (sentence, i.e., a number of words/expressions), or other elements[5].

Definition 3 (Simple/Composite Element) An element e is *simple* if $e.\zeta$ assumes either an atomic or composite textual value[6]. In XML trees, simple elements come down to leaf nodes. Content and value of simple element are used interchangeably in this paper. An element e is *composite* if $e.\zeta$ assumes other elements. In XML trees, composite elements correspond to inner nodes.

[4] RSS refers to one of the following standards: Rich Site Summary (RSS 0.91, RSS 0.92), RDF Site Summary (RSS 0.9 and 1.0), and Really Simple Syndication (RSS 2.0).

[5] We do not consider attributes in evaluating RSS item relatedness since they do not affect the semantic comparison process. Nonetheless, attributes will be considered in the merging phase

[6] In this paper, we do not consider other types of data contents, e.g., numbers, dates,…

Definition 4 (RSS Item Tree) An *RSS item tree* is an XML tree T having one single composite element, the root node r(usually with $e.\eta$ = 'item'), and k simple elements $n_1,\ldots,n_k$ describing the various RSS item components.

3.2 Knowledge Base

A *Knowledge Base* [16] (thesauri, taxonomy and/or ontology) provides a framework for organizing entities (words/expressions, generic concepts, web pages, etc.) into a semantic space. In our study, it is used to help computing relatedness and formally defined as $KB = (C,E,R,f)$ where C is the set of concepts (synonym sets as in WordNet [14]), E is the set of edges connecting the concepts,$E \subseteq CxC$, R is the set of semantic relations, $R = \{\equiv,\prec,\succ,<<,>>,\Omega\}$[7], the synonymous words/expressions being integrated in the concepts, f is a function designating the nature of edges in E, $f : E \rightarrow R$.
Following Definition 4, all elements of an RSS item (to the exception of the root) are simple, i.e., each composed of a label and a textual value (content). Hence, assessing the relatedness between (simple) RSS elements requires considering label as well as textual value relatedness. To that end, we introduce two knowledge bases: *(i) value-based*: to describe the textual content of RSS elements, and *(ii) label-based:* to organize RSS labels[8].

3.3 Neighborhood

In our approach, the *neighborhood* of a concept C_i underlines the set of concepts $\{C_j\}$, in the knowledge base, that are subsumed by Ci w.r.t. a given semantic relation. The concept of neighborhood, introduced in [6], is exploited in identifying the relationships between text (i.e., RSS element labels and/or textual contents) and consequently RSS elements/items.

Definition 5 (Semantic Neighborhood) The *semantic neighborhood* of a concept C_i (i.e. $N^R_{KB}(C_i)$) is defined as the set of concepts $\{C_j\}$ (and consequently the set of words/expressions subsumed by the concepts) in a given knowledge base KB, related with C_i via the hyponymy or meronymy semantic relations, directly or via transitivity.

[7] R underlines respectively the synonym ($\equiv$), hyponym (Is-A or $\prec$), hypernym (Has-A or $\succ$), meronym (Part-Of or $<<$), holonym (Has-Part or $>>$) and Antonym (Ω) relations, as defined in [6].

[8] Note that one single knowledge base could have been used. However, since XML document labels in general, and RSS labels in particular, depend on the underlying document schema, an independent *label-based* knowledge base, provided by the user/administrator, seems more appropriate than a more generic one such as WordNet (treating generic textual content).

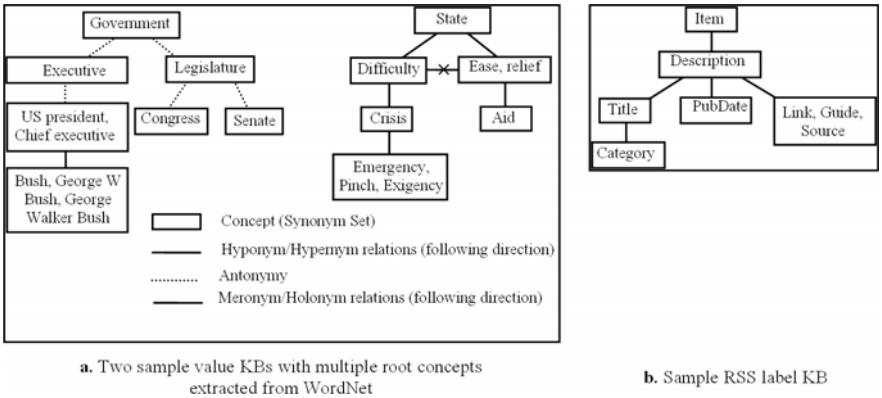


Fig. 3: Sample value and label knowledge bases.

Definition 6 (Global Semantic Neighborhood) The *global semantic neighborhood* $\overline{N_{KB}}(C_i)$ of a concept is the union of each semantic neighborhood w.r.t. all synonymy, hyponymy and meronymy relations altogether.

3.4 Text Representation

As illustrated previously, RSS (simple) element labels and contents underline basic text (cf. Definition 2). Thus, hereunder we define the idea of *concept set* to represent a piece of text. It will be exploited in representing (and consequently comparing) RSS element labels and contents.

Definition 7 (Concept Set) Consider a textual value t, composed of a set of terms $\{k_1, \ldots, k_n\}$, where n is the total number of distinct terms in t, i.e., $|t|$. The *concept set* of t, denoted as CS, is a set of concepts $\{C_1, \ldots, C_m\}$, where each C_i represents the meaning of a group of terms in $\{k_1, \ldots, k_n\}$, where m is the total number of concepts describing t, i.e., $m = |CS_t|$, having $0 \leq |CSt| \leq |t|$. Concept C_i is assumed to be obtained after several textual preprocessing operations such as stop-words removal [9], stemming[10], etc.

Definition 8 (Text Vector Space) Let t_i be a text value described by concept set $CS_i = \{C_1, \ldots, C_n\}$. Following the vector space model used in information retrieval [11], we represent t_i as a vector V_i in an n-dimensional space such as:

[9] Stop-words are very common words such as prepositions, demonstrative, articles, etc which do not provide useful information to distinguishing the content of the items (e.g. yet, an, but, ...).

[10] Stemming the process for reducing inflected (or derived) words to their stem or base form (e.g., "housing", "housed"→ "house"

$V_i = [\langle C_1, w_1 \rangle, \ldots, \langle C_n, w_n \rangle]$, where w_i represents the weight associated to dimension (concept) C_i. Given two texts t_1 and t_2, the vector space dimensions represent each a distinct concept $C_i \in CS_1 \cup CS_2$, such as $1 \leq i \leq n$ where $n = |CS_1 \cup CS_2|$ is the number of distinct concepts in both CS_1 and CS_2.

Definition 9 (Vector Weights) Given a collection of texts T, a text $t_i \subset T$ and its corresponding vector V_i, the weight w_i associated to a concept C_i in V_i is calculated as $w_i = 1$ if the concept C_i is referenced in the vector V_i; otherwise, it is computed based on the maximum *enclosure similarity* it has with another concept C_j in its corresponding vector V_j.

$$\text{enclosure\_sim}(C_i, C_j) = \frac{\overline{N_{KB}}(C_i) \cap \overline{N_{KB}}(C_j)}{\overline{N_{KB}}(C_j)} \tag{1}$$

enclosure_sim(C_i, C_j) takes into account the global semantic neighborhood of each concept. This measure returns a value of 1 if C_i includes C_j.

Example 3. Let us consider titles of RSS items *CNN2* and *BBC2* (Figures 1, 2). The corresponding vector representations V_1 and V_2 are shown in Figure 4. For the sake of simplicity, we consider that only these two texts make up the new items.

	Bush	*want*	*700m*	*emergency*	*food*	*aid*	*US president*	*offer*	*crisis*
V_1	1	1	1	1	1	1	**1**	0	**1**
V_2	**0.56**	0	1	**0.4**	1	0	1	1	1

Fig. 4: Vectors obtained when comparing title texts of RSS items CNN2 and BBC2.

Vector weights are evaluated in two steps. First, for each concept C in V_1 and V_2, we assign value of 1 if C exists in the concept sets corresponding to the texts being compared. Second, we update the weight of those concepts having value of zero with maximum semantic enclosure similarity value. Following the WordNet extract in Figure 3a, the concept 'US president' is included in the global semantic neighborhood of 'Bush', i.e., *US president* $\in \overline{N_{KB}}(Bush)$. Hence, *enclosure_sim(US president, Bush)* $= 1$. However, in V_2, *enclosure_sim(Bush, US president)* $= 0.56$. Likewise, 'Crisis' is included in the global semantic neighborhood of 'Emergency', i.e., *Crisis* $\in \overline{N_{KB}}(Emergency)$. Thus, *enclosure_sim(Crisis, Emergency)* $= 1$ but *enclosure_sim(Emergency, Crisis)* $= 0.4$.

4 RSS Relatedness Measure

As motivated in the beginning of the paper, a dedicated relatedness/similarity measure is needed as a prerequisite to merging RSS data. This section details the measures used for text, simple and complex element relatedness.

4.1 Text Relatedness

Given two texts t_1 and t_2, *Textual Relatedness* (TR) algorithm returns a doublet, combining the semantic relatedness *SemRel* value and the relationship Relation between t_1 and t_2. Formally, it is denoted as:

$$TR(t_1,t_2) = \langle SemRel(t_1,t_2), Relation(t_1,t_2) \rangle \quad (2)$$

SemRel value is computed using vector based similarity measure (e.g. cosine [11]) applied to text vector having weights underlining concept existence and enclosure in the concept set of both text inputs (definition 9).

The relationship between two texts t1 and t2 is identified as follows:

- *Relation*(t_1, t_2) = *Disjointness*, i.e.,$t_1 \triangleright\triangleleft t_2$, if there is no relatedness whatsoever between t_1 and t_2 i.e., $SemRel(t_1,t_2) = 0$.
- *Relation*(t_1, t_2) = *Inclusion*, i.e., $t_1 \supset t_2$, if the product of the weights of vector V_1 (describing t_1) is equal to 1, i.e., $\prod_{v_i}(w_p) = 1$. The weight product of V_1 underlines whether or not t1 encompasses all concepts in t_2.
- *Relation*(t_1, t_2) = *Intersection*, i.e., $t_1 \cap t_2$, if t_1 and t_2 share some semantic relatedness, i.e., $SemRel(t_1,t_2) > 0$, and the product of the weights of both vectors V_1 and V_2 are equal to zero, i.e., $\prod_{v_1}(w_p) \geq 0$ and $\prod_{v_2}(w_q) \geq 0$.
- *Relation*(t_1, t_2) = *Equality*, i.e.,$t_1 = t_2$, if corresponding vectors are identical, i.e., $SemRel(t_1,t_2) = 1$.

  Example 4. Considering Example 3, (t_1 of CNN2 and t_2 of BBC2), $SemRel(t_1, t_2) = 0.68$ and $Relation(t_1,t_2) = Intersection$. Hence, $TR(t_1,t_2) =< 0.68, Intersection >$.

4.2 RSS Item Relatedness

As shown previously, quantifying the semantic relatedness and identifying the relationships between two RSS items amounts to comparing corresponding elements. This in turn comes down to comparing corresponding RSS (simple) element labels and values (contents), which simplify to basic pieces of text (cf. Definition 2). The relatedness between two simple elements is computed applying *TR* Algorithm for both text content and label. *ER* algorithm accepts two elements e_1 and e_2 as input and returns doublet quantifying the semantic relatedness *SemRel* and the relationships *Relation* between e_1 and e_2 based on corresponding label and value relatedness. *SemRel*(e_1, e_2) semantic the relatedness value between elements, is quantified as *weighted sum* value of label and value relatedness as:

$$SemRel(e_1, e_2) = w_{label} \times LB_{SemRel} + W_{Valie} \times VR_{SemRel} \quad (3)$$

where $w_{Label} + w_{Value} = 1$ and $(w_{Label}, w_{Value}) \geq 0$. The relation between elements is computed based on rule rule-based method that combines label and value relationships as follows:

- Elements e_1 and e_2 are disjoint if either their labels and values are disjoint
- Element e_1 includes e_2, if $e_1.\eta$ includes $e_2.\eta$ and $e_1.\zeta$ includes $e_2.\zeta$
- Two elements e_1 and e_2 intersect if either their labels or values intersect
- Two elements e_1 and e_2 are equal if both their labels and values are equal.

Having identified the semantic relatedness and relationships between simple elements, Algorithm 1 evaluates RSS item relatedness. Given two RSS items I_1 and I_2, each made of a bunch of elements, *Item Relatedness (IR)* algorithm quantifies the semantic relatedness and identifies the relationships between I_1 and I_2 based on corresponding element relatedness (lines 7 - 12). Line 9 computes the relatedness between simple elements e_i and e_j and returns semantic relatedness eij_{SemRel}, and relationship $eij_{Relation}$. In line 10, semantic relatedness value eij_{SemRel} is accumulated to get grand total, and, in line 11, $eij_{Relation}$ is stored for later use. In line 13, the semantic relatedness value between I_1 and I_2 is computed as the average of the relatedness values between corresponding element sets I_1 and I_2.

Algorithm 1: ***IR Algorithm***	Line
Input: I_1, I_2: element // the two items (Complex elements)	1
Variable: eij_{SemRel}: double // semantic relatedness values e_i and e_j	
$eij_{Relation}$: string // relationship value between e_i and e_j	
$Eij_{Relation\_set}$: Set // would contain sub-elements relationship values	
Output: *SemRel*: double // relatedness value between I_1 and I_2	
Relation: String // relationship value between I_1 and I_2	
SumRel=0	
$Eij_{Relation\_set} = \varnothing$	
For each e_i In I_1	7
For each e_j In I_2	
$\langle eij_{SemRel}, eij_{Relation} \rangle = ER(e_i, e_j)$	9
$Eij_{Relation\_set} = Eij_{Relation\_set} \cup eij_{Relation}$	
$SumRel = SumRel + eij_{SemRel}$	
Next	
Next	
$SemRel = SumRel \,/\, \lvert I_1 \rvert \times \lvert I_2 \rvert$	13
$Relation = I_{Relation}(\{Eij_{Relation\_set}\})$ // $\forall i \in [1, \lvert I_1 \rvert], \ \forall j \in [1, \lvert I_2 \rvert]$	
Return <*SemRel, Relation*>	15

As for the relationships between two items, we develop a rule-based method $I_{Relation}$ (line 14) for combining sub-element relationships stored in $Eij_{Relation\_set}$ (which is the relationship between e_i and e_j) as follows:

- Items I_1 and I_2 are *disjoint* if all elements $\{e_i\}$ and $\{e_j\}$ are disjoint (elements are disjoint if there is no relatedness whatsoever between them, i.e., $SemRel(I_1, I_2) = 0$)
- Item I_1 *includes* I_2, if all elements in $\{e_i\}$ include all those in $\{e_j\}$

- Two items I_1 and I_2 *intersect* if at least two of their elements intersect
- Two items I_1 and I_2 are *equal* if all their elements in $\{e_i\}$ equal to all those in $\{e_j\}$.

  Example 5. Let us consider RSS items *CNN2* and *BBC2* (Figure 1, 2). Corresponding item relatedness is computed as follows. Notice that $w_{Label} = 0.1$ and $w_{Value} = 0.9$ is used while computing simple element relatedness (cf. 3). Below, each cell represent doublet returned by simple element relatedness ER algorithm.

ER	*title*$_{BBC2}$	*description*$_{BBC2}$
title$_{CNN2}$	<0.700, intersection>	<0.526, intersection>
description$_{CNN2}$	<0.483, intersection>	<0.435, intersection>

Using (line 13), *SemRel(CNN2, BBC2)* = $(0.700 + 0.526 + 0.483 + 0.538 + 0.435)/2 \times 2 = 0.671$. where $|I_1|$ and $|I_2|$ are equal to 2.

Relation(CNN2, BBC2) = Intersection since a number of their elements intersect.

Hence, IR(CNN2, BBC2) $=< 0.671,$ *Intersection*>

5 Experiments

We have conducted a set of experiments to conform the computational complexity and efficiency of our relatedness measure in comparison with current approaches. All the experiments were carried out on Intel Core Centrino Duo Processor machine (with processing speed of 1.73.0 GHz, 1GB of RAM). The experiments related to measure the efficiency of our relatedness measure won't be detailed due to the lack of space. We compared the efficiency of *xSim* [8], TF-IDF and our algorithm while identifying the relatedness of randomly generated synthetic news. The aim here was to compare our time processing with others. In all algorithms, relatedness identification is done without semantics as both *xSim* and TF-IDF do not consider semantics information. Figure 5 shows that our semantic relatedness method provides efficient result compared to *xSim* and it is less efficient compared to TF-IDF (which doesn't consider the structure of the RSS news item). Hence, our relatedness algorithm is efficient and also identifies relationship between text, element and items which is not the case in both *xSim* and TF-IDF.

6 Conclusions and Future Directions

In this paper, we have addressed the issue of measuring relatedness between RSS items, a pre-condition for merging. We have studied and provided a technique for texts, simple elements and items relatedness computation, taking into account

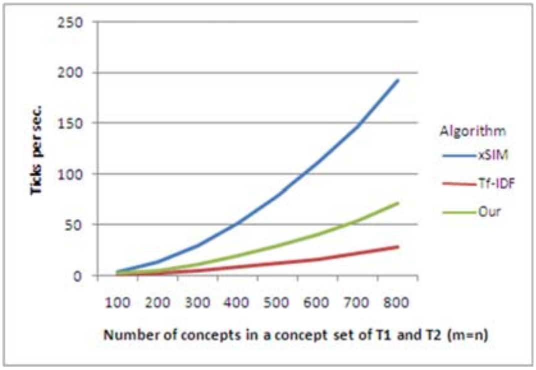


Fig. 5: Timing result obtained using three algorithms: *xSim*, TF-IDF and our algorithm.

different kinds of relationship among texts, elements and items. We have developed a prototype and compared the efficiency of our algorithm against *xSim* and TF-IDF. Our measure will help us in making decisions about merging rules to apply to clustered elements. Currently, we are investigating the relevance of the topological relationships in the construction of the merging rules. Later on, we are willing to extend our work so to address XML documents merging in multimedia scenarios (SVG, MPEG-7, etc.).

References

1. Bille, P.: A survey on tree edit distance and related problems. Theoretical Computer Science 337(1-3), 217–239 (2005)
2. Budanitsky, A., Hirst, G.: Evaluating WordNet-based measures of lexical semantic relatedness. Computational Linguistics 32(1), 13–47 (2006)
3. Chawathe, S.S.: Comparing hierarchical data in external memory. In: VLDB 1999: Proceedings of the 25th International Conference on Very Large Data Bases, pp. 90–101. Morgan Kaufmann Publishers Inc., San Francisco (1999)
4. Rahm, E., Bernstein, P.A.: A survey of approaches to automatic schema matching. VLDB Journal 10(4), 334–350 (2001)
5. Flesca, S., Manco, G., Masciari, E., Pontieri, L.: Fast detection of xml structural similarity. IEEE Transactions on Knowledge and Data Engineering 17(2), 160–175 (2005) (Student Member-Andrea Pugliese)
6. Getahun, F., Tekli, J., Atnafu, S., Chbeir, R.: Towards efficient horizontal multimedia database fragmentation using semantic-based predicates implication. In: XXII Simposio Brasileiro de Banco de Dados (SBBD 2007), Joao Pessoa, Brazil, October 15-19, pp. 68–82 (2007)
7. Grabs, T., Schek, H.-J.: Generating Vector Spaces On-the-fly for Flexible XML Retrieval. In: Proceedings of the ACM SIGIR Workshop on XML and Information Retrieval, Tampere, Finland, pp. 4–13. ACM Press, New York (2002)

8. Kade, A.M., Heuser, C.A.: Matching XML documents in highly dynamic applications. In: Proceeding of the eighth ACM symposium on Document engineering, Sao Paulo, Brazil, pp. 191–198 (2008) ISBN:978-1-60558-081-4
9. La Fontaine, R.: Merging XML files: A new approach providing intelligent merge of XML data sets. In: Proceedings of XML Europe 2002 (2002)
10. Lin, D.: An Information-Theoretic Definition of Similarity. In: Proceedings of the 15th International Conference on Machine Learning, pp. 296–304. Morgan Kaufmann Publishers Inc., San Francisco (1998)
11. McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill, New York (1983)
12. Nierman, A., Jagadish, H.V.: Evaluating structural similarity in XML documents. In: Proceedings of the Fifth International Workshop on the Web and Databases, WebDB 2002, pp. 61–66. University of California (2002)
13. Porter, M.F.: An algorithm for suffix stripping. Program 14(3), 130–137 (1980)
14. Princeton University Cognitive Science Laboratory. WordNet: a lexical database for the English language, http://wordnet.princeton.edu/
15. Resnik, P.: Semantic Similarity in a Taxonomy: An Information-based Measure and its Application to Problems of Ambiguity in Natural Language. Journal of Artificial Intelligence Research 11, 95–130 (1999)
16. Richardson, R., Smeaton, A.F.: Using wordnet in a knowledge-based approach to information retrieval. Technical Report CA-0395, School of Computer Applications, Trinity College, Dublin, Ireland (1995)
17. RSS Advisory Board. RSS 2.0 Specification, http://www.rssboard.org/
18. Tekli, J., Chbeir, R., Ytongnon, K.: A hybrid approach for xml similarity. In: van Leeuwen, J., Italiano, G.F., van der Hoek, W., Meinel, C., Sack, H., Plasil, F. (eds.) SOFSEM 2007. LNCS, vol. 4362, pp. 783–795. Springer, Heidelberg (2007)
19. Wu, Z., Palmer, M.: V Verbs semantics and lexical selection. In: Proceedings of the 32nd annual meeting on Association for Computational Linguistics, Morristown, NJ, USA, pp. 133–138. Association for Computational Linguistics (1994)
20. WWW Consortium. The Document Object Model, http://www.w3.org/DOM

An Efficient Simulator for Reactive Web Caching

Georgios Kastaniotis and Christos Douligeris

Abstract Simulation is a very powerful and widely-accepted tool in the research of reactive web caching. This paper first gives a brief overview of the functionality of reactive web caching and pays special attention to the creation of representative traffic patterns for effectively modeling the real workloads a caching application deals with. An upgraded version of the Wisconsin web cache simulator is presented as a case study. This upgraded version constitutes a light and intelligent testing environment of the largest created database of replacement strategies, as it is able to handle effectively even the most complex of the proposed strategies.

1 Introduction

The scientific research on reactive web caching is to the largest extent based on the well-known technique of trace-driven simulation. The majority of this research is targeted at the development, assessment and validation of innovative object replacement strategies [1]. Web caching architectures [2] and cache consistency [2] have been meticulously studied as well. The Wisconsin simulator [3] is a pioneer for the trace-driven simulation of the basic functionality of reactive web caching, specializing, in particular, in replacement strategies. Its intelligence stems from its simplicity - i.e. the ease of upgrading and updating - and lightness - i.e. the ability to hide the complicated parts of the transfer and network level and focus on the application level effects. Because of its intelligence and qualifications, the Wisconsin simulator constitutes a good solution to support the extensive evaluation of replacement strategies [4]. Unfortunately, the original version of the Wisconsin simulator was confined to simulate a small number of simple algorithms of mostly low complexity, under the relatively light web activity of the mid-90s. It is obvious that for today's complex environment the Wisconsin simulator presents some dissuasive flaws and deficiencies in its original design. This paper analyzes the architecture

Department of Informatics, University of Piraeus, Greece

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 65–75.
springerlink.com

and the functionality of an upgraded version of the Wisconsin web cache simulator. Before that, a brief overview of the basics of the trace-driven simulation technique on reactive web caching is presented. Concerning the layout of this paper, section 2 makes a brief introduction to reactive web caching. Section 3 focuses on issues that arise in the study of the collection and preprocessing of the empirical traces, and on the generation of synthetic traces, which are used to model the workload of the simulated environments. In section 4, the motivation behind the decision to upgrade the Wisconsin web cache simulator is put forward and the flaws of the original version are revealed, in an attempt to determine the specifications of the upgraded version. Section 5 analyzes the architecture of the upgraded Wisconsin simulator, in detail. Section 6 experimentally evaluates the efficiency of the simulator, which stems from its ability to handle effectively and efficiently even the most complex of the replacement strategies. Section 7 presents the related work and covers the most important simulators that have been developed to support research on reactive web caching. Finally, section 8 concludes this paper.

2 Reactive Web Caching

In this section, we present the basic functionality of reactive web caching, since the design of a simulation tool requires the acquisition of a thorough understanding and a deep knowledge about the system to be simulated. Figure 1 depicts the various steps of a transaction on the web, where a caching proxy stands between the client and the origin server. The dashed lines denote the internal functionality of a machine, while the solid ones represent the network functionality between the machines. In the most general scenario, the web client sends a request to the proxy server, which in turn, tries to serve it from its local cache. A likely inability to serve the request triggers the proxy to forward this request to the origin server, on behalf of the client. Finally, the proxy stores a copy of the requested object in its local cache and forwards the response back to the client. In the case of a full cache, a replacement strategy is activated during the insertion of a new object, in order to decide on a candidate object for eviction.

In complex architectures of caching proxies, the above scenario becomes more involved. In the special case of caching hierarchies, a cache miss at the low level leads the request to the caching proxy at the immediately higher level and so on until the request reaches the origin server. Caching architectures may also be more distributed. In this case, the proxies at each level cooperate with each other, by using communication protocols like ICP [5] and HTCP [6]. An interested reader on caching architectures can find a detailed analysis in [7]. Figure 2 shows a simple example of a two-level architecture of cooperating caches.

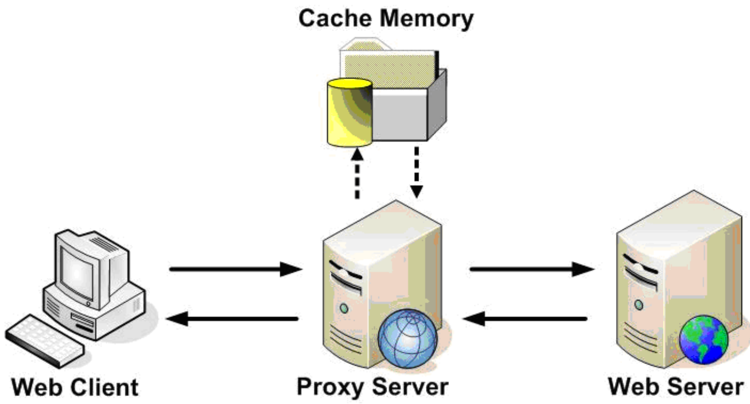


Fig. 1: The basics of the functionality of reactive Web caching.

3 Empirical vs. Synthetic Workloads

In order for the results of a simulation experiment to be considered valid and useful, there is a need to correctly represent the workload the caching application will face. For example, a caching application running on the client deals mostly with the requests from only one user towards servers over the Internet, while a proxy deals with requests from a larger group of users. There are two different approaches in rendering the input workload for a simulation experiment. The first approach is the collection and preprocessing of traces of real internet activity, derived from systems identical to the one that is to be simulated. The input files with the empirical traces are then formatted according to the specifications of the simulator. The second approach dives deeper into the preprocessing of the empirical workloads in an attempt to generate synthetic workloads. Files with synthetic traces of activity are mostly used for evaluating the impact of the specific features of the workload - e.g. the popularity and the size of web objects - on the caching application. Barford and Crovella [8] reported two methods in the generation of synthetic workloads. The analytic method comprises the characterization of the empirical workloads, leading to the appropriate statistics and the corresponding distribution models of their main features, and the development of special engines that incorporate these statistics and generate the synthetic workloads. The trace-based method refers to the direct alternation of the empirical workloads, through sampling, request swapping etc.

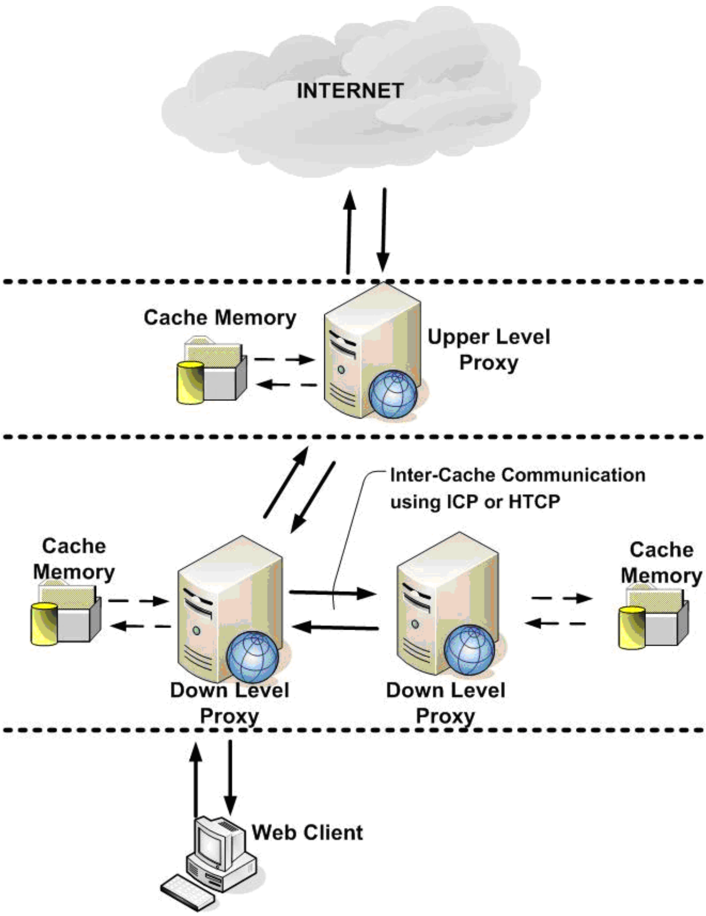


Fig. 2: Example of a two-level architecture of cooperating caches.

4 The Upgraded Wisconsin Web Cache Simulator - Motivation, Specifications and comparison with the original version

The reason behind the proposed upgrade of the Wisconsin simulator is to support the evaluation of the object replacement strategies that have been proposed over the last fifteen years [4]. Among our basic specifications on the choice of a simulator was its simplicity - i.e. the ease of upgrading and updating from a programmer's point of view, - its lightness - i.e. the ability to focus on the application level effects and hide the complicated parts of the transfer and network level - and of course its ability to simulate a variety of existing or to be proposed replacement strategies. All these requirements are met by the Wisconsin web cache simulator, and thus, the upgraded version inherits these properties from the original one. Unfortunately, the original version of the Wisconsin simulator was confined to simulate a small number of simple algorithms of mostly low complexity, under the relatively light web activity of the mid-90s. Thus, some dissuasive flaws and deficiencies in the old design can be observed, which formed the basis of the more detailed specifications of the proposed upgraded version. At first, during the encoding of new replacement strategies, we were faced with the deficiency in supporting the randomized strategies. The lack of a reliable mechanism to support the strategies of higher complexity and reduce the simulation time was one of the major shortcomings of the original version. Furthermore, the inability of the original simulator to provide the user with an interactive interface hindered the progress for the addition and the management of the new policies. Perhaps the most important drawback of the original version of the simulator was its inability to correctly prioritize the sequence of responses, with respect to the time series of the corresponding requests and the observed delay. This flaw led the simulator to a false perception of the cached content, which, in turn, resulted in misreports concerning cache hits and misses. A final drawback of the original version that needs to be addressed was the dependence of the performance metrics on the transient period, when the cache fills with objects. We found the incorporation of a warm-up mechanism quite handy in solving this problem.

5 Architecture of the Upgraded Wisconsin Simulator

Figure 2 pictures the detailed analysis of the architecture of the core of the simulator and the interaction between its main parts. The core consists of the following parts:

- *Control Unit*: it takes over, after invoking the simulator, and generates the other basic architectural parts of the system. Some of the tasks of the control unit are the execution of the infinite cache simulation instance for the input trace-file, he printing of the result into the output file, and the responding to the user input through an interactive interface that is analyzed below.
- *Trace Storage*: it keeps the input traces in the main memory of the system in order to avoid the time-consuming I/O calls to the trace-file.
- Random Number Generator: it is based upon the 'rand()' standard routine of the C language. This part of the simulator is utilized in the simulation of the randomized object replacement strategies [9].
- *Hash Table*: it is used to swiftly search the cache memory for an object.
- *Response Priority Queue*: it prioritizes the responses, according to the expected time of their entrance in the cache, with respect to the time series of the corresponding requests and the observed delay. This mechanism prevents the simulator from acquiring false information regarding the cached content, a situation that may result in misreports concerning cache hits and misses.
- *Data Structures for Cache Management*: the simulator uses three different data structures, according to the needs of the strategies: a simple queue, a multiple queue system and a priority queue.

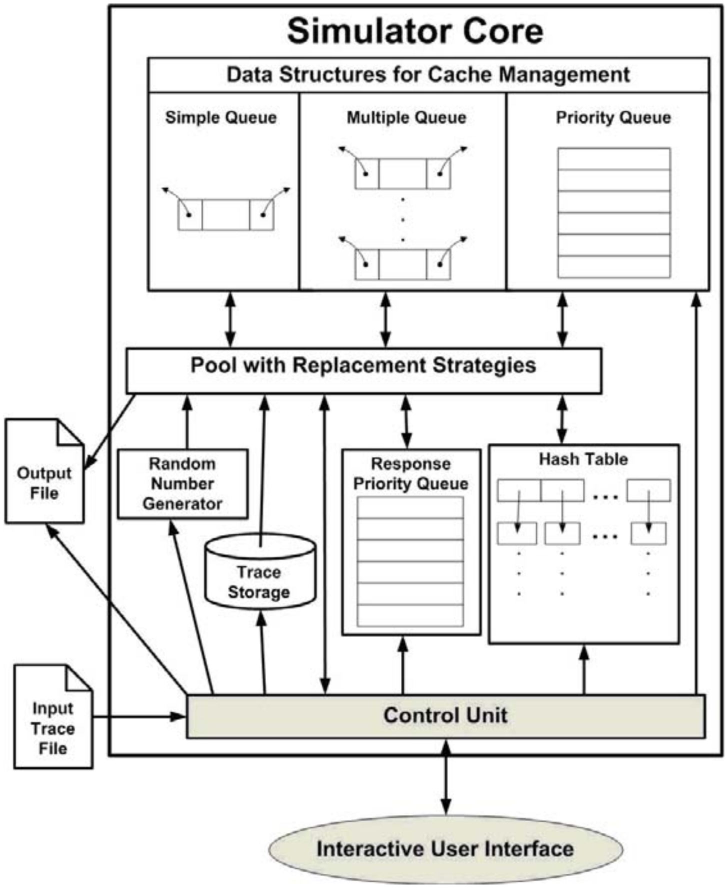


Fig. 3: The architecture of the core of the upgraded Wisconsin web cache simulator.

- *Pool with Replacement Strategies*: it provides the simulator with the instructions on how to simulate each replacement strategy.

One of our additions to the simulator is the interactive user interface. The interface gives the user the ability to isolate and focus upon a specific replacement strategy, which is particularly useful in the debugging of the simulator, while encoding new algorithms. The layout of the interface is presented in figure 3. At first, the interface offers the main menu, from which the user may choose to change the basic parameters of a simulation - e.g. the simulated cache size, - to navigate through the submenus that contain the replacement strategies, to simulate all the strategies and to exit the environment of the simulator. The submenus with the replacement strategies are organized in accordance to the classification model proposed in [9]. Thus, each submenu corresponds to one of the following classes of strategies: recency-based, frequency-based, recency/frequency-based, function-based and randomized.

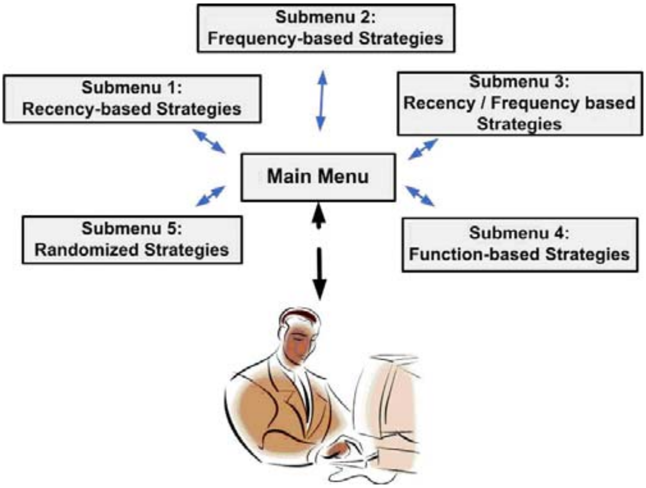


Fig. 4: The Layout of the interactive user interface.

The uniqueness of the upgraded version of the Wisconsin web cache simulator stems from its large database of supported replacement strategies. More than forty strategies and strategies variants are currently supported. The process of updating, which is still in progress, aims at accumulating all the proposed strategies in the bibliography, in order to make the simulator the ultimate testing environment for new and innovative ideas in the field. Table 1 presents the supported replacement strategies for each class, according to the classification model in [9]. The detailed analysis of each strategy is beyond the scope of this article. Thus, the interested user is advised to refer to [9] for a brief study on the supported strategies.

Table 1: Supported Replacement Strategies

Recency	FIFO, LRU, LRU-MIN, LRU-THOLD, SIZE, PSS, PSS-LATENCY, EXP1, PART-LRU
Frequency	LFU, LFU-AGING, A-AGING, LFU-DA
Recency/Frequency	Segmented LRU, LRU*, LRU-SP, LRU-SP-LATENCY
Function-based	LAT (LLF), HYBRID, Bolot/Hoschka, LRV, GDS, GDSF, GDF, GD*, g-GDSF, M-Metric, MIX, TSP, LUV
Randomized	RAND, LRU-S, LRU-C, LRU-CS, Rand-MIX

6 Evaluation of the Efficiency of the Simulator

All the high-complexity replacement strategies that are supported by the upgraded Wisconsin simulator owe their complexity to the creation and management of a priority queue during the eviction of an object from the cache. In order for the simulator to deal with these strategies, we have incorporated a low watermark parameter that aims to reduce the simulation time, by moderating the need for triggering the eviction process. The low watermark is expressed as a percentage of the cache size. During an eviction process, the cache keeps evicting objects until this low watermark is reached. It is sensible that the number of calls to the complex evictions process increases, as the low watermark gets closer to the cache size. The purpose of the experiment presented in this section is twofold. On the one hand, we assess the complexity of the supported strategies and on the other hand, we underline the usefulness of the low watermark in reducing the simulation time, especially for the most complex strategies that owe their complexity to the eviction process. In an attempt to meet the needs of this experiment, we used the ’clock()’ routine of the C language to estimate the time during which the CPU was busy with the simulation of a specific strategy. We used traces that were collected from the ’pa.us.ircache.net’ proxy, at Palo Alto, in California, between the 9^{th} and the 10^{th} of January 2007. The trace-file is composed of 407,166 unique web transactions. We also set the cache size to 636 MB. This value corresponded to the 15% of the infinite cache size for the specific trace-file. This simulation aims to compare the most complex of the supported strategies in theory. These are the LRU-MIN, MIX, M-METRIC and TSP strategies [9]. Figure 4a exhibits the measured time complexity against the low watermark, for each of the strategies above, under the Palo Alto traces. The time complexity was computed in seconds of CPU activity. We should note, in the figure, that it was logarithmically transformed for better visual results, and that the low watermark is expressed as a percentage of the cache size. MIX, M-METRIC and TRP are found to have the same behavior, which was expected because they follow the same eviction patterns. In all cases, for a watermark value close to 95% of the cache size, the simulation time drops substantially. On the other hand, LRU-MIN is found to be resource-demanding, but this behavior does not follow the increase in the value of the low watermark.

The eviction process in LRU-MIN is based on the repetitive traversal along the simple queue until the necessary number of objects is expelled from the cache. On the one hand, the increase in the value of the low watermark results in the reduction of the number of the evicted objects, and this can be further translated into fewer chances of many traversals along the queue. On the other hand, the number of candidate objects for eviction depends on the size of the input object. This means that the extremely big input objects - compared to the size of the in-cache ones - lead to an eviction process of high cost in resources, irrespectively of the value of the low watermark. We believe that the combination of these two contradictory situations explains the behavior of LRU-MIN in figure 4a. At this point, the question that arises is whether there is any effect of the low watermark on the basic performance metrics, especially as far as the evaluation process is concerned. Figure 4b presents

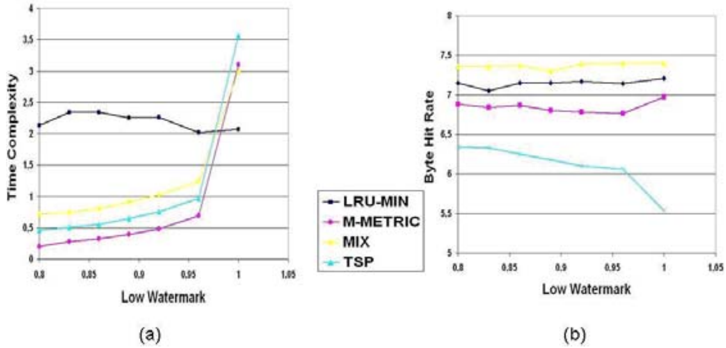


Fig. 5: Logarithmically transformed time complexity (a) and byte hit rate (b) against low watermark, for the LRU-MIN, M-METRIC, MIX and TSP strategies, under the Palo Alto traces.

the byte hit rate against the low watermark for the four policies, under the Palo Alto traces. The negligible impact of the low watermark on the performance evaluation of the strategies, in terms of byte hit rate is clear. A negligible impact on the hit rate was observed as well, but the corresponding visual depiction was impossible due to lack of space. Since the size of the trace is a crucial factor of utmost importance when measuring simulation time, the same experiment was repeated for larger traces and different replacement strategies. Once again, the results of these experiments verified the initial findings.

7 Related Work

A significant number of simulation tools have been designed and presented in the scientific literature. Each one has its own peculiarities concerning the degree of detail in emulating the reactive web caching behavior. DavisSim [10] was made from scratch in an attempt to improve over the standard functionality of the Wisconsin simulator. Another simulator of the same kind is WebCASE [11], which claimed to be a flexible and userfriendly tool that is able to closely monitor the evolution of the simulation experiments. The WebTraff software [12] includes a simulator of some of the salient replacement policies. Gonzalez-Canete et al. [13] have recently designed and implemented a windows-based simulator for object replacement strategies. PROXIM [14] and NCS [15] dive deeper into the protocol stack, in order to assess the impact of the special characteristics of the transport and network levels on the performance of the caching applications. As long as the simulation of caching architectures is concerned, Williamson [16] mentions the design and implementation of two simulators that are targeted at the study of load balancing and total performance. ProxyCizer [17] is a set of tools capable of emulating caching proxy architectures and testing real systems, with the ability to distinguish among multiple, different input patterns. In [18], the integration of ProxyCizer with the network simulator NS [19] is analyzed. The combination of both tools enables the assessing of the performance of the surrounding network as well. In [20], a simulator for distributed web caching systems is presented, which is able to run on systems with parallel processing. Finally, Cardenas et al. [21] designed and implemented a web cache simulator that is able to evaluate different replacement strategies and caching architectures, providing an innovative technique in assessing the cost of a cache-miss. The simulator was validated in comparison to the standard functionality of a real caching application.

8 Conclusions

This paper was dedicated to issues that arise during the development and running of simulators for reactive web caching. The fundamental parts of the functionality of reactive web caching, which must be considered by a prospective simulation tool, were presented. Having in mind the significance of the appropriate modeling of the workload, a caching application deals with, we also emphasized on matters about the collection and preparation of the empirical traces, and the generation of synthetic workloads. An upgraded version of the Wisconsin web cache simulation was presented and evaluated, which to the best of our knowledge comprises the testing environment for the largest database of object replacement strategies. A number of simulation experiments revealed the efficiency of the simulator, that is, its ability to handle even the most complex of the replacement algorithms, without affecting the performance evaluation procedure.

References

1. Selvakumar, S., Sahoo, S.K., Venkatasubramani, V.: Delay sensitive least frequently used algorithm for replacement in web caches. Elsevier Computer Communications 27(3), 322–326 (2004)
2. Pan, J., Hou, H.T., Li, B.: Retrieval and freshness thresholds in hierarchical caching systems. Computer Networks 44(2), 235–246 (2004)
3. The Wisconsin Web Cache Simulator. Original Version, `http://pages.cs.wisc.edu/cao/webcache-simulator.html`
4. Kastaniotis, G., Maragos, E., Dimitsas, V., Douligeris, C., Despotis, D.K.: Web proxy caching object replacement: Frontier analysis to discover the 'good-enough' algorithms. In: Proceedings of MASCOTS 2007, pp. 132–137 (2007)
5. Wessels, D., Claffy, K.: Internet Cache Protocol (ICP), version 2. RFC 2186 (1997)
6. Vixie, P., Wessels, D.: Hyper Text Caching Protocol (HTCP/0.0). RFC 2756 (2000)
7. Wang, J.: A survey of web caching schemes for the Internet. ACM SIGCOMM Computer Communication Review 29(5), 36–46 (1999)
8. Barford, P., Crovella, M.: Generating representative web workloads for network and server performance evaluation. ACM SIGMETRICS Performance Evaluation Review 26(1), 151–160 (1998)

9. Podlipnig, S., Boszormenyi, L.: A survey of web cache replacement strategies. ACM Computing Surveys 35(4), 374–398 (2003)
10. Barnes, J.F.: DavisSim: another web cache simulator, http://pdclab.cs.ucdavis.edu/qosweb/DavisSim.html
11. Zhang, J., Izmailov, R., Reininger, D., Ott, M.: WebCASE: A simulation environment for web caching study. In: Proceedings of the 4th International Web Caching Workshop (1999)
12. Markatchev, N., Williamson, C.: WebTraff: A GUI for web proxy cache workload modeling and analysis. In: Proceedings of MASCOTS 2002, pp. 356–363 (2002)
13. Gonzalez-Canete, F.J., Casilari, E., Trivino-Cabrera, A.: A windows based web cache simulator tool. In: Proceedings of SIMUTools 2008 (2008)
14. Caceres, R., Douglis, F., Feldmann, A., Glass, G., Rabinovich, M.: Web proxy caching: The devil is in the details. In: Proceedings of WISP 1998 (1998)
15. Davison, B.D.: NCS: Network and cache simulator - An introduction. Technical Report DCS-TR-444, Department of Computer Science, Rutgers University (2001)
16. Williamson, C.: Advanced Networking Applications (ANA) - Final report, http://www.cs.usask.ca/faculty/carey/projects/finalreport.html
17. Gadde, S.: ProxyCizer, http://www.cs.duke.edu/ari/cisi/Proxycizer/
18. Sosa, V.J., Navarro, L.: A new environment for web caching and replication study (2001), http://personals.ac.upc.edu/leandro/pubs/jcc2001.pdf
19. Network Simulator (NS), http://www.isi.edu/nsnam/ns/
20. Brandao, R.F., de Oliveira Anido, R.: A parallel simulator for distributed and cooperative web caches. In: Proceedings of DS-RT 2001, pp. 113–120 (2001)
21. Cardenas, L.G., Gil, J.A., Sahuquillo, J., Pont, A.: Emulating web cache replacement algorithms versus real system. In: Proceedings of ISCC 2005, pp. 891–897 (2005)

Constructing World Abstractions for Natural Language in Virtual 3D Environments

Carlos León, Salvador de la Puente, Daniel Dionne, Raquel Hervás, and Pablo Gervás

Abstract Human-Computer interaction can be clearly leveraged by using Natural Language systems, specially in interactive environments. In this paper a system which interprets a 3D virtual world in terms of human abstractions like “corridor” or “crossroad” and differences between world items basing on the properties the user refers to, is presented. This interpretation is oriented to human orders understanding, making it possible for the system to be interactively guided by the user. A partial implementation of the algorithm and its results are shown, and discussion about how the system could be empowered for better Natural Language Processing is detailed.

1 Introduction

Human-Computer interaction is known to be a challenging field in Computer Science, and there exist several approaches for leveraging the interactive experience when a human user communicates with a machine. Among them, Natural Language Processing is perhaps one of the most interesting approaches because most humans use natural language for ordinary communication.

Usually, interactive systems which implement Natural Language Understanding receive user input which is not explicitly represented in the computer database. For instance, the computer could be showing the user a 3D world internally represented with *tiles*, and the user could ask for information about a *corner* he sees. In this situation, even if the corner really exists from the user’s *point of view*, the computer would not understand the order.

Carlos León, Salvador de la Puente, Daniel Dionne, Raquel Hervás, and Pablo Gervás
Departamento de Inteligencia Artificial e Ingeniería del Software, Universidad Complutense de Madrid, Spain e-mail: {cleon,raquelhb}@fdi.ucm.es, e-mail: {neo.salvador,dionnegonzalez}@gmail.com, e-mail: pgervas@sip.ucm.es

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 77–86.
springerlink.com

Alternatively, if the system has to describe a certain location to the user, or to give him a particular instruction, it may be restricted to using the conceptual vocabulary that is explicitly represented within the reference 3D world. For instance, the system may describe the location or express the instruction in terms of the tiles it covers or the sequence of tiles that has to be traversed. However, for the user, it would be much more intuitive to talk about abstract concepts such as corridors, doors, corners, or rooms.

It is possible to encode a hard-wired world representation by which all possible definitions of every part of the world are already included in the representation. Doing this by hand is very effort consuming and it may not be valid in worlds that change dynamically over time. For these environments, a dynamic, computerized system can help the machine to understand a larger set of commands or dialog parts, or to be able to formulate instructions or descriptions in human abstractions, and either way create a better user experience.

For the communication to be more human-like, thus improving the quality of the system, it is necessary to recognize input messages expressed in abstract terms, or to formulate output messages in correspondingly abstract terms. In this paper a system for Natural Language abstractions processing in virtual 3D environments is shown. The user has visual 3D information, and the system has explicit knowledge about the world in terms of low level graphical representation (tiles, objects, valid positions). The system is designed to follow a set of commands like moving the camera representing the user or focusing on something interesting, and/or to guide the user through the 3D environment by providing instructions. Orders from the user and descriptions from the system could possibly include high level abstractions (from the low level graphical information point of view) like *door*, *corridor* or *corner* as the user sees them. Correctly understanding these orders involves an abstraction process in which the 3D graphical world representation is *interpreted* in a human-like manner. The system should also be capable of *formulating* its instructions in abstract terms along the same lines.

Figure 1 shows a comparison between an architecture with direct Natural Language Recognition (top) and an architecture with the presented system in the design. The main benefit of including this system for order recognition consists on the fact that some properties which are expressed by a human user are not explicitly present on the virtual world, and thus they can not be identified by the simpler version. Basically, the main addition to the system is the *extension* module, which receives a world representation and expands it to automatically contain more information, following algorithms like that shown in Figure 1.

2 Previous Work

Natural Language Understanding has been a research topic for many years, starting with the very beginnings of AI. Since then, it has gone to several stages, full of high hopes at the start, followed by disillusions when problems proved much more

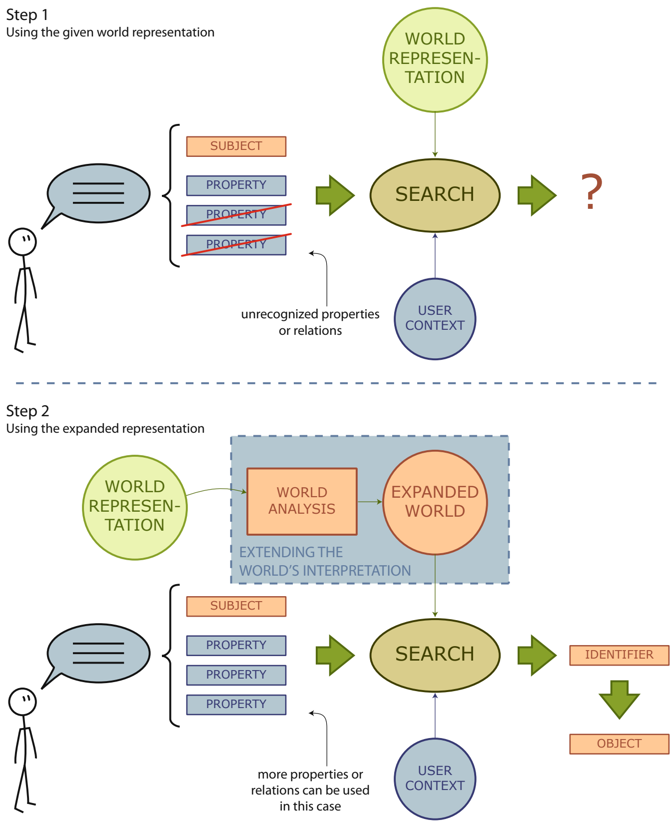


Fig. 1: Architecture comparison between a simple order recognizer (top) and the Natural Language Understanding in Virtual Environments System (bottom).

difficult than expected. The field has reappeared as a dignified research topic in recent times, partly as a result of the improved results in parsing achieved by statistical parsers, and the surge of new solutions arising from the application of machine learning techniques to the area. In general terms, this transition in practice pushed the state of the art past the original obstacles related to parsing, to be stopped again by problems related to semantics.

In this context, natural language interaction in virtual environments provides an interesting case study because the 3D representation of the world that is being used to graphically render the environment can double up as an explicit representation of the semantics of the world.

Solving Natural Language Understanding in Virtual Environments problems has been tackled from several different points of view. In [8], how people refer to similar objects using disambiguating characteristics, and how this can be applied to a recognizing system is studied. Other Natural Language systems are applied robotics, [12], where spatial utterances recognition are solved by using state information gathered by the robot's sensors.

For referencing elements in Natural Language in a 3D environment (whether it is real or virtual) several projects try to add logic information about certain relations between items in the space. Creating a logic description for *near* or *by* usually helps systems to better understand human utterances about 3D spaces [11]. Referring expression generation studies not directly related with virtual environments also apply in this field, as many Natural Language Generation issues are general enough for including this kind of systems [7, 10].

Ontologies play a significant role in this field. Formally describing what is present in a virtual world is usually needed, and related work can be found in the field of spatial ontologies [9, 4, 3, 5]. Most of them try to create a reasoning system based on ontologies which are not necessarily constrained by three dimensions.

Logical spatial reasoning shown in [6] presents a logical framework for identifying spatial properties of objects (or between objects). [2] presents basic spatial relations between object which set the base for spatial logical reasoning.

All this systems are related to Natural Language and interactive environments in some sense. However, they assume a fixed, explicit or implicit representation of objects and a "labeling" for them. Next sections explain an approach for identifying these labels for a subset of 3D constructions which could help these systems to understand new different parts of a 3D space.

3 Constructing World Abstractions from Graphical 3D Information

This section shows a system which exchanges instructions in Natural Language with a user, and either identifies correctly what the user wants to say when he is referring to complex terms like *corner* or *wide corridor*, or is capable of using such terms when giving instructions to the user. The system relies on ground graphical 3D information from the world and the current user state to correctly compute what parts of the world the user wants to interact with, or how to refer to the parts of the world that the system has to mention.

3.1 Internal World Representation

The system is designed to handle a very low level world representation based on a 2D tile grid, which can be considered to be unbound (although real computational requirements will demand the opposite). This space discretizes space into four directions (North, South, East and West) that the user can face (Figure 2).

On this 2D grid, there are defined a set of walls whose layout can be considered to form rooms and corridors (and thus corners, doors, and so on). Discrete items (like lamps or chair, for instance), are also laid out all along the virtual world. These items have properties which difference them from other items of the same class in the world. The representation also includes a set of rules which precise how the world dynamically changes when the user performs some actions: *the user pushes a button → a wall moves, uncovering a new corridor*. Figure 2 shows a schematic diagram representing the virtual world and its elements.

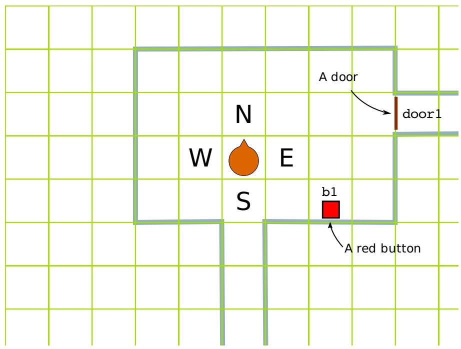


Fig. 2: Walls, items like the door and the button, and properties like the red colour of the button.

Two basic actions for this world are defined: *movement* and *interaction*. The user can **move tile by tile** forward, backward, to the left and to the right. He can **turn left or right** too, facing each of the four discretized orientations explained before. The user can **focus on an world element** and interact with it. Interactions include pushing buttons and taking objects.

Note the world does not include any information related to different rooms, corridors, corner and other human space concepts like these. All this information must be inferred from the basic world representation.

3.2 Abstracting 3D Information in Understanding

When the system receives an order in Natural Language, and after a simple pre-processing stage where the order and the parameters are identified, the recognition stage starts.

The goal is to match human-like orders including basic actions, which will be interpreted using the world information. The system pre-processes complex orders by simply splitting them into simple ones, which will be issued to the recognizing algorithm one by one. The first stage, **instruction matching**, applies simple first-order logic like templates to the received order and fills some slots for it to be useful. For instance, an instanced template could have the following form: $go(corner)$ which will be interpreted by the system as the desire of the user of going to some corner.

As said before, there are world elements which are not explicitly declared as world structures. However, the user could refer to them because a human can recognize these elements if they are inside his visual field. The system uses some predefined **keywords** from the instanced action template to trigger certain types of **search agents**. This agents are just simple threads which look for items (lamps, for instance) or world parts (corners) which match some properties present in the user's order.

Different types of search agents return different sets of matching items. These agents could find items (buttons, red things, lamps...) which are simple, referenced things in the world whose properties are used to disambiguate between them, and world parts like corridors, corners, and so on, which are physical abstractions which have to be inferred from the world layout.

For instance, consider the next sentence: "*Go to the left corner*". The word *"corner"* will trigger a spatial search agent to identify possible referred corners and the word "*left*" will trigger another spatial search agent discerning between elements at left of right from the user's position and orientation.

Consider another instruction: "*Push the red button*". The word "*red*" will trigger a property-based search agent to discover red elements. The "*button*" word triggers a button-oriented search agent which will find all buttons inside the user visual field.

Several geometrical search agents for rooms, corridors and corners identifying have been developed. Algorithm 1 shows as an example the basic procedure for rooms. When the user refers to a "room" and a search agent finds a room in his visual field, for instance, the algorithm maps that reference ("room") with the set of tiles, and then this information can be used for the order execution: once the algorithm knows that the user is referring to the subset of tiles which are enclosed inside a square bounded by the coordinates (x_i, y_i), (x_j, y_j), the system can move the camera to the center tile inside that room.

A small trace in a square room with a small corridor could be like this: first, the cursor is set in any valid location and orientation: $(2,2)$, facing positive y. Then, it advances until $(2,4)$. A new room is added, and the recognition starts: the cursor turns right to $(4,4)$. The to $(4,0)$, then to $(0,0)$, and then to $(0,4)$. Now, the cursor finds a gap in $(1,4)$ which end in $(2,4)$, the starting point. With this process, a new room has been identified.

Once every search agent is done with its search, they all have a set of matching elements: a "button" agent has a set of buttons, a "red" agent has a set of red items. The intersection of these set lets the system obtain a new set which will contain zero, one or more references to items or parts of the world. In case this intersection contains only one element, the matching process has succeeded and there is only one possible item the user is referencing to. If the result are zero or more than one items, the system will have to ask the user for a more detailed order.

Search agents can be applied following no particular order. A successful set of agents always come up with just one element but the order affects performance. If the set of agents does not find one element only there is an **ambiguity** and the

Algorithm 1 Looking for rooms.

```
cursor ← user position
orientation ← user orientation
advance cursor facing orientation until it finds a wall
start location ← wall location
turn right
add new unexplored room to the room list
while there are unexplored rooms in the room list do
    while cursor is not at start location do
        advance cursor facing orientation until it finds a corner
        if cursor comes up with a concave corner then
            turn right
        end if
        if cursor comes up with a convex corner then
            if there is a door at cursor location then
                add a door at cursor location
                add new unexplored room to the room list
            else
                turn left
            end if
        end if
    end while
    add discovered border to the room
end while
```

system has to give feedback to the user for him to specify what he wants in more detail.

Let us say that a Natural Language processor receives an order: "go to the corner near the window". After processing, it outputs the recognized structure, in the form: $go(corner), isnear(corner, window)$. Several search agents are created. One of them, following a process analogous to the one shown in Algorithm 1, tries to find a corner inside the user's visual field. Another one looks for a window. If the search agent finds two corners, following the expression, the system will identify the corner which is nearer the window as the referred corner by intersecting the two returned sets from the search agents.

The system stores a set of rules for interpreting the valid set of verbs and references. Meanings for predicates like *go* and *isnear* are hard-coded in the interpretation system, so no previous processing is needed. Although this is far from the best possible alternative, it has been proved valid for this prototype. Improving it is a main part of the future work.

Applying the algorithm in a real system could produce, for instance, if the camera represented the user position in the virtual environment (imagine a first person system), camera movement towards the corner. A planning system, in this case, would also be needed, but this functionality is beyond the scope of this paper.

3.3 Abstracting 3D Information in Generation

These algorithms for world parts and item identification can also be applied to Interactive Systems which need Natural Language Generation where, for instance, a planning algorithm must show its solution in terms of high level Natural Language.

For this to be possible, the orientation of the system must be inverted. Instead of having an order to be executed, the system can have a plan to be explained to some user. This plan could be represented as a list of steps the user must follow in order to reach some point (for instance, when guiding a person the system could know where and how to go, and it could have to explain it to that person in Natural Language). Each step should be realized without ambiguities. This option is carried out by substituting search agents with referring agents, which receive a low level order (like *push button number 5*) and has to perform a search in the user visual field and consider all possible objects to correctly identify the desired button. If the system detects three buttons in the user's visual field and they all have different colours, the system could choose to say "push the green button". The algorithm is complementary to the previous one: each referring agent looks for a property or item ("button" or "red"). Then, from these sets, the system chooses the definition that uniquely defines the desired item.

This process, however, is not necessarily an inverse mapping from the understanding process. Several other issues must be taken into account. For instance, the plan or the task the user is to do, should be clearly, briefly explained in the beggining for the user to have a general idea of the task. Also, a guiding system like this one must consider the possiblity of the user being lost, and it has to create a solution which has to be explained to the user in terms of the abstracted world.

For demonstration of this being possible, the abstraction module has been implemented and tested in the GIVE Challenge (*Giving Instructions in Virtual Environments*) [1]. In this challenge, the main objective is to guide the user in a virtual interactive changing 3D environment. In this virtual world, the system only receives low level graphical information, and it is necessary to feed the user with high level orders, pretty much the inverse to understanding high level user orders. This challenge has clearly demonstrated that correctly identifying graphical information and creating abstractions for communicating with human users leverages the interactive experience.

4 Discussion

The presented system tries to fill a small but important gap in the field of interactive systems which communicate with the user using Natural Language. The presented algorithm has been created to add new functionality to Natural Language systems, specially those related with dialogs and orders in dynamic, interactive environments.

Abstracting information in real-time is a must for dynamic virtual environments. Offline world reconstruction for finding abstractions is only useful in static worlds, where everything remains the same independently of the user action. For a Natural

Language recognition system to be fully deployable in an interactive environment, online capabilities are a hard requirement. However, offline addition of Natural Language information can still be useful in static systems (or even partially useful in dynamic systems). However, creating this virtual information by hand is tedious and time consuming. Using the algorithm for this purpose automates the process and can be applied in many different domains: equivalent low level representation is present in many different interactive systems, and Natural Language Understanding can be used for many tasks.

The whole paper references *graphical information* as the internal representation of the 3D virtual world. As show in Section 3.1, this information is not purely graphical. The system is handling *tiles* and not just plain points in a 3D space. Although many virtual environments can be accessed assuming this level of information can be accessed, for a system to fully understand Natural Language sentences in a wide set of environments, this assumption can not be always made. Further research about how connecting low level graphical algorithms with this recognition systems can be very interesting.

In case of a uninterpretable order such "do something with that thing", the system discards it considering it has found an **instruction ambiguity**. The system, then, must give feedback to the user, asking him for a more detailed explanation about the action he is really referring. When the action is identified we can move to the next stage. It is however possible that such a sentence is not lacking meaning if the user just used some item (which could be mapped to *that thing*) and that this item only permits one operation (thus *something* could be mapped to that operation). This level of Natural Language Understanding is planned as future work.

The agent oriented design that has been followed during the development of the system tries to create a framework that scales well. Agents must implement a simplistic interface, which is basically composed by functions receiving the world state and return new high level elements. Thus, the scalability complexity is dependent on the high level concept (it might be more difficult to identify a round door than a square window) and not on the architecture itself. However, this has not been exhaustively tested and it is planned as further study.

It is important to note that the current implementation of the algorithm can recognize many typical 3D patterns, but not every possible layout which could be referenced by humans. This first approach has been proved to be useful in certain environments which are not extremely complex. For instance, the algorithm would not recognize sinuosity in a corridor. Studying different environment constructions is a fundamental part of the future work.

5 Conclusions and Future Work

A system which generates world abstractions (in the form of *corridor* or *corner*) from basic 3D information (like *there is a wall in (x, y)* has been detailed. The system is conceived as a part for Natural Language Interaction systems which receive orders in virtual interactive 3D environments.

Several future research lines are being developed to empower the system. Linking the system output with a spatial ontology for better understanding of concepts expressed by the user could leverage the success rate, thus creating a more useful system. Also, a better Natural Language Processing system can be plugged into the system in order to obtaining a more complex, richer interactive system. Identifying 3D constructions which are referenced by human users is fundamental to build a reliable and power system. Current ongoing research on this is on process. It would be very interesting to carry out a deep study about what specific parts of the world humans refer to when interacting in a virtual 3D environment.

References

1. Generating Instructions in Virtual Environments (GIVE) (2009), http://homepages.inf.ed.ac.uk/v1akolle/proj/give/
2. Clarke, B.L.: A calculus of individuals based on connection. Notre Dame Journal of Formal Logic (1981)
3. Coenen, F., Beattie, B., Bench-capon, T.J.M., Shave, M.J.R.: An ontology for linear spatial reasoning, pp. 718–727. Springer, Heidelberg (1996)
4. Coenen, F., Visser, P.: A core ontology for spatial reasoning
5. Coenen, F., Visser, P.: A generic ontology for spatial reasoning. In: Research and Development in Expert Systems XV, Proc. of ES 1998, pp. 44–57. Springer, Heidelberg (1998)
6. Cohn, A.G., Bennett, B., Gooday, J., Gotts, N.M.: Representing and reasoning with qualitative spatial relations about regions
7. Dale, R., Reiter, E.: Computational interpretations of the gricean maxims in the generation of referring expressions. Cognitive Science: A Multidisciplinary Journal 19(2), 233–263 (1995)
8. Gorniak, P.: A visually grounded natural language interface for reference to spatial scenes. In: Proceedings of the International Conference for Multimodal Interfaces, pp. 219–226. ACM Press, New York (2003)
9. Grütter, R., Bauer-Messmer, B., Hägeli, M.: Extending an ontology-based search with a formalism for spatial reasoning. In: SAC 2008: Proceedings of the 2008 ACM symposium on Applied computing, pp. 2266–2270. ACM, New York (2008)
10. Piwek, P.: Meaning and Dialogue Coherence: A Proof-theoretic Investigation. J. of Logic, Lang. and Inf. 16(4), 403–421 (2007)
11. Reddy, V., Neumeier, K., McFarlane, J., Cothren, J., Thompson, C.W.: Extending a natural language interface with geospatial queries. IEEE Internet Computing 11(6), 82–85 (2007)
12. Stopp, E., Gapp, K.P., Herzog, G., Laengle, T., Lueth, T.C.: Utilizing spatial relations for natural language access to an autonomous mobile robot. In: Dreschler-Fischer, L., Nebel, B. (eds.) KI 1994. LNCS, vol. 861, pp. 39–50. Springer, Heidelberg (1994)

Standardized Mobile Multimedia Query Composer

Mario Döller, Tobias Mayer, Kim Lim Fong, Stefan Beck, Harald Kosch,
David Coquil

Abstract This paper presents an extension to inViPla, a video search framework designed for operation on Personal Digital Assistants (PDA). The internal query system of the original framework has been replaced with a new query engine which generates multimedia queries using the MPEG Query Format (MPQF) standard to achieve platform independence. To formulate such queries the Mobile Multimedia Composer has been developed for mobile devices. Its touchscreen-friendly user interface provides extensive search functionalities including keyword search, technical parameters and semantic concepts. Additionally, the system features an editor which lets users formulate complex multimedia queries while offering a visualization in the form of a tree structure.

1 Introduction

The first ever mobile phone, was the Motorola DynaTAC 8000x (see Figure 1). Measuring 25 cm (not including antenna) and weighing 793g, there was never any hope of carrying this phone in the pocket.

Today, the Apple iPhone[1] is set to redefine the term "mobile phone" as the DynaTAC once did. Boasting a multi-touch user interface, almost all the iPhone's functions can be used without having to utilise a stylus or keypad. Some may actually use the iPhone for making calls, but its other features such as fast mobile internet, GPS navigation and multimedia capabilities certainly explain its stellar sales figures. The iPhone may be smaller, but is packed with more features and a battery which is capable of up to 6 hours of web browsing, 7 hours of video playback, or 10 hours of talk time. The above comparison shows clearly that mobile phones of today are

Department of Distributed Information Systems
University of Passau
Germany

[1] http://www.apple.com/iphone/

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 87–101.
springerlink.com

shifting away from their traditional role as communication devices to all-in-one personal content devices with internet capabilities. Network connection is an important feature as it enables users to watch videos streamed from the internet and also to produce and upload their own content to the World Wide Web. These new mobile phones using third generation (3G) technology and standards can achieve speeds of up to 14.4Mbit/s (download) and 11.5Mbit/s (upload)[2], making it possible to watch and upload videos comfortably. According to Nielsen Mobile in the US [5], mobile internet has become a mass medium. As of may 2008, there were 91 million subscribers who owned a video-capable phone with 13,9 million (6% of U.S mobile subscribers) paying for mobile videos services. Given these numbers, it is obvious that the amount of video content available on the internet today is enormous. The market potential is the issue that puts the spotlight on mobile video search applications. Current hardware is more than adequate to run multimedia applications smoothly on mobile devices, but the sheer volume of multimedia content prevents users from being able to find content that interests them effectively. A second more subtle problem also comes into play when one considers the nature of the mobile devices. Most mobile phones are geared towards the use of proprietary multimedia solutions as an iPhone needs the iTunes software to organize multimedia content.

2 Related Work

With the increasing number of video-capable mobile devices the interest in accessibilty of digital media grows as well. The following examples outline the extensive work performed in trying to make digital videos accessible on small mobile devices.

Fig. 1: A Motorola DynaTAC 8000x (l) and an Apple iPhone (r). (Source: tuaw.com)

[2] Theoretical maximum speeds for High-Speed Down/Uplink Packet Access protocols.

2.1 Research projects

There are numerous scientific publications available dealing with mobile multimedia retrieval and browsing (e.g. [9] [10] [19] [15] [11]). Two of them are shortly presented here. These proposals also cover mobile querying and presenting of videos.

One work introduces the Candela platform, a personal video retrieval and browsing for mobile users [15]. As "*an infrastructure that allows the creation, storage and retrieval of home videos with special consideration of mobile terminals*" it covers many parts of the lifecycle of digital videos. Metadata are available for all videos and represented by means of MPEG-7 media description schemes. Low-level (e.g. color histograms) as well as mid-level metadata (e.g. structural data like video segments) are automatically generated. High-level metadata describe the semantic content and are partly provided using personal ontologies, annotated by the user itself. As user interface an Apache COCOON-based[3] web application is available for both, mobile and stationary (PC-) users. To improve the usability to mobile users the layout is splitted into three windows. The web interface lets users browse through ontology catalogues and search videos. Summarizing all these features, Candela focuses on the infrastructure as a whole and in particular with regard to finding videos it provides only poor search functionalities (fulltext keyword search). Another work of Kamvar et al. [11] especially targets specially at the user interface. The MiniMedia Surfer is presented, a software for searching and browsing video segments of interest on small displays. The basic interface principle are transparent layers to support browsing tasks like (keyword) queries, browsing through keyframes of search results and video playback. Seperate layers shall clearly present information while the (adjustable) transparency creates a connection between them.

2.2 Commercial approaches

2.2.1 YouTube for Mobile

YouTube is a website that needs no introduction. It has been available to mobile users since June 2007. Last year, the mobile website[4] was completely revamped to allow access via 3G or WiFi to most of YouTube's regular video catalog, which contains tens of millions of videos. Another option, which is currently still under beta testing, is the YouTube for Mobile Java application, which offers YouTube with a user interface designed for mobile devices without relying on a mobile web browser (see Figure 2).

To accommodate mobile devices, YouTube for Mobile streams videos in 3GP format using the H.263/AMR [6] codecs instead of using the Flash video format. Most mobile devices do not have Adobe Flash installed meaning they cannot play

[3] http://cocoon.apache.org
[4] http://m.youtube.com

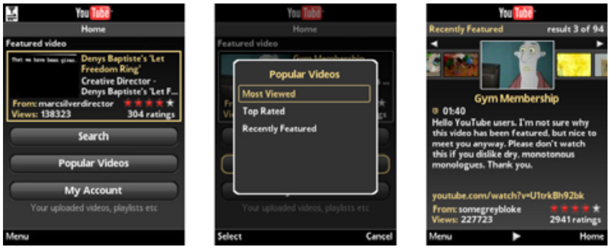


Fig. 2: (from left) home screen, popular videos menu, recently featured (Java app.). (Source: allaboutsymbian.com)

the videos from the desktop version of YouTube without prior conversion. This conversion problem explains why YouTube for Mobile only offers a subset of the videos available on the main website.

2.2.2 veveo vTap, Avot mV - Mobile Content Publishing

A new concept designed for mobile devices is the mobile content publishing. Instead of relying directly on user generated content, these solutions aggregate videos from sites such as YouTube, MetaCafe and GoogleVideo and allow users to search through this content, while also allowing content owners to publish more content on their networks. Two such solutions are mV and vTap, which have been available for beta testing since late 2007. Similar to YouTube for Mobile, these applications are available in the form of websites designed for mobile users and client applications that can be installed directly on the mobile device. mV is geared towards smartphones and PDAs, and the official website currently lists over a hundred supported devices, while vTap already has a Java ME client, a native implementation for newer Windows Mobile devices and an AJAX application for the iPhone, making it available on a wider variety of mobile devices. Both solutions offer an incremental character-based search, which means that results are displayed as the user types (on the mobile websites) and that relevant results can be offered before the whole keyword has been entered (e.g. type in ”beet” and results on Beethoven will be displayed). This filtering takes place on a remote server and all results are relayed back to the mobile device in real time, meaning that a relatively fast internet connection is a must. In vTap content is indexed primarily using the metadata found on the web pages on which the content appears and is then compared with relevant metadata from other sources. Users can choose from a general list of categories or add specific concepts from a searchable list of more than three million keywords generated by vTap during its content indexing process.

3 The MPEG Query Format

The MPEG Query Format (MPQF) [1] [8] became international standard early 2009. It specifies a format for the interaction of multimedia clients and multimedia retrieval systems (MMRS). In detail, the standard defines the message format for multimedia requests (e.g. Query by Example or Query By Text) to heterogeneous MMRS and the message format for their responses. Furthermore, a management part provides features such as service discovery (service is a synonym for MMRS) and service capability description. Figure 3 shows the individual message types and a possible usage scenario, in which a content provider undertakes the distribution and aggregation of MPQF messages to individual MMRS.

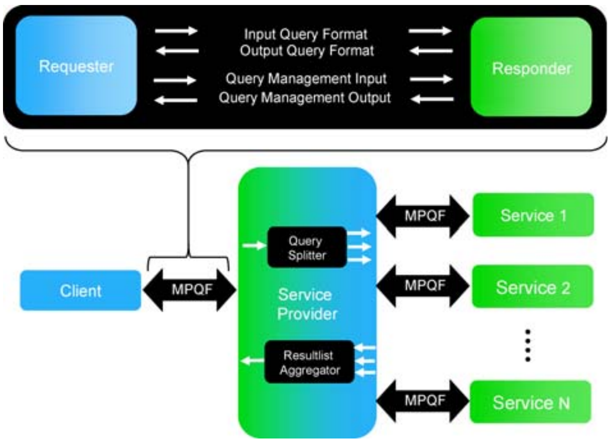


Fig. 3: Possible scenario for the use of the MPEG Query Format.

MPQF came from the MPEG-7 activities but it is important to note that it is not tied to MPEG-7 at all. In fact any XML based metadata format can be integrated. This versatility allows groups of users to define there own metadata sets (e.g. as folksonomies and informal tags) while retaining interoperability in the vital search interface. However, research is necessary to investigate how to distribute and aggregate requests that are addressed to heterogenous MMRS (e.g. a MPEG-7 MMRS and a Dublin Core MMRS).

4 System Architecture

The contribution for approaching a standardized MMRS described in this work is the Mobile Multimedia Composer communicating with the inViPla video search framework. This framework with its client/server architecture has been extended by

MPQF functionalities. At client side the Mobile Query Composer is responsible for generating MPQF queries being sent to the server as well as for handling MPQF query results.

Its architecture supports the retrieval in video collections. that have been annotated by MPEG-7 [12]. The application domain of our test data, its annotation and its ontology have been tailored to soccer movies. However, the concepts and tools are open and adaptable for any topic. An overview of our current overall system architecture is presented in Figure 4.

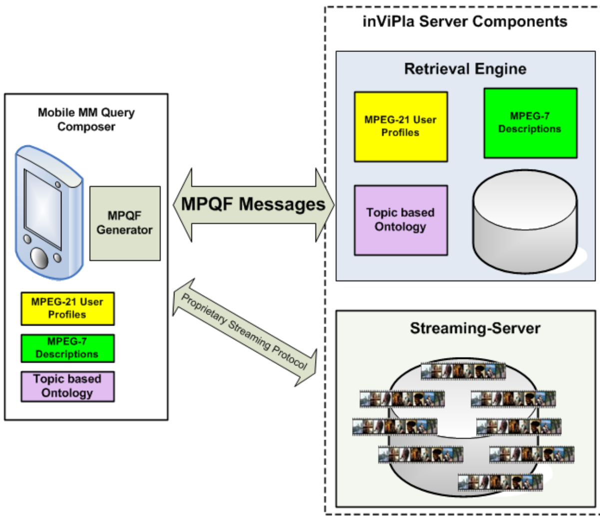


Fig. 4: System Architecture

Thanks to the use of the MPQF standard the individual components can be exchanged and developed independently. Furthermore, an additional abstraction layer can be introduced by the integration of MPQF aware middleware layers such as described in [7]. In the following, the individual components are introduced whereas the main focus of this paper is on the description of the mobile device targeting on an MPQF aware multimedia query composer.

4.1 inViPla Server Components

The system architecture of inViPla, shown in Figure 4, consist of two major components, the retrieval engine and a streaming server. These are seperate applications, thus they can be deployed on different machines. The streaming server contains all videos in a database with streaming accessibility through RTSP [16]. The retrieval engine provides search functionalities for those videos via MPQF messages. Basing on metadata descriptions (MPEG-7) and a topic based ontology, common queries (full-text search, criteria mapping etc.) as well as semantic ones are possible. Furthermore the retrieval engine contains a component for handling user profiles to realize personalized queries.

4.2 Mobile Query Composer

The Mobile Query Composer features a user interface to create and/or edit queries, which afterwards are formulated as MPQF requests. These are sent to the inViPla server that performs the actual search and delivers the results in MPQF format as well, containing a subset of MPEG-7 descriptions. A user can load a profile from the server for more personalized queries. Therefore, MPEG-21 user profiles [3] are used that contain information about digital item history, user rights, personal information like needs and preferences (language, video length, codec etc.). The usage of standards like MPEG-7 and MPEG-21 is a common approach to personalize mobile multimedia queries that can be found in many works (e.g. [2] [17] [18]).

Special attention was devoted to the ease of use. One example of following the goal of usability are view patterns. They are used to reach similar layouts, and thus to avoid confusing the user. All pages are clear as well as self-explaining and a global navigation bar (see Figure 5) is visible at any time. The different pages and their functionalities to compose a query are now described in more detail.

5 Mobile Search Paradigm

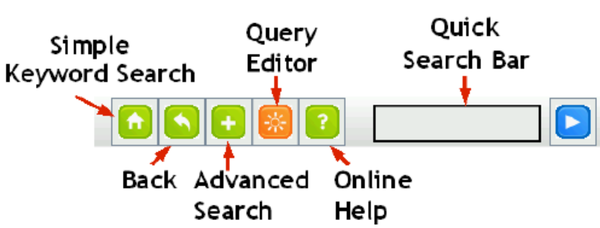


Fig. 5: Toolbar with description of the buttons.

5.1 Keyword Search

The keyword search is the most basic of the search functions. It allows the user to type in keywords to search for. Users can search using multiple keywords or phrases. The last 10 searches for the current session are recorded in a list below the text field. For added convenience, this view can be reached by pressing the home button in the toolbar (see Figure 5) and keywords entered in the toolbar's quick search field will also be recorded in the history for easy access.

5.2 Advanced Keyword Search

The advanced search view (see Figure 6) consists of the advanced keyword search, the technical search and the semantic search pages. By choosing the corresponding option in the combo box, users can switch between these three pages. Like the simple keyword search, the advanced keyword search is also aimed at casual users and the fields are labelled clearly to illustrate their purpose. Using the boolean operators from MPQF, the user can now formulate more interesting queries by choosing to include/exclude certain keywords. The text fields work as boolean AND, OR and NOT operators respectively. One can choose to use the text fields individually or combine them to perform more complex searches.

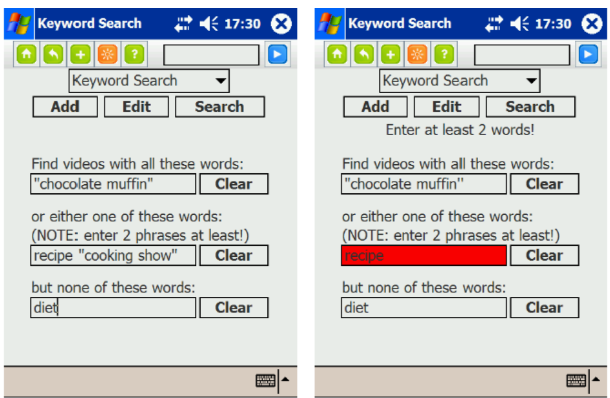


Fig. 6: Advanced keyword search with correct (l) and incorrect (r) input.

5.3 Technical Search

Due to space constraints of the display, the technical search (see Figure 7) has been reduced to four parameters that have been deemed to be most useful in searching videos: format, resolution, duration and file size. To provide individual use, they can be combined by activating the appropriate checkbox. Each parameter comes with a list of standard values for the query. As an alternative, the user may also enter individual values. Thereby more precise queries and results are possible. These values are checked for errors before accepting. Once accepted, they appear as choices in the standard list and are available throughout the session.

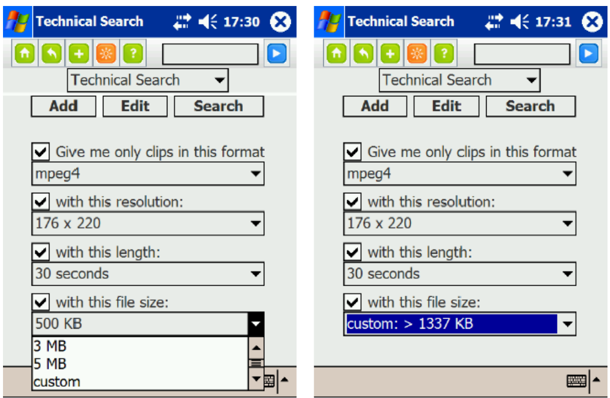


Fig. 7: Technical search: standard choices (l) and a custom value (r).

5.4 Semantic Search

The former interface from inViPla has been extended in the Mobile Query Composer by semantic querying capabilities. The current interface allows one to specify the domain (semantic context, e.g. football or romance) and up to three identifiers (object, person, action). Thus, the semantic context as well as its content are described to formulate queries such as ”show me all scenes where Oliver Kahn is the goalkeeper and a goal is scored” (see Figure 8). With such a way of semantic description, a domain containing identifiers, this work follows the paradigma *taxonomy* that groups semantic information to different domains. Technically the composer uses XML-templates inside MPEG-7 descriptions that are known to both, the server and the Mobile Query Composer. The identifier are restricted to three due to the mobile screen space. To enable formulating more complex queries, this restriction is not existent in the Query Editor view.

Basing on the experience using the concept of taxonomy, enhanced semantic relationships (semantic information are related to each other) can be realized, e.g. by the integration of ontology mechanisms. With this stepwise introduction of semantic search functionalities the particular increase of benefits can be observed.

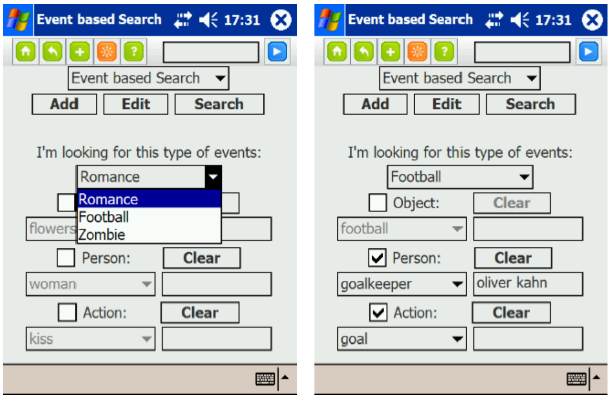


Fig. 8: Semantic search: choosing a template (l) and entering an identifier (r).

5.5 *Query Editor*

The search methods introduced in the preceding sections allow relatively basic queries without any means to formulate complex relationships. As comprehensive search interface, the Query Editor view allows the user to combine all search methods introduced above (keyword, technical and semantic) and visualizes the query using a tree structure. Using a point and click interface, users can choose to show/hide parts of the tree. Nevertheless any complicated query will definitely be rather long and take up a lot of screen space. To support such complicated queries, the Query Editor offers both horizontal and vertical scrolling, so that users can scroll around to keep track of their query tree even when it exceeds the size of the display. The power of the Query Editor lies not only in the different types of conditions that can be used in the query, but also the fact that one can now nest complicated boolean conditions. As the use of boolean operators is unrestricted in the Query Editor, one can write a query consisting of many logical parts and also express complex relationships by adding boolean conditions to specific parts of the query.

A query can be set up in two ways. As first option the corresponding search view (keyword, technical or semantic) can be used to enter search values. Pressing the add button will then copy these information to the system's clipboard and transfer the user back to the Query Editor view. Placing the information into the tree is then simply done by selecting a node and pasting the clipboard information.

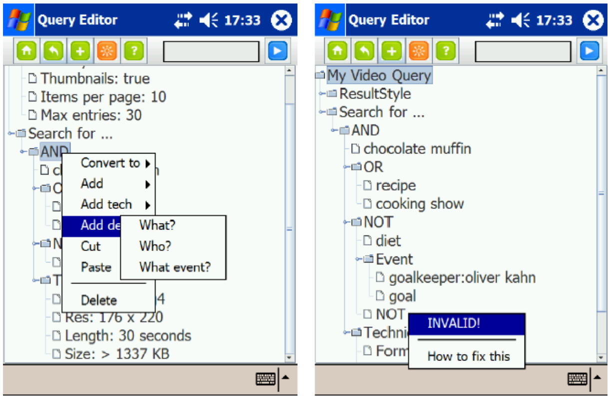


Fig. 9: Sample context menu (l) and error message pointing out an empty NOT node (r).

As second option the view does not switch to other search forms but keeps in Query Editor view. Here, new nodes can be inserted directly into the tree using the context menu as shown in Figure 9. Furthermore it provides options to edit nodes that are already present in the query tree. Leaf nodes are very simple, only their values can be changed. The other ones such as boolean nodes are more complex, hence for these additional operations (e.g. conversion to another boolean type) are available. Moving nodes around can also be done by using the context menu to cut a node and then pasting it elsewhere in the query tree. Each search type has its own clipboard, which is limited in size. Once inserted, nodes can also be deleted if it is permitted by the constraints of the parent node (e.g. the context menu does not allow deletion of the only child of a NOT node, one has to remove the NOT node itself). Finally an MPQF document is only produced when the validation is completed without any syntax errors to avoid producing meaningless queries.

6 Result Presentation

An integral part of any search interface is the presentation of the search results. Often keyframes are shown to reach the goal of a user-friendly presentation. However, using the same base technique, solutions may follow completely different ideas [4] [14]. Due to space constraints of the display, mobile result presentation is not as flexible as for standard PCs. The approach of Mobile Query Composer is the browser view concept, which tries to list the results in a familiar top-down way while still providing an overview of available segements of the selected video. Therefore, videos are listed vertically and for each video the keyframes of its segments are arranged horizontally as shown in Figure 10 and in Figure 11 with example videos. Results are returned in MPQF format by the inViPla server, being parsed by the

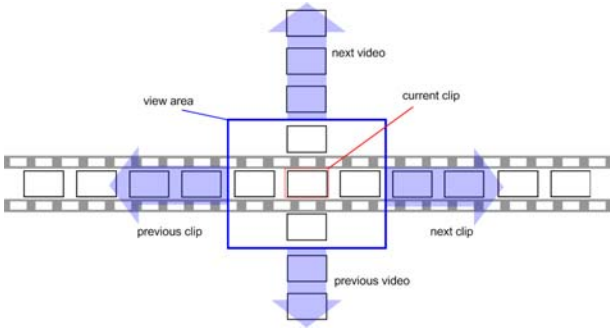


Fig. 10: Result presentation concept - clips (horizontal) and videos (vertical).

Mobile Query Composer which then adds found videos to the current result set. To view these results, one only has to close the search interface to return to the browser view in inViPla. Tapping the stylus on the video clip shown in the middle of the browser will call up a context menu. In this menu one can either start playback using an external video player or view detailed information about the video clip (format, resolution, duration, etc.) as shown in Figure 12 for a semantic search with a result set of one video.

Fig. 11: Browser view - clips (horizontal) and videos (vertical).

Fig. 12: Sample results for a semantic search (foul).

7 Conclusion and Future Work

7.1 Conclusion

This system, the Mobile Query Composer, consists of two parts: a mobile client running on the PDA and a video server. The mobile client is able to generate valid MPQF queries in the form of XML documents and transfer them to the video server using WiFi and a TCP connection. Using that mobile client, it is possible to formulate queries using keywords, technical parameters, semantic concepts and also a combination of these elements. The highlight of the mobile client is the Query Editor view. This view is an interface for more advanced users, which enables query entry using all the aforementioned options and also the nesting of query conditions in boolean statements to produce complex queries. The main feature of the Query Editor is the ability to visualize the query in the form of a tree structure which is similar to the logical structure of the MPQF query itself.

7.2 Future Work

As described in subsection 2.1.2, the management part of MPQF provides mechanisms for service discovery and describing service capability among other features. The Mobile Query Composer is designed to communicate directly with a single remote query server. Its address must be known to the user, which is a rather simplistic assumption. A more realistic solution would be to *improve the connectivity* by implementing some sort of service discovery mechanism. This could find available server automatically, thus improving the ease of use.

Another future work could be a *ontology integration*. Mobile Query Composer still uses a taxonomy, that can be considered only as independent information (not related to each other), describing semantics of a scene and being grouped to domains like "football". To realize semantic concepts to combine those independent information and thus enhance semantic search functionalities, ontology-based datastructures could be used. One possibility are RDF (Resource Description Framework) graphs coupled with the query language SPARQL [13]. Relational information between semantic descriptions would be provided and queries could be formulated more precisely in a bigger semantic context.

Acknowledgement

This work was conducted in the framework of the Multimedia, Distributed and Pervasive Systems (MDPS) network. The MDPS is a French-German doctoral college supported by the Université Franco-Allemande (CDFA-05-08) and the Deutscher Akademischer Austausch Dienst (D/08/10806).

References

1. Adistambha, K., Döller, M., Tous, R., Gruhne, M., Sano, M., Tsinaraki, C., Christodoulakis, S., Yoon, K., Ritz, C., Burnett, I.: The MPEG-7 Query Format: A New Standard in Progress for Multimedia Query by Content. In: Proceedings of the 7th International IEEE Symposium on Communications and Information Technologies (ISCIT 2007), Sydney, Australia, pp. 479–484 (2007)
2. Andreou, G., Karpouzis, K., Maglogiannis, I., KolliasAthens, S.: Mpeg-21 concepts for personalized consumption of heterogeneous multimedia. In: SMAP 2006: Proceedings of the First International Workshop on Semantic Media Adaptation and Personalization, Washington, DC, USA, pp. 141–145. IEEE Computer Society, Los Alamitos (2006)
3. Burnett, I.S., Pereira, F., Van de Walle, R., Koenen, R.: The MPEG-21 Book. John Wiley & Sons, Chichester (2006)
4. van Gemert, J., Mark, J., Koelma, G.D., Nguyen, G.P., de Rooij Cees, O., Snoek, G.M., Worring, M., Seinstra, F.: Mediamill: Exploring news video archives based on learned semantics. In: ACM International Conference on Multimedia, pp. 225–226 (2005)
5. The Nielsen Company. Nielsen's three screen report - telephone, internet and mobile usage in the U.S. (May 2008)
6. Cote, G., Erol, B., Gallant, M., Kossentini, F.: H.263+: Video Coding at Low Bit Rates. IEEE Transactions on Circuits and Systems for Video Technology 8(7), 849–866 (1998)
7. Döller, M., Bauer, K., Kosch, H., Gruhne, M.: Standardized Multimedia Retrieval based on Web Service technologies and the MPEG Query Format. Journal of Digital Information 6(4), 315–331 (2008)

8. Döller, M., Tous, R., Gruhne, M., Yoon, K., Sano, M., Burnett, I.S.: The MPEG Query Format: On the way to unify the access to Multimedia Retrieval Systems. IEEE Multimedia 15(4) (2008)
9. Hürst, W., Meier, K., Götz, G.: Timeline-based video browsing on handheld devices. In: MM 2008: Proceeding of the 16th ACM international conference on Multimedia, pp. 993–994. ACM, New York (2008)
10. Hürst, W., Merkle, P.: One-handed mobile video browsing. In: UXTV 2008: Proceeding of the 1st international conference on Designing interactive user experiences for TV and video, pp. 169–178. ACM, New York (2008)
11. Kamvar, M., Chiu, P., Wilcox, L., Casi, S., Lertsithichai, S.: Minimedia surfer: browsing video segments on small displays. In: CHI 2004: CHI 2004 extended abstracts on Human factors in computing systems, pp. 1371–1374. ACM, New York (2004)
12. Martinez, J.M., Koenen, R., Pereira, F.: MPEG-7. IEEE Multimedia 9(2), 78–87 (2002)
13. Prud'hommeaux, E., Seaborne, A.: SPARQL query language for RDF. W3c recommentation, W3C (January 2008)
14. Sachinopoulou, A., Mäkelä, S.-M., Järvinen, S., Pietarila, P., Peltola, J., Jurvansuu, M.: Personal Video Retrieval and Browsing for Mobile Users. In: Multimedia on Mobile Devices, Proceedings of SPIE-IS&T Electronic Imaging, vol. 5684, pp. 219–230 (2005)
15. Sachinopoulou, A., Mäkelä, S.-M., Järvinen, S., Westermann, U., Peltola, J., Pietarila, P.: Personal video retrieval and browsing for mobile users. In: Proc. of the SPIE - Multimedia on Mobile Devices, pp. 219–230 (2005)
16. Schulzrinne, H., Rao, A., Lanphier, R.: RFC 2326: Real time streaming protocol (RTSP), Status: PROPOSED STANDARD (April 1998)
17. Tseng, B.L., Lin, C.-Y., Smith, J.R.: Using mpeg-7 and mpeg-21 for personalizing video. IEEE MultiMedia 11(1), 42–53 (2004)
18. Vlachogiannis, E., Gavalas, D., Anagnostopoulos, C., Tsekouras, G.E.: Towards itv accessibility: the mpeg-21 case. In: PETRA 2008: Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments, pp. 1–6. ACM, New York (2008)
19. Xu, C., Tian, Q.: Semantic analysis and personalization for mobile media applications. In: IEEE International Conference on Multimedia and Expo, July 2007, pp. 1047–1050 (2007)

Computer Animation and Communicability in Multimedia System: A Trichotomy Evaluation

Francisco V. Cipolla Ficarra and Miguel Cipolla Ficarra

Abstract We present the results obtained through a heuristic assessment of the 2D and 3D animations in commercial off-line and on line multimedia systems. Additionally, a set of quality metrics known as CAC -Computer Animation and Communicability, has been established as well as a communicability guideline, which is a guide that can be applied by the communicability expert at the moment of designing contents for multimedia/hypermedia systems aimed at long distance education, cultural heritage, and video games, for instance. The costs to reach these results have been minimal because a laboratory or specialized equipment have not been required thanks to the involvement of the communicability heuristic assessor for multimedia systems.

1 Introduction

The verb *to animate* implies the action of giving life. In the classical sense many regard animation as moving something in space. However, in the context of computer animation it is a wider notion because it entails a series of variables or components to be considered, such as the changes that produce a visual effect, including location in time, shape, the changes of light on the objects, environmental lighting, the special effects (FX) in the context, the shape and transformation of the objects, the colour, the position of the camera lens, the rendering of the scene, etc. [1] [2] [3]. Now we consider the number of spatial coordinates, that is to say, bi-dimensional, (x-y), or also three-dimensional (x, y and z), which are used to depìct animation. In

HCI Lab. - F&F Multimedia Communic@tions Corp.
ALAIPO: Asociación Latina de Interacción Persona-Ordenador
AINCI: Asociación Internacional de la Comunicación Interactiva
Via Pascoli, S. 15 - CP 7, 24121 Bergamo, Italy
e-mail: ficarra@alaipo.com
e-mail: ficarra@ainci.com

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 103–115.
springerlink.com

the first case it is based mainly on the movement of the lines and surfaces, sometimes with colour, and which occupy a steady and flat zone in the computer screen. Currently, a classical commercial program to carry out this kind of animation is Flash [5]. In three-dimensional animations it is necessary to consider other variables, such as the position of the camera, the approaching or distancing movement of its goal, whether the camera has a static or dynamic position within the scene, the special effects that may surround the objects (fire, explosion, morphing, etc.), textures of the objects, the different kinds of lighting, the interfaces -traditional [6] or organic [7], etc. With the purpose of obtaining a better result inside the assessment of communicability, a diachronic analysis of the assessment of communicability has been carried out through the incorporation of off-line multimedia systems. Our field of study comprises a random selection of off-line and on-line multimedia systems aimed at pastimes (videogames), education and tourism. Out of a total figure of 120 off-line systems (CD-ROMs and DVDs) and 120 websites, a sample of 20% has been chosen for the study and elaboration of the heuristic guide aimed at communicability in computer animation. The animations in some off-line and on-line multimedia systems may refer to famous characters of the mass media that have not used the computer animation techniques in their creation [8]. However, the speed of acceptance of these characters in a hypermedia system and the contents it presents in the communicative process may be much greater if the character is known by the user beforehand, since there is an empathy with him/her. Consequently, knowing each one of the main components of the computer animation design especially for teaching is important for a fast qualitative communication. All of this eases the interaction process with the on-line and/or off-line hypermedia system, especially in teaching and entertainment environments. In contrast to the animations of the computer videogames of the 90s, where the quality of the images was inferior in many cases to gain speed in the movement, the last generations in hardware platforms (Xbox, Playstation and Wii) widely overcome these limitations in the entertainment context [9]. However, one of the main problems in communicability is achieving the perfect union between realistic animations, the potential user and the purpose of the interaction, that is to say, teaching, consulting, entertainment, business, etc. [10] [11] [12] [13]. This is the main reason for the focus of the current work: the intersection of realism and the role of animations in the teaching process with special interest in some design and graphics aspects such as can be lighting and its special effects, the validity of the animation in the face of other audiovisual dynamic means, and the trichotomy relation between objects and/or real/virtual characters animation, their behaviour, the different methods of communicating (gestures and speech in the main) and the influence of the context in animation and communicability. All these aspects analyzed under the technique of heuristic assessment of the observation [14].

2 Realism and Animation

There are multimedia/hypermedia systems which have the whole content of the hyperbase in 3D. However, the content can be presented or visualized in different ways. The usage or not of the 'z' coordinate in the presentation of the content of the system is related to the greater or lesser realism of the image, that is to say, the criteria of quality of the richness and notion of prospect inside the set of primitives used in the heuristic assessment of communicability [15]. For instance, the content of the Brancity multimedia system of the figures 1 and 2 is presented mainly in 3D [16], but in a lesser proportion it also incorporates 2D. The bidimensionality of the topography and the arrows used serves to indicate elements of a 3D object. In this way a kind of breakdown takes place in the coherence of the context at the moment of the presentation of the information. The purpose of the use of 2D can be to attract the user's attention.

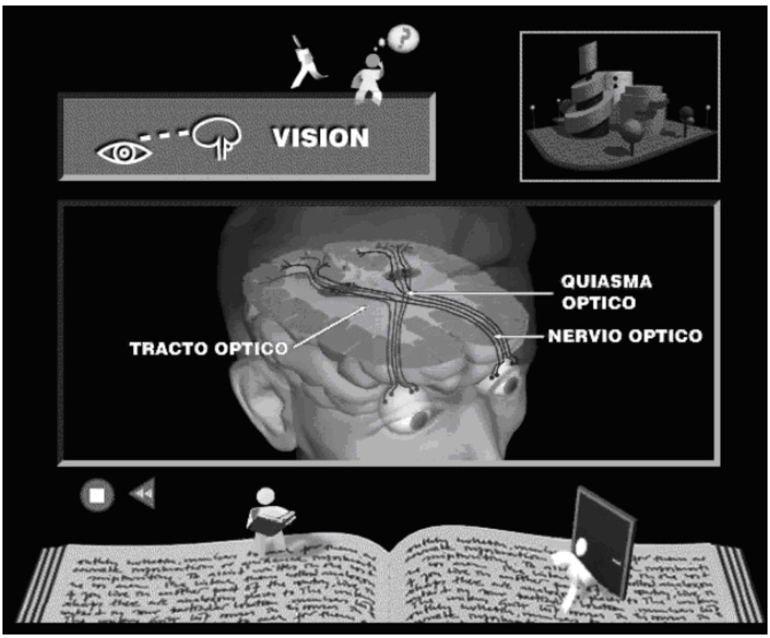


Fig. 1: Animation 2D and 3D.

The figures 1 and 2 have a perfect rendering of the objects that make up the interface. The rendering is achieved through a series of techniques, procedures and algorithms belonging to computer graphics which calculate colour, texture, lighting, etc. of each one of the pixels that make up an image. The goal of these techniques and procedures is to achieve the maximum level of realism in the image. Some hypermedial systems make it possible that the 3D objects are presented to the user

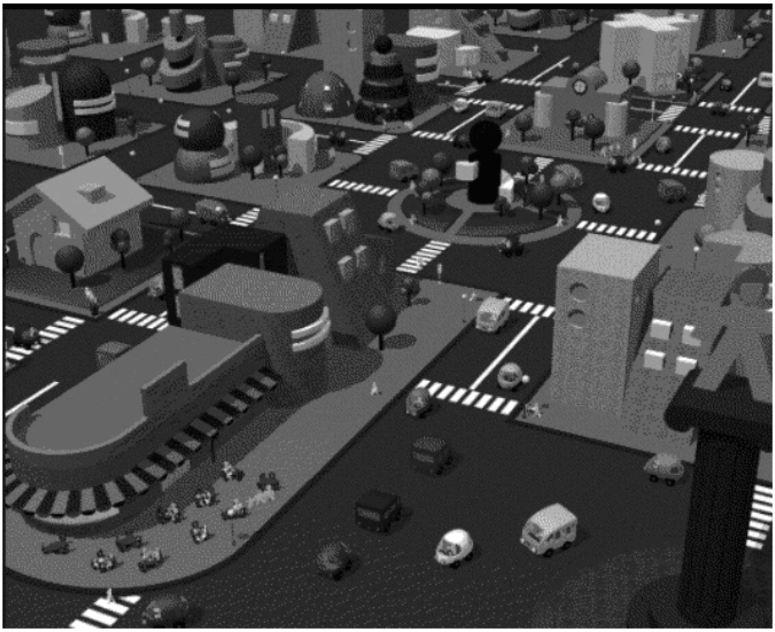


Fig. 2: 3D objects with rendering.

with or without rendering. Next two examples about the use of the ’z’ coordinate and the perfect rendering in the Object Browser system [17]. The first example in wire format, and the second with a perfect rendering in 2D.

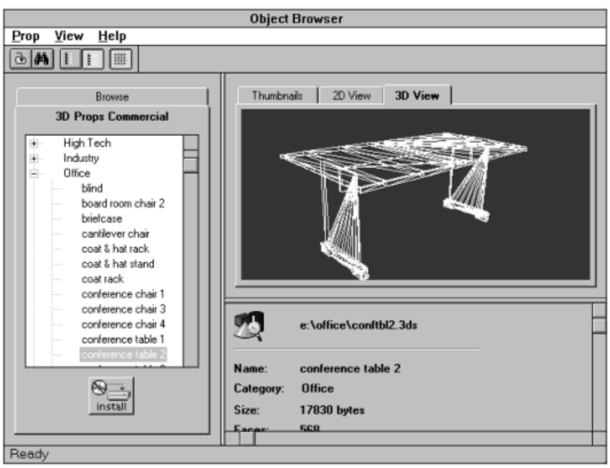


Fig. 3: Object in a wire format: 3D.

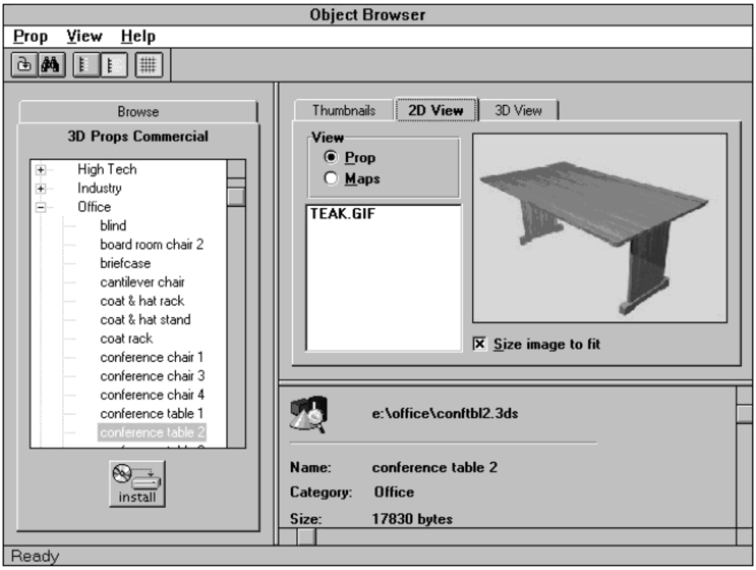


Fig. 4: Object with rendering: 2D.

The animations may have the same ways of visualization as the objects, that is to say, the rendering, with or without rendering in a wire format, in 2D, etc. The animations in a hypermedia system can be classified in the following way: 2D, 3D, metamorphosis, combination of 2D and 3D. The combination of 2D and 3D have a high acceptance rate by young users. Metamorphosis is one of the animation resources that has a high teaching value. The teaching aspect of metamorphosis is due to the possibility of explaining simultaneously all the stages of a process through an origin image and a target image [18]. These two images may be different. Another attribute of metamorphosis is that it draws the attention of the user to objects and elements that are depicted on the screen. In all kinds of animation it is necessary that the speed of frame emission remains constant in the hypermedia system, and that they are synchronized to the audio and/or video. Through synchronic and diachronic metrics it is easy to detect how in the hypermedia systems the bidimensional and tridimensional animations have been occupying bigger space in the dynamic media in order to show the contents to international users. Obviously, the evolution of the storage supports, such as the transition of CD-ROM to DVD, or the speed of the Internet have made this evolution process easier. Simultaneously, there has been a reduction of costs in the making of bidimensional, tridimensional animations or the combination of both, thanks to the appearance of commercial programs used worldwide, such as Flash, 3D Max, Maya, Softimage, etc. aside from the first TV series or animation films in the mass media. That is to say that a bidirectional relationship was generated of the contents of the educational interactive systems with the classic means of audiovisual communication.

3 Computer Animation and Education

Since the early nineties, the education sector has been an ideal field for computer animations. However, there is not much use in interactive teaching content, apart from the virtual agents used in some teaching multimedia systems [19]. The reason for this lack of interest in the multimedia industry lies in the fact that commercial software companies have aimed their production at the entertainment sector, especially videogames. In the past decade, some off-line multimedia systems have included three-dimensional animation as a means to draw attention, through special effects. Some examples are following educational multimedia systems [20] [21]. In all these cases, the animations belong to the quality criteria known as richness. Richness in hypermedia systems consists of the correct combination of static and dynamic means. This is why computer animations were used to embellish the interactive contents in the personal computers. Some contents were presented in an attractive way in the form of speed action games, with the presentation of a visual efect or sound to motivate or seduce the student to follow a study planned beforehand, for instance, in the classic CAE (Computer Aided Education) systems. At first the intention was to replace the paper support, information or books by the computer screen in the classrooms, others, however, have seen a coexistence inside the classroom. Even some manufacturers and designers of personal computers favoured including the traditional elements of the classroom as main metaphors in the first off-line and on-line interfaces of the operative systems. Now, once the pupil was attracted by animation, an additional animation could be used to generate a fruition of agreeable content. It was the moment at which the structure and the access to the information of the multimedia system on-line allowed the user to connect to the Internet. For instance, from an atlas in CD-Rom form that talked about the globe, to the NASA website, to see real simulations of the earth spinning on its rotation axis. Another attractive feature aimed at the artistic activties of computer animation is the degree of realism which can be offered in the reproduction of artistic works, such as painting. The degree of realism is higher, because computer screens have a higher resolution as compared to the classic photo slides, or the pictures in analogical support. Very subtle color nuances, when used by non-expert users, can be created and recreated with great ease by any user. Besides, it is possible to foster the change of colors, increasing the chances of drawing the user's attention, as the angle of the artistic fruition changes, or at the moment of increasing or softening the lighting of the work, to mention a couple of examples. Without any doubt, in the nineties the European producers of multimedia systems off-line, public and private, related to art, have consistently used each one of the resources stemming from computer animation, not only for educational purposes, in order to better understand art, but also to attract the potential visitors to the real museums, after a virtual visit. Some examples of these multimedia systems are: Le Louvre [22], Muse d'Orsay [23], etc. The content in the hypermedia systems posses a main visual element for the development of visual culture in the educational context such as: the ability to store the digitalized high quality images with the chance that the user has access to the works he wants to see and additional explanations, through bidimensional animations such as the perspective of the looks of the characters in a picture, the real dimension of a work as compared with the height of an average person, the prevailing lines, the light focus, the colour relationships, iconographic schemes, etc. Besides, the right combination of text and images may be a boost to promote and motivate the fruition of the stored information in the database of the interactive system. It is also important to consider the approaching and distancing actions at the moment of seeing an image in a bitmap or vectorial factor. Through the use of the magnifying glass they allow to discover peculiarities that usually escape the human eye. Obviously, the possibilities of use of animations as a teaching instrument are boundless.

4 Dynamic and Static Components in Multimedia Systems

The data that are emitted or received through the communication channels of a multimedia/hypermedia system make it possible to split the channels into static or dynamic medias. Some examples of static media are the maps, the graphics, the texts, the pictures and the other animation free figures. Whereas in some examples of the dynamic media they are animations by computer, video and sound. The difference between static and dynamic media is in the temporal variable in the presentation of the data. However, there is a mistaken tendency in some countries in Southern Europe in branding all dynamic data as audiovisual data, particularly in the encyclopedia systems [24]. The audiovisual data must be related only to the filmed images, that is to say, those images that have been recorded with a video camera, a cinema camera, mobile phone, etc. This conceptual failure has brought about even the confusion of dynamic data with the static data, as can be seen in the left margin of the figure below:

The differentiation among the data that currently prevail in the on-line and off-line databases of a multimedia/hypermedia system are not the same as those of the last three decades. This is important not only for the correct design of the interface, contents, navigation, etc., but also in the suitable grouping of the multimedia/hypermedia systems at the moment of the assessment of communicability. The lack of provision of the time factor may have a negative influence on the research results, since the databases of the multimedia/hypermedia systems were not the same in the 1980s as in the current systems. Taking a look at the evolution of the first hypermedia systems as compared to the current hypermedia systems, it can be seen how the dynamic data have increased their presence in the hyperbase contents, and even the presentation of this content has been fragmented at several levels [25]. The levels of content can be defined as the amount of information to which the user has access inside the system. This is a characteristic feature of the systems aimed at computer assisted teaching and entertainment [26]. The selection of the level is usually done at the beginning of the interaction with the content. The contents, which are related to the type of user of the multimedia system, may be divided into three levels: low, intermediate and high. In the lower level the user has access to a bigger amount of information than in the higher level, and vice versa. It is at the moment of

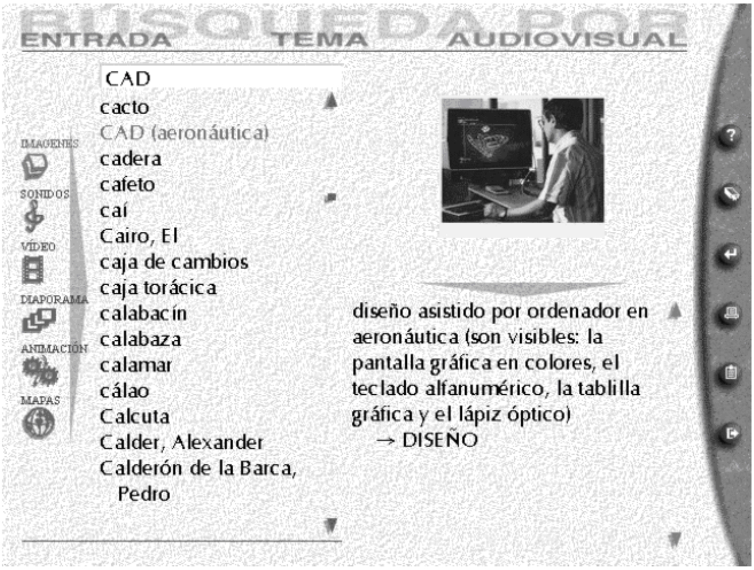


Fig. 5: Incorrect use of the audiovisual notion by combining dynamic and static data.

design when the communicability expert determines the amount of information for every level, the use of one or several dynamic levels simultaneously, (audio-video, audio-animation, video-animation, etc.), the amount of nodes in the guided links, the overall number of components in each entity, etc. Once the level of content has been defined, and knowing beforehand the potential user of the system, it is not correct to subdivide a same level into other three. A solution to this problem, and one which was successful in a E-learning system, was of a dychotomical kind, that is, to cut down the levels from three to two through a question based interrogation of the user. This fractioning of the content into several levels has an influence on the extension, the organization of the kinds of data, mainly of the textual kind, and the possibility of disabling the dynamic media in hypermedia systems. For instance, the text which is presented on the screen must be organized in such a way that regardless of the level chosen by the user you can have access from the start to a summary of it.

5 Virtual Characters and Real Actors: Trichotomy Evaluation

Sometimes to speed up the presentation process of art content, tourism, education, etc, virtual guides are inserted. These guides can be 2D, 3D or 2D + 3D: the video of real people. The success of inserting several kinds of guides in a hypermedia system will depend on the communicability of the contents in relation to the potential users. In this case, when we talk about communicability, we mean previous assessments of the quality of the contents before the massive production of the off-line multimedia supports or of setting on line the hypermedia contents. What follows are some examples of hypermedia systems where correct [27] and incorrect [28] use of virtual and real guides can be seen -circle area.

Fig. 6: A correct use of virtual character.

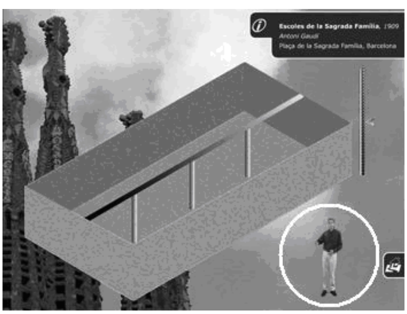


Fig. 7: An incorrect use of real actor.

In the case of inserting video of real characters as guides in systems whose contents are related to cultural heritage, it has been demonstrated that there is a series of technical drawbacks at the moment of synchronizing the static means and the dynamic means or very small, for example, in the off-line multimedia system Gaud [28]. In this case, a 2D character would have solved the synchronism problem. That is why in the use of characters it is important to decide at the design stage of the system whether it will be real or imaginary. A character with a realistic look should also be animated in a realistic manner. If the goal is to design and animate a realistic character, the reference to the real is essential. Usually, a real character is that in which the human figure keeps the proportion of 6 or 8 heads as average in height to achieve a realistic character. Imaginary characters are only two or three heads tall. Imaginary characters can be designed and animated more easily than a realistic human character. However, these characters are not necessarily synonymous with poverty in the animation and in the context in which they move. The most important thing in a character is the personality and the degree of communicability with the user. For instance, a big character with a negative or mean role usually has wide shoulders and bright, small eyes. At the same time, big eyes in a fat character implies goodness [3]. Whereas a character with big muscles and a small head denotes intelligence. In contrast, a big head in a small body may imply genius. Regardless of the dimensions or shapes of the character, to enhance the visual features which define the character's personality in the modeling stage, it is essential that its surface is easy to deform or manipulate in the animation process, to lower the production costs of animation. Here are some reasons why video might be used rather than animations. The animation of characters is the study of movement, time and action. The components' originality, simplicity and universality are some of the principles or quality attributes which guide the creative process to obtain quality animations. In our case we have worked with a set of quality attributes of Computer Animation and Communicability -also known as CAC. These attributes are a trichotomy [29] between: acteme (the most basic unit in the analysis of communicative behavior), acteur (an 'actor' or character/object at the surface level of a narrative discourse or action) and behaviorism (the role of the environment in determining behavior). Each one of these attributes has been transformed in metrics in order to assess heuristically the design of multimedia systems in the following categories: content, panchronic (from Greek *pan-* and *chronos*, time) and presentation, and using the primitives which are the result of the intersection between multimedia/hypermedia systems, human-computer interaction, software engineering and semiotics. The results obtained in the assessment of the design in the universe of study (see annex #1) have made it possible to make the communicability guidelines (see annex #2).

6 Conclusions

The costs of creation of computer animations have been cut down in the last decade through the democratization of graphic software and hardware. The quality of certain animations has such realism that it can now beat video and digital photography, especially in the case of scientific visualization. The 2D and 3D animations have a very high educative function in the current multimedia and virtual reality systems, aimed at education and entertainment. The diachronic analysis of the off-line multimedia systems has made it apparent that in 13.3% of the studied cases the quality is superior with regard to on-line animations, since one of the main problems remains the synchronisation of the dynamic means due in many cases to the use ofbroadband. The quality attributes and the generated metrics have allowed us to easily detect the presence or absence of the quality of communicability and the computer animation. Additionally, the observation technique used in heuristic assessment to generate the guide has been economic and reliable. The use of the technique by the communicability expert has made it possible to detect systematically those details in the design stage that boost communicability. This was one of the purposes of the elaboration of the guide. Nowadays and in regard to the content of the hypermedia systems, the animations may be essential to overcome the legal copyright barriers that some artistic works have for their reproduction, cutting down considerably the manufacturing costs.

Acknowledgments

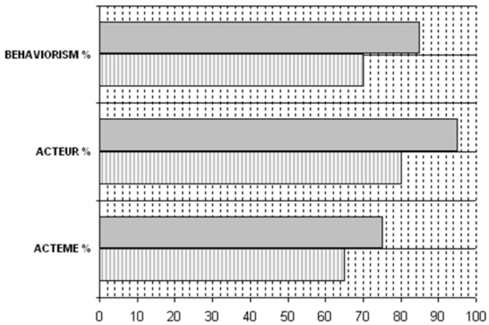
A special thanks to Emma Nicol (University of Strathclyde), Maria Ficarra (Alaipo & Ainci - Italy and Spain) and Carlos for their helps.

References

1. Newman, W., Sproull, R.: Principles of Interactive Computer Graphics. McGraw Hill, New York (1979)
2. Demel, J., Miller, M.: Introduction to computers graphics. PWS Engineering, Boston (1984)
3. Maestri, G.: Character Animation. New Riders, Indianapolis (2000)
4. Birn, J.: Digital Lighting & Rendering. New Riders, Indianapolis (2000)
5. Fernandez, I.: Macromedia Flash Animation & Cartooning: A Creative Guide. McGraw Hill, New York (2001)
6. Billinghurst, M., Grasset, R., Looser, J.: Designing Augmented Reality Interfaces. Computer Graphics 39, 17–21 (2005)
7. Colman, D., Vertegaal, R.: Organic User Interfaces: Designing Computers in Any Way, Shape, or Form. Communications of the ACM 51, 48–53 (2008)
8. Lord, P., Sibley, B.: Cracking Animation. Thames and Hudson, London (1998)
9. Höysniemi, J., et al.: Children's Intuitive Gestures in Vision-Based Action Games. Communications of the ACM 48, 44–50 (2005)
10. Obrenovicc, Z., Gasevic, D., Eliens, A.: Stimulating Creativity through Opportunistic Software Development. IEEE Computer Software 25, 64–70 (2008)
11. Chatham, R.: Games for Training. Communications of the ACM 50, 36–43 (2007)
12. Van Welbergen, H., et al.: Presenting in Virtual Worlds: An Architecture for a 3D Anthropomorphic Presenter. IEEE Intelligent Systems 21, 47–53 (2006)
13. Van Lent, M.: The Business of Fun. IEEE Computer 41, 97–103 (2008)

14. Nielsen, J., Mack, R.: Usability Inspections Methods. Wiley, New York (1994)
15. Cipolla-Ficarra, F., Cipolla-Ficarra, M., Harder, T.: Realism and Cultural Layout in Tourism and Video Games Multimedia Systems. In: Proc. of the 1st ACM Workshop on Communicability Design and Evaluation in Cultural and Ecological Multimedia System, New York, pp. 461–470 (2008)
16. Braincity CD-ROM. Digital Illusion, Barcelona (1995)
17. 3D Props Commercial CD-ROM. Autodesk, San Rafael (1995)
18. Goes, J., et al.: Warping & Morphing of Graphical Objects. Morgan Kaufmann Publishers, San Francisco (1999)
19. Gulz, A., Haake, M.: Design of animated pedagogical agents -A look at their look. International Journal of Human-Computer Studies 64, 322–339 (2006)
20. Enciclopedia de la Ciencia CD-ROM. ZetaMultimedia, Barcelona (1998)
21. Cómo funcionan las cosas CD-ROM. Zeta Multimedia, Barcelona (1998)
22. Le Louvre CD-ROM. Montparnasse Multimedia, Paris (1995)
23. Musée d'Orsay CD-ROM. Montparnasse Multimedia, Paris (1996)
24. Enciclopedia Multimedia Larusse CD-ROM. Espasa-Calpe, Madrid (2001)
25. Frohlich, D.: The Design Space of Interface. In: Proc. Eurographics Workshop Multimedia: Systems, Interaction and Application, pp. 53–69. Springer, Stockhom (1991)
26. Woolf, B., Hall, W.: Mutltimedia Pedagogues: Interative Systems for Teaching and Learning. IEEE Computer 28, 74–80 (1995)
27. Little Monsters CD-ROM. Ramson, Winchester (2000)
28. Gaudí CD-ROM. Barcelona Multimedia (2001)
29. Sebeok, T.: Global Semiotics. Indiana University Press, Bloomington (2001)
30. Cipolla-Ficarra, F., Cipolla-Ficarra, M.: Interactive Systems, Design and Heuristic Evaluation: The Importance of the Diachronic Vision. New Directions in Intelligent Interactive Multimedia, pp. 625–634. Springer, Heidelberg (2008)

Annex #1: Trichotomy Results –Multimedia On-line and Off-line



	ACTEME %	ACTEUR %	BEHAVIORISM %
OFF LINE	75	95	85
ON LINE	65	80	70

Annex #2: Comunicability and Computer Animation Guidelines

Table 1. Components, Attributes –Acteme [Ae], Acteur [Ar], Behaviorism [Bm] and Design Categories: Content (Co), Panchronic (Pa), and Presentation (Pr)

• Modeling: 2D, 3D, 2D and 3D.	[Ar, Bm] (Pr, Co)
• Illumination: Point (omnidirectional) lights, spotlights, directional lights, area lights (spherical area lights, flat area lights, linear lights, models serving as lights), etc.	[Bm] (Pr, Co)
• Lighting: softness, intensity, color, throw, animation (moving the light and animation light parameters).	[Ar, Bm] (Pr, Co)
• Shadow: Single-shadow scenes, shadows from fill light, shadow clutter, omitting shadows, and shadow areas.	[Bm] (Pr, Co)
• Color: Contrast, meanings of colors (contextual associations, culturally specific associations), color and depth (i.e., warm and cool colors, saturation, etc.), black and white.	[Bm] (Pr, Co)
• Time and Context (scenario or setting): Modern or actual (i.e., city, country, etc.), past (historical), future (science fiction).	[Bm] (Pr, Co)
• Materials –Textures (natural or artificial), transparency, opacity, etc.	[Bm] (Pr, Co)
• Rendering –shading surfaces, raytracing, global illumination, etc.	[Bm] (Pr, Co)
• Camera position and movement: Angles (i.e., perspective, high-angle and low-angle shots, POV shots), zoom in, zoom out, travelling, rostocopie, etc.	[Ac, Ar] (Pr)
• Educational objects: Chalkboard, globe, rule, pencil, etc.	[Ae] (Pr, Co)
• Character Static Realism: Face, skin, hair, beards, etc.	[Ae] (Pr, Co)
• Character Dynamic Realism (movements believable): Running, walking, etc.	[Ae, Ar] (Pr, Co)
• Character/s Fashion: Clothes and accessories –i.e., watches, jewels, etc.	[Br] (Pr, Co)
• Character/s Real or Imaginary –simulation based on real actors, literature, etc.	[Ar] (Pr, Co)
• Voice: normal, fast talking, booming, whining, melodic, robotic, alien, weird, etc.	[Ae, Ar] (Pa)
• Movement: normal, acceleration and deceleration, etc.	[Ar] (Pr)
• Special movement: lip synchronization, expression and gestures, etc.	[Ae, Ar] (Pa, Pr)
• Static Media: Photography, map, graph, draw, picture, etc.	[Bm] (Co, Pr)
• Dynamic Media: Sound, music, audio FX, video, etc.	[Ae, Bm] (Co, Pa, Pr)
• Deformation objects and/or body: metamorphosis, translation, scale, rotation, etc.	[Bm] (Pr)
• FX natural: Water, fire, smoke, rain, snow, wind, melting, freezing, etc.	[Bm] (Co, Pr)
• FX disaster: Coalition, explosion, earthquake, tornado, flooding, etc.	[Bm] (Co, Pr)

Virtual Learning Environment: Quality Design for Foreign Languages in Multimedia Systems

Francisco V. Cipolla Ficarra

Abstract We present a communicability guide applied to distance learning. The guide is the result of the triadic relationship between a set of commercial multimedia systems offline, a university website for E-learning, and the diachronic vision of the multimedia applications carried out for a Spanish virtual campus, whose results have demonstrated to be excellent in the costs-quality equation. The heuristic analysis of the main components makes it possible to obtain high quality in the design of hypermedia systems as well as in the communicability between the participants in the educational process, and with the cheapest use of a usability laboratory. The guide for the quality of multimedia content in a virtual learning environment represents the intersection of the following areas: human-computer interaction, usability engineering, software engineering, pedagogy and the primitives used in the design models for multimedia systems.

1 Introduction

There is a tendency to use the terms e-learning and on-line learning as if they were synonymous. However, e-learning has a wider meaning, and generally refers to an educational opportunity based on telematic and computer systems. The first university to make use of these systems was the British Open University in 1994 [1]. Online learning is a concept that refers to distance teaching (where there is vertical communication between students and teacher/tutor), where the interaction and cooperation among all the agents participating in education is included (here the horizontality of communication is essential) [2]. Aside from these questions of meaning

HCI Lab. - F&F Multimedia Communic@tions Corp.
ALAIPO: Asociación Latina de Interacción Persona-Ordenador
AINCI: Asociación Internacional de la Comunicación Interactiva
Via Pascoli, S. 15 - CP 7, 24121 Bergamo, Italy
e-mail: ficarra@alaipo.com

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 117–127.
springerlink.com

between both notions, there is a learning methodology that allows one to create a real learning community which breaks away from the bi-directional relationship between teacher-student through the relations with a group: classmates, teachers, tutors, etc. The goal is to implement and value the social component in the learning process, through the new technologies. A set of quality attributes in interactive teaching in multimedia/hypermedia systems both on-line and off-line, is necessary to boost the individual growth which is stimulated by the interaction with others. In a virtual environment, learning is a process where the individuals, through interaction and feedback, verify their points of view with the others, challenging their ideas and building actively their thoughts and conclusions.

In the current work the state of the art of distance teaching is described along with its technological and pedagogical evolution in time. The main reasons for which users, i.e. teachers, students and academic authorities, resort to these multimedia systems are also examined. The goal is to analyze the communicative design of a set of multimedia off-line and on-line systems aimed at the teaching of the English language. Simultaneously, a diachronic study has been made of the current on-line interface components and off-line interactive products of two virtual campuses. The first is for the university of UNSAM (National University of San Martin -Buenos Aires, Argentina) and the second is the UOC (Open University of Catalonia -Spain). These results have made it possible to compose a guide of heuristic assessment of communicability in the following categories: presentation or layout, panchronic (from Greek *pan-* and *chronos*, time), content, navigation and structure. Communicability entails quality. Quality in interactive software, the same as the notion of beauty in art or design, is very easy to detect when it is absent. Quality design is an excellent interrelation between communicability, usability and users. Quality design requires a set of methods, techniques, metrics and specialized staff to exist in the software of multimedia systems. Our guide is the result of the implementation and assessment of educative multimedia systems over a period of 12 years. Using a communicability expert it has been possible to obtain excellent results in the design with a reduction in costs since working in this way does not require the use of a laboratory, specialized technical equipment, or staff .

2 New Technology, Languages and Learning

Education and health are two cornerstones of any country of the different continents of our planet. The increasing affordability of personal computers, whether it is software or hardware, is allowing the realisation of the famous global village announced by Mc Luhan [3]. The Internet was presented in the 90s as the digital library of Alexandria [4]. Therefore, users could freely access contents not only to make consultations, but also to study via distance learning. Distance coupled with climate has been one of the main problems in societies in ensuring that their populations can have access to education. This is the main reason why E-learning has developed more quickly in the north of Europe, especially in the Scandinavian peninsula, than in the European South. Going back centuries, distance learning methods were used by traditional couriers; the English speaking countries of America and Europe being pioneers in this.. Currently the European website devoted to E-learning (i.e., elearningeuropa - www.elearningeuropa.info) presents the need to count with technologies to the service of permanent teaching. Obviously such an overall objective entails a series of secondary goals: to create a common platform for learning through computers, to give incentives to the teachers to foster the use of this technology with teaching purposes (an essential point indicated by the greatest US software manufacturer, which is aimung to develop an operating system [5], to unite all the classrooms and all the schools through the Internet, to train the staff in the environment of digital culture [6], and to make adequate education programmes right from the early years of teaching, especially in the field of English and computers. Remarkably enough in this last regard some of Piaget's premises are not necessary, that is, to learn while playing [7]. In our research work we focus on learning but including the entertainment aspect for children, adolescents and adult users, especially for the learning of English. One of the shortcomings in the field of E-learning is the almost total absence of significant experiences in the use of the technologies in the field of teaching and education. That is why there are in Europe some guidelines tending to boost the learning of languages. Linguistic pluralism is one of the most characteristic aspects in Europe when compared with the American continent, for instance. Besides, it is a decisive factor for future generations, if they are to build a common European identity, or in the access to labour markets throughout Europe, and from other continents. The study of languages does not only concern the traditional teaching centres such as schools, secondary schools, colleges, etc., but also an important sector inside the multimedia industry aimed at teaching. In these systems the intersection between the off-line and the on-line offers new communication opportunities between those who learn and those who teach, or it also serves to give support to presential learning in the case of languages. Currently there is a wide range of possibilities to attend an on-line course virtually, thus facilitating access to education for all those who can't do it on the grounds of work, time, adverse climatic conditions, etc. Interactive courses and virtual classrooms are being extended quickly through international geography, bringing about an evolution or revolution in the context of elaboration and transmission of data and information in the telematic networks [8].

3 Analog and Digital Information for E-Learning

There are disagreements as to the origins of distance teaching [1] [2]. On the one side there are those who maintain that its genesis is in the Swedish and English universities, when in the mid-nineteenth century they decided to teach calligraphy by correspondence. In the industries there was a growing demand of staff for secretaries. It was then that the first models of distance teaching activities were created.

Others in contrast place its origins in the early 1900s, when through correspondence, material was received and sent in paper support [1], between teachers and students, in a continuous process of bidirectional communication (feedback), with an emitter-teaching, receptor-student dynamic. Later on, the social media of communication were included such as the radio, in the first place, and television later on. In 1969 the British Open University started to offer distance university teaching and it was the first in possessing the three multimedia systems needed for distance teaching, that is to say, sequential, partially interactive and totally interactive [2]. Aside from the temporal component and analyzing the history of software and the diffusion of informatics in the 80s, we see how in the office environment the era of the computer occurred thanks to such programs as Wordstar, Wordperfect, Word, etc. Therefore, it is in the diffusion of textual contexts that we place the origin of this evolutionary process, as it also happened with the hypertextual systems which broke the linear sequence of textual contents [9]. Over time, open universities have developed a set of diversities related to teaching and the methodology, whether it is in the kind of teaching material, as in the modality of putting forward the distance teaching process. Consequently, these days there are four modes of delivery:

1. Single mode, when the university has the main function of distance education and resorts to the telematic/multimedia systems on-line and off-line, as also do the traditional post system for the transportation of CD-ROMs, DVDs, books, exercises notebooks, etc.
2. Dual mode when the university devotes itself to face to face teaching and incorporates distance education, supporting multimedia systems generated inside or outside the teaching centre (outsourcing).
3. Mixed mode, which is a combination of the two former categories.
4. Virtual mode, those universities that have their own courses in a virtual mode, resorting to dedicated software.

The virtual university is a system of online learning that makes possible a teaching atmosphere in the community as well as the construction of interdisciplinary knowledge. Its main feature is the flexibility of the integrated environment, joined to videoconferencing systems and other pedagogical resources: ad boards, virtual secretaries, chat, etc. [1] [2]. Now, in all these cases it is necessary to take into account the technological reality of the potential users. For instance, the speed of the net, the kind of computers and peripherals, to which they can have access, and the geographical location of these students. Obviously, the off-line hypermedia systems may even today help in some aspects for the learning of languages, when the area does not have a connection to broadband Internet, for instance. In the strategy for the design of the first multimedia systems in the Open University of Catalonia in 1995, the floppies were chosen to store the multimedia system in a compressed way [10]. Later on, it was switched to the installation of that system in the user's hard drive. Although at the time it was the era of the CD-ROM, many users did not have a CD-ROM reader in their personal computers because of the high cost of this.

4 Quality Desing in Virtual Classroom

Through the current methodology the intention is to implement an activity such as the heuristic assessment of the design of hypermedia systems in each one of the main categories, which are: content, structure, navigation, presentation and panchronism (from Greek pan- and chronos, time). This activity is made by specialists in communicability for on-line and off-line multimedia systems [11]. Some of these off-line commercial systems with an international distribution (see annex #1) have demonstrated that the young user stops interacting with them quickly. That is to say, the advantage of free access tothe learning of languages in the form of CD-ROM or DVD is now meeting competition in the shape of the use of virtuality in the teaching process, where there is no attendance obligation. It has been seen in many of these systems that they have an excellent level of communicability in the design and that they respect the structural rules of this kind of contents related to the learning of languages. Generally, these courses in off-line and on-line formats are used as a reinforcement to the face to face aspects of the course teaching. In the study that was carried out for this paper it was detected that many students need a real teacher for the learning of a language, while some adapt to the virtual tutors. In both cases, the students are looking for safey and confidence in the assimilation of the new language. A real or virtual tutor can attract the attention of the student to correct the pronunciation or the structuring of the sentences [12] [13]. Virtual classrooms are one of the main elements of E-learning, and have the purpose of fostering social activity. In contrast to other models of distance teaching, there is not a high number of students and they have tutors, experts and observers who are dedicated exclusively to them. The production of the educational material may have a lower cost in the whole life cycle when compared with the costs in paper support, for instance. Apparently, at first the costs of the multimedia information may be more competitive. However, with the reusability the information in digital format, onemake these adapt themselves to the new needs of the students. This means that costs are in reality lower, especially if there is a constant recycling of that multimedia content. In the virtual classroom there are two ways of communication: vertical and horizontal. The first it is related to the relationship between professor/tutor and students, where the content can be adapted to the advance of the contents or difficulties that appear in the teaching process. The communication among students through virtual communities, discussion forums or chat is described as horizontal communication. The adaptability of the content is essential in continuous training. Continuous training these days can be developed inside university centres, or in private enterprises, such as training courses for the handling of new machinery. With regard to this, some virtual tutors achieve excellent results, such as in the case of Cosmo, Steve and Hernan [12]. Finally, it has been observed that those multimedia systems that have virtual tutors available are more frequently accepted as the age of the user decreases. The older that the user grows, these are used only infrequently as an aid in the cases of aimless navigation, pronunciation correction, searches for meaning of the words and for synonyms.

5 Virtual Community and Open Software

The virtual community among students can result in genuine reciprocal teaching, via open software and the process of interchange of mutual experiences, for example. Chat facilitates this process because there is a communication among the students in real time. From a technical point of view the communication is superior in many cases when compared with videoconferencing. The use of videoconferencing is one of the most significant in training via the Internet, extranet and intranet. The teachers and professors who participate have the feeling of sharing the same space when there is panchronism. However, for reasons of speed the telematic net does not have a great worldwide diffusion yet among end users. The novel aspect of the virtual community in the educational context lies in the interaction with real society. The students can be developers of products, services, projects, etc. There is also an interaction between professors inside the same country, and with those of other nations. Each one of them can participate in the formation of the virtual community to which they belong. Now the categorization of the content and presentation within the design of these multimedia systems are essential. That is why it is necessary to analyze and design correctly the teaching materials, the content of the courses and the way that they work on-line. It is not enough to include on-line slides made with PowerPoint, for instance. It is necessary that this content isfrequently updated. In E-learning or on-line learning, the community of students and teachers are to be located on-line, therefore, it is impossible to generate a whole series of communicative strategies to foster the interaction of motivating character. Motivation is the essential element of attention, memory and learning [1] [14]. It is necessary to tell with a rich interaction (forum, chat, videostreaming, etc.). In the following figure we can see the interface of the videoconferencing system for the virtual campus in UNSAM (National University of San Martin), whose communicability analysis has made it possible to insert components in the communicability heuristic assessment guide.

In the case of languages, all these interactive resources help in the process of strengthening knowledge. Besides, the virtual community, through the use of videoconferencing, facilitates the learning of everydayexpressions, and considering the synchronic aspect of a language. The Internet offers to the young the chance to have free access to on-line language courses. Many of the courses currently available have been made with open source software and in accordance with state-of-the-art virtual campus models. Therefore, these users, directly or indirectly, are going ahead, technologically of the tools that they will encounter in their future university studies. In those university virtual learning environments for the teaching of languages to ad hoc developments have been resorted to, but not without high costs. The costs arise from: the maintenance and the constant updating of personalized interactive systems, the yearly expense of the user licences of the operating system, etc. The ideal alternative these days is to resort to open source software to cut down as far as possible the licensing costs of the software used for programming, for instance. Another of the novelties of open source software is the possibility of creating one's own virtual environment aimed at distance teaching and learning[15].

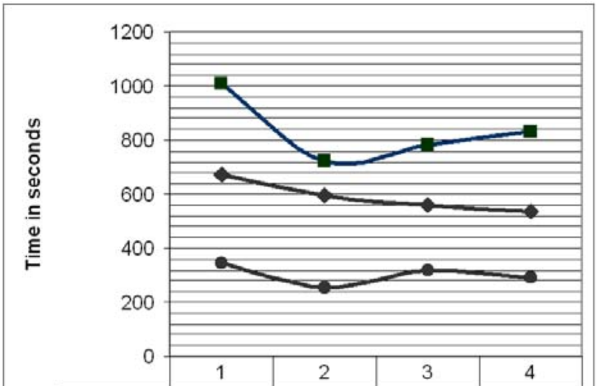
Fig. 1: Virtual Campus UNSAM -Videoconferencing for E-learning.

Currently Moodle (Modular Object-Oriented Dynamic Learning Environment) is a Course Management System (CMS) also known as a Learning Management System (LMS) or a Virtual Learning Environment (VLE). University centres are spending an significant percentage of their financial resources onsoftware licenses and the maintenance of the hardware of large servers [16]. In contrast Moodle is a free web application that educators can use to create effective on-line learning sites.

6 Heuristic Evaluation

With the purpose of increasing the validity of the presented assessment we regard it important to analyze in detail the strategies used in the generation of off-line multimedia systems for a Spanish virtual campus (UOC -Universitat Oberta de Catalunya) [10]. It is a diachronic study with reference to two aspects: communicability and usability. Having a diachronic vision allows us to be more objective in the assessment of the design of these interactive systems. For instance, the greater time required to navigate a computer aided education system on a PC from the early nineties can't be evaluated as compared to current systems without also considering the evolution of the hardware and the operating systems. The high number of explanations in the multimedia system was due to the fact that there was no such wide circulation as nowadays of these concepts. Now each one of the used strategies and the obtained results will be considered. In the applications for the UOC, three groups of adults who were non-expert in the use of computers were involved. Two assessors participated and the technique used was direct observation and the questionnaire [17] [18]. One of the assessors belonged to the group of the "'double"' experts in usability [17], since he took part in the design of the system. The test consisted of the complete navigation through each one of the nodes which make up the system to measure the time which each one of the users needed to go through all the subjects. The assessment can be divided into four stages: heuristic assessment of the system by an expert; the first interaction of the users with the system; the second interaction with the system and completion of the questionnaire; the third interaction with the system (previously changes have been introduced in the design). In the following diagram the results are compared between the assessment made by a specialist in usability and the other two made with the users for the UOC. It is observed how the changes introduced in the design of the system have favoured navigation, since the times have been reduced (the results of the second measurement are nearer to the results of the usability aspect and the communicability strategies applied in these systems carried out in a record time: 1 month).

Depiction of the results of the assessments carried out by two groups of users non-expert in the use of computers: (see series 1 and 2.In the series 2 modifications in the design had been introduced previously). and a usability expert (series 3). Reference: 1= operating system, 2=word processor, 3=calculation sheet, and 4=electronic mail. The average improvement incorporated in the system reached almost 25%, with which the efficacy of the proposed methodology is demonstrated, especially when one works in the heuristic assessment, with users whose previous knowledge in the handling of computers is nil. The communicability and usability in the system was excellent and enabled the users who had never before used a computer to participate in the E-learning process of the virtual campus. In some cases, the students learned in less than a week to use each one of the services of the virtual campus: forum, advertisements board, calculation sheet, text processor, etc. In less than a month the virtual community for the multimedia communication course had been set up. A very positive result since at that time there was no Moodle.



	1	2	3	4
Serie1	675	594,4	557,7	538,4
Serie2	1010,2	723	781,5	832
Serie3	345	253	316	289

Fig. 2: Communicability and Usability -UOC, Heuristic Evaluation Results .

7 Conclusions

The combination of the analysis of communicability and usability in the design of multimedia systems aimed at long distance education have demonstrated excellent results in the last few years, regardless of theformat of the digital information. In the analyzed off-line multimedia systems a good communicability level has been detected. At the same time, the quality of these off-line systems during adolescence, fosters an interest towards the learning of foreign languages via the Internet. An early familiarity with text processors, the search for information on-line, communication via chat or e-mail, favours the generation of a virtual community. Therefore we regard as important the inclusion of some of the resources of the virtual classroom to boost quick integration of potential college students, desirous of learning languages. In the universe of study of children it has been seen that content in the learning process of English refers largely to prepositions, words, phrases, verbs, adjectives, pronunciation, vocabulary, expressions, dialogue, the hours of the day, frequent questions and numbers, all of it under the principle of learning by playing. Tthe degree of complexity of the design of textual content for the learning of a foreign language is usually superior to that of mathematics, for instance. The reason for this complexity lies in the fact that a superior richness is necessary in each one of the design categories. It is necessary to resort to the repetition of the content but from other perspectives, using the reusability of the information to strengthen grammar, correcting the pronunciation mistakes of the teenagers, for instance.

Acknowledgments

The author wish to acknowledge producers and distributors of hypermedia systems: DeAgostini (Milano, Italy), Electronic Arts (Madrid, Spain), Microsoft (Madrid, Spain) and ZetaMultimedia (Barcelona, Spain). Also thanks to Emma Nicol (University of Strathclyde), Jorge Hernández (Universidad Nacional San Martin), Maria Ficarra (Alaipo & Ainci -Spain and Italy) and Gloria for their help.

References

1. Paymal, N.: Pedagogia 3000. Editorial Brujas, Cordoba (2008)
2. La Noce, F.: E-learning. Franco Angeli, Milano (2001)
3. McLuhan, M., Power, B.: The Global Village: Transformations in World Life and Media in the 21st Century. Oxford University Press, Oxford (1992)
4. Eco, U., Radziewicz, J.: Travels in HyperReality. Harvest, San Diego (1990)
5. Gates, B.: The Road Ahead. Viking, New York (1995)
6. Druin, A.: Designing Online Interactions: What Kids Want and What Designers Know. Interactions 14, 42–44 (2008)
7. Piaget, J.: The Children Machine. Basic Books, New York (1993)
8. Kleinberg, J.: The Convergence of Social and Technological Networks. Communications of the ACM 51, 66–72 (2008)
9. Horn, R.: Mapping Hypertext. Lexington Press, Waltham (1989)
10. Cipolla-Ficarra, F.: Evaluation and communication techniques in multimedia product design for on the net university education. In: Multimedia on the Net, pp. 151–165. Springer, Vienna (1996)
11. Cipolla-Ficarra, F.: Communicability Design and Evaluation in Cultural and Ecological Multimedia Systems. In: Proc. of the 16th ACM International Conference on Mutimedia, New York, pp. 1171–1172 (2008)
12. Cipolla-Ficarra, F.: Heuristic Evaluation of Animated Help in Hypermedia. In: CD Proceedings of the HCI International, Las Vegas (2005)
13. D'Mello, S., Graesser, A., Picard, R.: Toward and Affect-Sensitive AutoTutor. IEEE Intelligent Systems 22, 53–61 (2007)
14. Stephenson, J.: Teaching and Learning Online: Pedagogies for New Technologies. Kogan Page, London (2001)
15. Schlesinger, D.: Working with Open Source: A Practical Guide. Interactions 14, 35–37 (2007)
16. Riehle, D.: The Economic Motivation of Open Source Software: Stakeholder Perspectives. IEEE Computer 40, 25–32 (2007)
17. Nielsen, J., Mack, R.: Usability Inspections Methods. Wiley, New York (1994)
18. Bevan, N.: Practical Issues in Usability Measurement. Interactions 13, 42–43 (2006)

Annex #1: Multimedia Off-line

Multimedia off-line (CD-ROM/DVD)	[c = children, a = adolescent and A = Adult]
- Blue Tortoise. Corel, Ottawa (1995)	[c]
- Business English. DeAgostini, Novara (2001)	[A]
- Corel: Nikolai's Trains. Corel, Ottawa (1995)	[c]
- English Business. ZetaMultimedia, Barcelona (1999)	[A]
- English Communication, ZetaMultimedia, Barcelona (1999)	[a, A]
- English Course I, ZetaMultimedia, Barcelona (1999)	[a, A]
- English Course II, ZetaMultimedia, Barcelona (2000)	[a, A]
- Inglés con Pipo. Cibal Multimedia, Mallorca (1999)	[c]
- Interactive English Academy. DeAgostini, Novara (2001)	[a, A]
- Interactive English Junior. DeAgostini, Novara (2000)	[c]
- Interactive English. DeAgostini, Novara (2005)	[A]
- Interactive English Learning. Microsoft Encarta, Madrid (1999)	[a, A]
- L'inglese. DeAgostini, Novara (2004)	[A]
- Mis Primeras Palabras en Inglés con Pipo. Cibal Multimedia, Mallorca (2006)	[c]
- Passport: Un frasario turistico interattivo. DeAgostini, Novara (1999)	[a]
- The Interactive Alphabet. Corel, Ottawa (1995)	[c]
- The Theatrix Interactive. Stirling Technologies, Mirrabooka (1995)	[c]
- Travel Talk. Libra Multimedia, Windosor (1994)	[A]

Annex #2: Guide of Components for Qualitative English Learning in Multimedia

Design categories: Content = C, Layout = L, Navigation = N, Panchronic = P, Structure = S	
- Access to interactive publications (for instance, digital libraries in the UK and USA).	(S)
- Advertisement board.	(S)
- American English vs. British English.	(C)
- Chat and video chat.	(S)
- Classical games: crossword puzzles, trivia quizzes, jigsaw puzzles, etc. and special developments (action and adventure, driving and racing, strategy and simulation, sports, etc.).	(C)
- Cultural references.	(C)
- Explanations of slang, idioms and the most common proverbs.	(C, L)
- Forums.	(S)
- Integrated dictionary (visualization of the content by pressing on the word and even being able to listen to its sound).	(C, L. P)
- Interactive theatre –possibility of choosing the characters, their voices, the stage setting, etc	(N, S)
- Karaoke system for reading.	(C, L, P)
- Linguistic lab: voice matching, that is to say, pronunciation is compared until a correct pronunciation is reached.	(N, P, S)
- Error analysis –distinction between orthographic and content errors.	(C)
- Personalization of the contents.	(C, L)
- Real situations or real contexts for the use of the language.	(C, L)
- Self-assessment tests to define a personalized study plan.	(S)
- Simultaneous translation.	(P, S)
- Smart pitch control: to improve the speed of pronunciation.	(S)
- Teachers on-line to provide the corrections.	(S)
- Replacement of a character in a spoken dialogue.	(N, S)
- Typical expressions in every situations.	(C)
- Video conferences.	(S)
- Virtual tutors.	(S)
- Videoteque.	(S)
- Voice Pilot (to speak with the virtual tutor).	(N, S)

Recognizing objects in images using visual attention schema

Hoai-Bac Le and Tran-Thanh Thang

Abstract Behavioral studies of visual attention have suggested two complementary modes of selection. In a space-based mode, locations in the visual field are selected; in an object-based mode, organized chunks of visual information - roughly, objects are selected, even if the objects overlap in space or are spatially discontinuous. Although the two modes are distinct, they can operate in concert to influence the allocation of attention. In this paper we unify these two attention modes and create our attention schema to recognize objects in images. The object-based attention mode works after the space-based attention mode. With the method of rotating objects at their centrals, we define object parts and create our object-based networks. We have experimented with 27 objects. Each object consists of two to six parts. Our program can recognize objects in color images. The accuracy of the chosen object from our attention schema is higher than from program that only uses space-based network.

1 Introduction

Behavioral studies of visual attention have suggested two distinct and complementary modes of selection, one involving space and the other objects. In a space-based mode, stimuli are selected by location in the visual field [9]. In contrast to the space-based mode, evidence has also been found for an object-based mode in which attention is directed to organized chunks of visual information corresponding to an object or a coherent form in the environment, even if objects overlap in space or are spatially discontinuous. All visual features of an attended object are processed concurrently, and features of an attended object are processed faster and more accurately than features of other objects [2]. Object-based attention operate on a structural description of an object - a description that decomposes an object or scene

Faculty of Information Teachnology, University of Science,
HCM City, Vietnam
e-mail: `lhbac@fit.hcmuns.vn` and e-mail: `thanhthangbmt@yahoo.com`

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 129–144.
springerlink.com

into its parts, and that characterizes the relationships among the parts in terms of multiple all centric frames of reference [1]. Both of these attention modes coexist in the visual system and influence one another. Many studies continue to address the relationship between these two modes of selection as if one mode of selection is more important than the other. In this paper, we use Lavie and Driver's suggestion [5] that is a space-then-object account by demonstrating that object-based effects occur only within the focus of spatial attention.

We use Hierarchical Temporal Memory - HTM [3] to create these attention modes. When the HTM network is used in image processing, it becomes the *HTM space-based network*. In order to create the object-based in our schema, we create our objects in 3-Dimensional space and divide them into parts. Then we rotate them at their centrals to make the our data sets. We create two kinds of HTM networks: HTM space-based network (HTM-SBN) and HTM object-based network (HTM-OBN). We use HTM-SBN to identify object names and filter a number of objects for the HTM-OBNs. HTM-SBN is trained by image of object that consists all parts. We built one HTM-OBN for each object. This kind of networks is trained from the data sets of object parts. After all, we combine these two kinds of networks to make our attention schema and experience it.

All HTM systems have some type of sensory input, even if the data is coming from a file. There are two essential characteristics of sensory data. *First*, the sensory data must measure something that is directly or indirectly impacted by the causes in the world that we might be interested in. *Second*, the sensory data must change and flow continuously through time, while the causes underlying the sensory data remain relatively stable [3].

In order to make the causes change while the object is still stable, the original method try to move that object from left to right then top to bottom in the image space [6]. This method has some advantages: It is very simple to make training data sets and we can copy data from trained nodes to untrained nodes because the object coincidences do not change. However, this method has some disadvantages: the number of object coincidences is less and they are often drawn by hand, so that they do not express every coincidence for a real object. In this paper, we introduce a method to move our object also satisfy HTM system's conditions. We created our objects in 3-Dimensional space and determined their centrals. Then we rotated them at their centrals to make the changing and got our image data sets. Because of making in 3-Dimensional space, our objects have most of the real object attributes, and this method creates our data sets automatically. By this way, we can divide our whole object into parts, and get data sets for each part separately. These data contain correlation information between parts of each object. From these data sets of object parts, our network will learn the object attributes. We use HTM network to train the object parts so that we call them *HTM object-based networks*.

We create our attention scheme by combining HTM space-based network and HTM object-based networks. HTM-SBN uses the whole object in image to predict object names. It gives us some objects result from an input image. Beside of that, we segment that input image into several parts. HTM-OBNs correlates to those object names will calculate their part values. After all, we calculate the value for each object from its part values. The wining object is an object that has greatest value. The structure of this paper has four main sections. In the first section, we describe the method of rotating our objects at their centrals and get our data sets. In the next section, we create the HTM space-based network (HTM-SBN) and HTM object-based networks (HTM-OBNs). After that, we combine these two kinds of networks to create our attention schema then explain its work. In the third section, we experience from this attention schema results for recognizing objects in images. At the end, we conclude and discuss the further works for this study.

2 Rotate the objects at their centrals

We create our objects in 3-Dimensional spaces. Each object consists of two to four parts. Each part has different color from the others. For example, the object 'Table' has two parts: the 'Table top' is red and the 'Table legs' is yellow as describe in Fig. 2. We determine each object central in space Oxyz. After that we rotate it around Oy-axis, the total rotating angle is 360^o. Our capturing camera moves from angle 0^o to 45^o on the axe Ox as describe in Fig. 1. All output frames are of 64 pixels by 64 pixels. The data set for each object (or each object part) contains 2400 continuously frames.

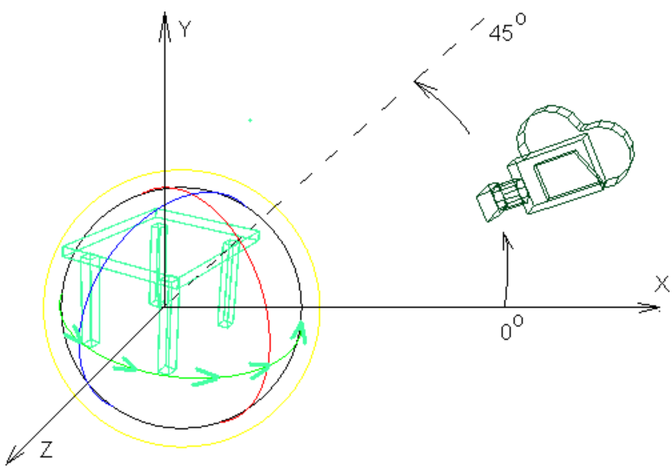


Fig. 1: Rotate the object at its central and get image data sets.

We can add or remove each object part easily from the scene in order to create its data set. We make the color image data sets for testing. From these data sets, we convert them to binary image data sets (the object-pixels are black and the background pixels are white as showed in Fig. 2). We use these binary data sets to train our HTM networks. Our binary image data sets are divide into two types: data set Object_Full contains image data of objects consisting all their parts (e.g. 'Table_full') and data set Object_Part contains image data of objects consisting only one object part (e.g. 'Table_top' or 'Table_legs'). We use data set Object_Full to train HTM-SBN and data set Object_Part to train HTM-OBNs

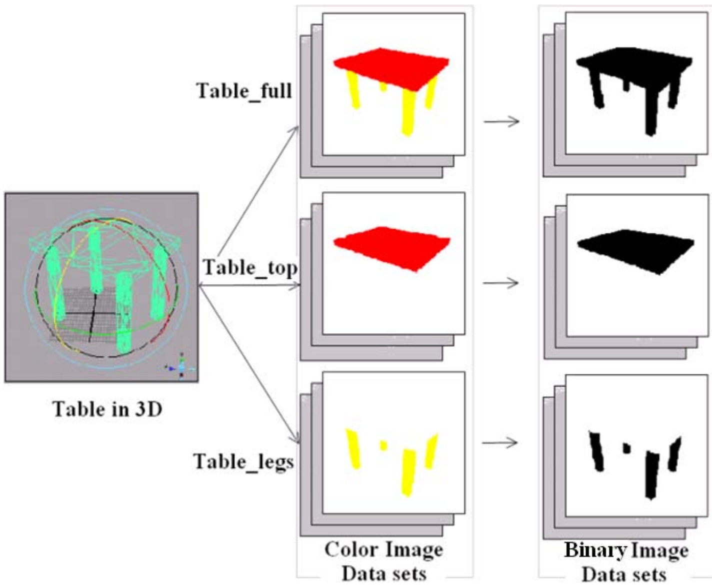


Fig. 2: : The object 'Table' rotates in 3-Dimensional space and creates three color data sets 'Table full', 'Table legs' and 'Table top' then we convert them to binary image data sets.

Fig. 3 shows an example output images using our rotating method then we convert them to binary format. We rotate object 'Table' at its central. The 'Table' rotate from 0^o - 0^o- 0^o to 0^o - 79.8^o - 0^o, the camera moved from 103^o - 0^o - 0^o to 103^o - 7.7^o- 0^o. This Figure shows nine frames between 0001 and 0267.

2.1 Attention schema

This section we create HTM-SBN and HTM-OBNs. Both HTM-SBN and HTM-OBNs have same structure. HTM-SBN is train from data set Object_Full, this network is used for recognize the object names. We create one HTM-OBN for each object. These networks will calculate their object parts. After all, we create our attention schema by combining these two types of networks then we describe its work.

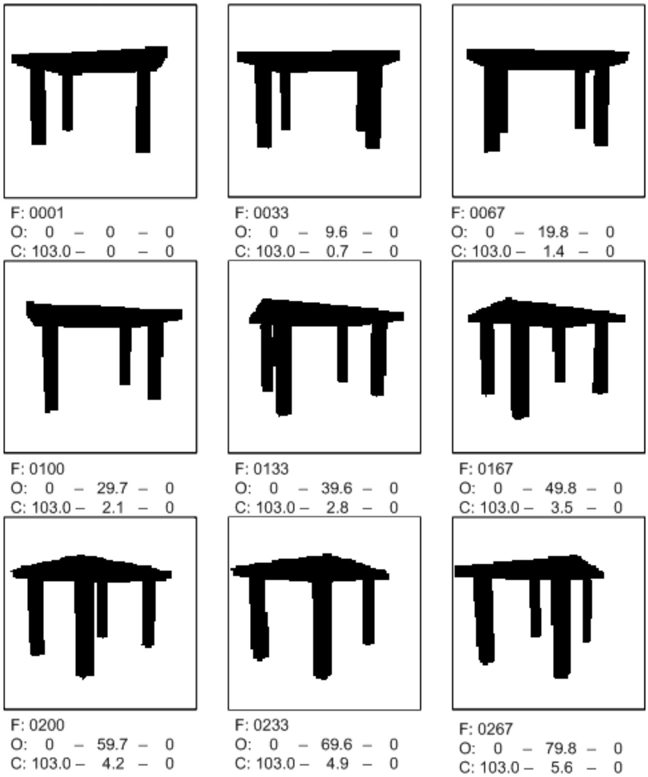


Fig. 3: An example output frames of object 'Table': F (frame): from 0001 to 0267, O(Object): the object movement, C (Camera): the camera movement.

2.2 HTM network structure for recognizing image

This HTM structure has six hierarchical layers. Its input is fed in at the bottom level. Nodes are shown as squares. The top level of the network has one node, lower levels has 4, 16, 64, 256 and 1024 nodes. The size of an input image is 64 pixels by 64 pixels. This image is divided into adjoining patches of 4 pixels by 4. Each bottom-level node's input corresponds to one such 44 path. A node at a higher level receives its input from several nodes at the lower level. As in Fig. 4, a Level_6 node receives its input from the output of four Level_5 nodes.

The size of the receptive field for the Level_1 nodes is 4x4. The Level_2 nodes have a receptive field of size 88 because their inputs come from four Level_1 nodes, and so on for each upper level. The size of the receptive field of a node increases as we go up in the hierarchy. The node at the root of the tree covers the entire visual field, by pooling inputs from its child nodes.

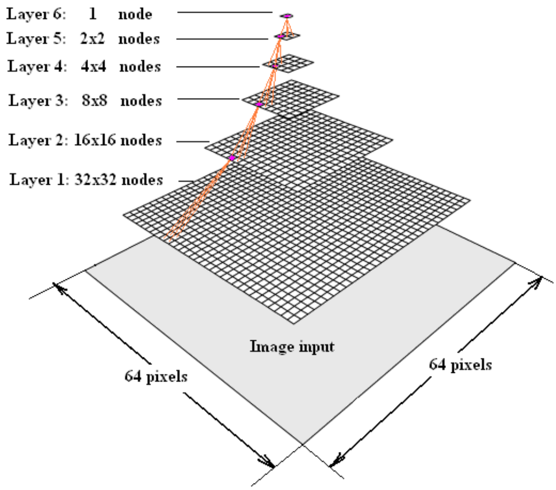


Fig. 4: HTM structure for image 64x64.

All nodes in this network use learning and inference algorithms based on the same principles. Each node contains two modules to perform the above two steps: spatial pooler and temporal pooler. These both modules must undergo a learning phase before they switch to an inference phase. Ref [7] details the interface and implementation of algorithms for learning and inference. These algorithms are performed by Zeta1 nodes. These nodes use temporal information during learning. They look at sequences of input patterns and remember which patterns occur together in time. All learning and inference processing is handled in a feed-forward manner, and no downward links are used in this hierarchy.

Zeta1 nodes support both supervised and unsupervised learning. All nodes except the top node learn in an unsupervised fashion. All algorithms are used in this network as following:

- *Unsupervised learning at lower and middle levels*: In spatial pooler, we used algorithm Gaussian for nodes in Layer_1 and algorithm Dot for nodes in other layers; in temporal pooler, we used algorithm maxProp for all nodes.
- *Supervised learning at the top of a hierarchy*: we used Dot Spatial Pooler algorithm for this top node.

2.3 HTM space-based network

HTM-SBN has HTM structure as describe in sub-section A. The purpose of this network is that it can recognize object consisting all its parts. We used data set

Object_Full to train this network. The output of this network is a prediction vector. The length of this vector is a number of objects in our data sets. In Fig. 5, we trained HTM-SBN with 27 objects, each of them has all its parts. The pyramid is our HTM network. The output of HTM-SBN is a prediction vector. We will choose m top object (which have highest values) from this vector for our attention schema.

2.4 HTM object-based network

HTM-OBN has HTM structure as describe in sub-section A. We create one HTM-OBN for each object. The purpose of this kind of networks is that they can recognize object parts. When we input an image contains object consisting only one part to this network, it will output a prediction vector. From it, we can calculate the degree this part belongs to that object. We used data sets Object_Part to train these networks. The output of these networks is a predition vector. The length of the vector output is the number of object parts. As describe in Fig. 6, we create one HTM-OBN for the object 'Computer'. This object has three parts: 'Computer_Moniter', 'Computer_CPU' and 'Computer_Keyboard'. So that, we have three data sets to train this network. The output of this HTM-SBN is a vector of length three. In this figure, we input an image containing a 'Computer_Moniter', so that in our vector output, the value of 'Computer_Moniter' is the highest.

2.5 Attention schema

In this section, we combine HTM-SBN and HTM-OBNs to make our attention schema, and then we explain its work. In section four, we experience the result from this schema. At first, HTM-SBN recognizes m top objects from a number of objects in our data set. After that, the HTM-OBN of each object in these m objects will calculate their part values. Fig. 7 describes our attention schema. Attention field A has many circles having the same centre, but they have different radius. The space in these circles has different values to the others. The circle that has shorter radius has greater value. The circle space value increases from 10 to 100. We have six blue circles in attention field A. These circles are our objects $O_i(i = 1|m)$. Each of these objects consists of two to four parts. We describe them as black circles $P_j(j = 2|4)$. HTM-OBN calculate the value for each object part P_i and then put this part to the right position at circle space in attention field A corresponding to its output value. We calculate each object value from its part values. The object will be attention (or chosen for the output) if its circle central position is nearest to the attention field center A. It means that, that object has the highest value.

When we input an image, HTM-SBN will give us m objects to our attention field A. These m objects are $O_i(i = 1|m)$. Each object O_i uses its HTM-OBN to calculate its parts value $P_j(j = 1|k)$ (O_i has k parts). After that, we calculate the value for each object from its part values.

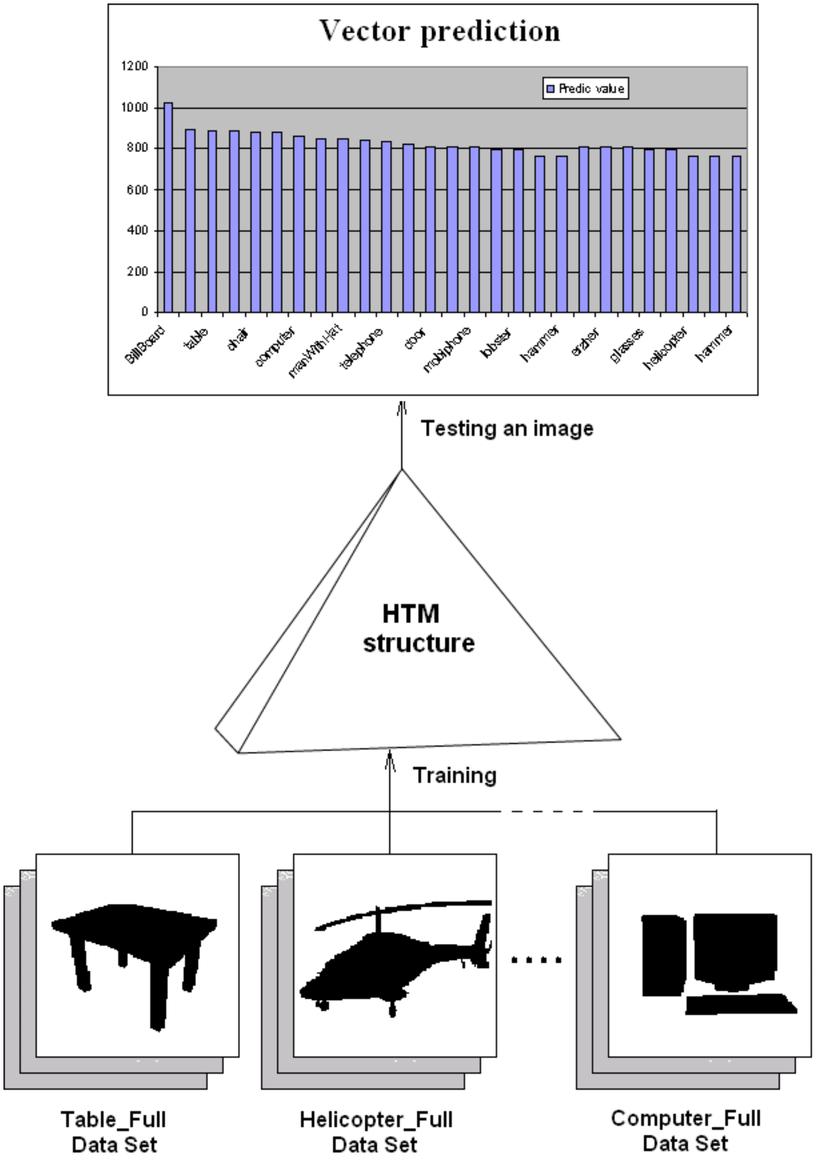


Fig. 5: Training HTM-SBN.

The output of HTM-OBN of object O_i is a prediction vector $\overrightarrow{v} = a_1, a_2 ... a_k$. The value for each part P_j is the highest value in vector $\overrightarrow{v}$:

$$Value(P_j) = max(a_t)_{t=1}^{k} \tag{1}$$

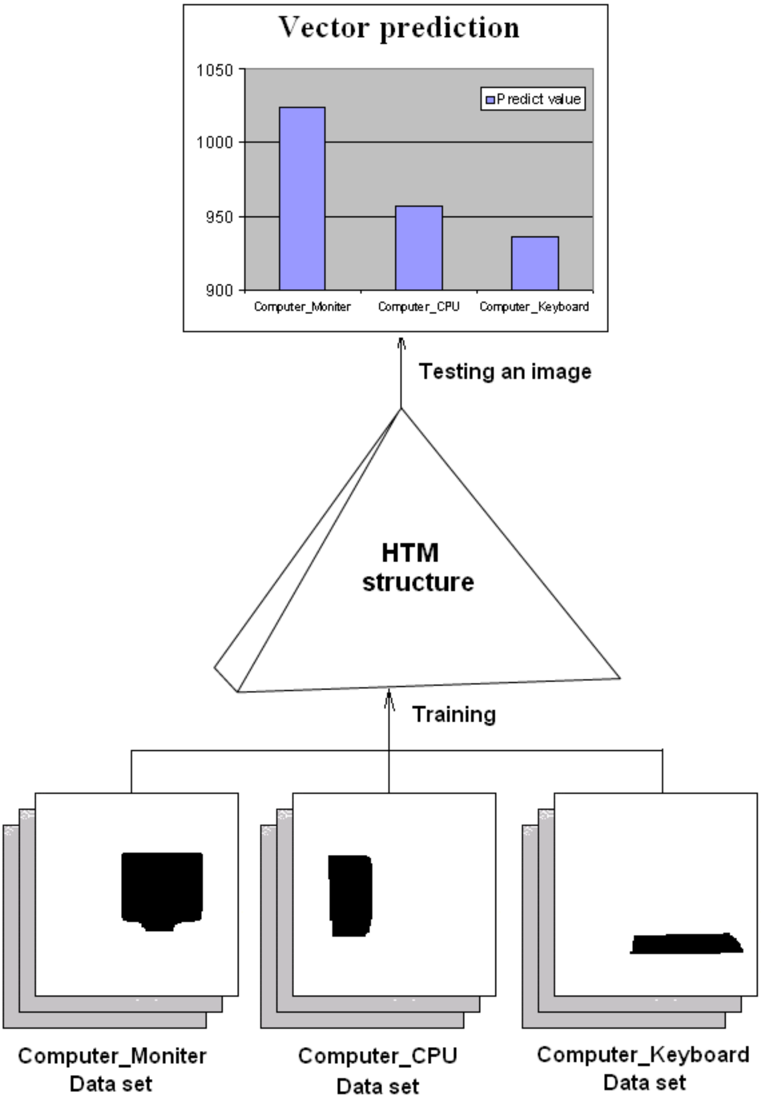


Fig. 6: Training HTM-OBN.

The value of object O_i will be calculated by:

$$Value(O_i) = \frac{1}{k}\sum_{j=1}^{k} Value(P_j) \tag{2}$$

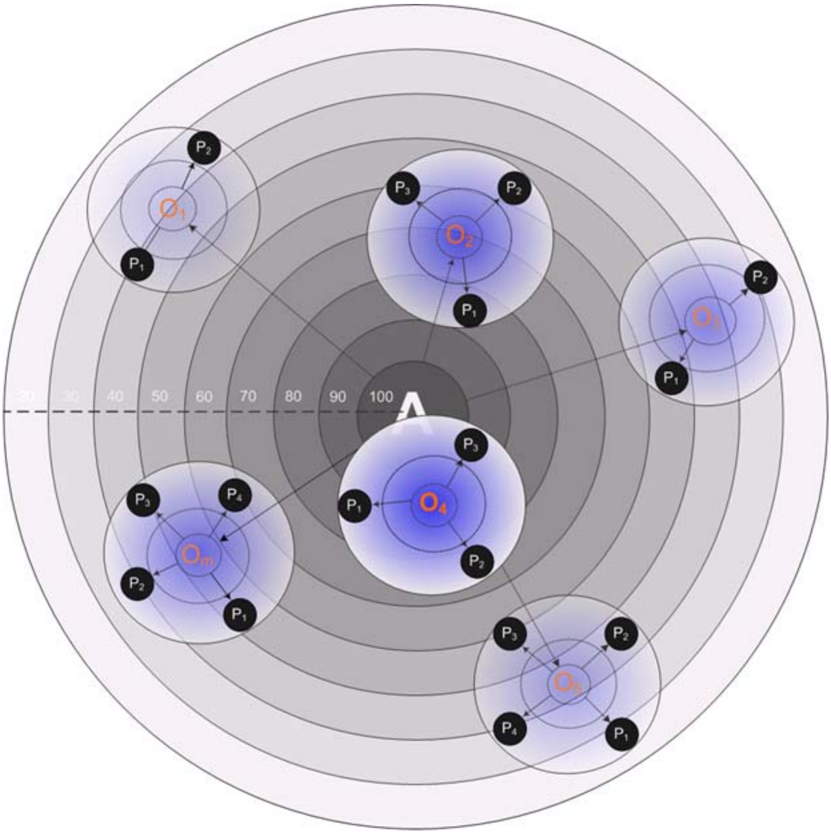


Fig. 7: Attention schema.

At the end, the object that has greatest value will be the object result output. In Fig. 7 object O_4 has three parts, all its parts have high values, so they are put near to the central of attention field A. This object has greatest value. Its position is the nearest to the central of attention field A. Object O_4 is our output result.

3 Experience

In this section, we create our program using attention schema. We compare our result to the program that only use space-based network. It recognizes our objects better when the number of objects part increases.
We experienced our attention schema with 27 objects. Each object consists of two to four parts. Each part has different color from the others. Each object part has 2400 images. In order to segment our color images, we used the implementation of pyramids segmentation algorithm [8] by OpenCV. The links between any pixel *a* on level *i* and its candidate father pixel *b* on the adjacent level are established if the Euclidean RGB color distance is below 'Threshold_1'. After the connected components are defined, they are joined into several clusters. Any two segments A and B belong to the same cluster, if the Euclidean color distance between the average of the connected components is below 'Threshold_2'. We follow it with a connected-components procedure to get separate continuous segments. See more details in the OpenCV manual.

After that, we converted color image data sets to binary image data sets to train our HTM-SBN and HTM-OBNs. We used color image Object_Full data sets to test our program. The recognition accuracy on this training data set was 100 percent. The accuracy for the chosen object from the vector output is higher than the program that only uses space-base network method. We input an image of object 'Computer'. This object has three parts: 'Monitor', 'CPU' and 'Keyboard'.

- We segment this image into parts, then we convert them into binary images.
- We convert our original input image to binary image to input to our HTM-SBN.

The output of HTM-SBN is a prediction vector. We choose top m objects that have highest values from this vector. These m object will determine m HTM-OBNs $O_i(i = 1|m)$.

Each part of object in our image data has different color from the others, so that, we segment that image into parts. After that, we test each part in HTM-OBNs. The Fig. 9 explains in more detail. HTM-OBN of each O_i calculates its object parts values from equation (1). The value of object O_i will be calculated by equation (2). After calculating for all object $O_i(i = 1|m)$, we compare the object values. The object that has greatest value will be the object result output. As described in Fig. 9, the object result output is 'Computer'. We compare our program results with the space-based network program. The recognition accuracy of space-based network program on training data set was 100 percent. However, the accuracy for the chosen object from its vector output is lower than from our attention program. The accuracy for each object from the vector output $\overrightarrow{t} = O_1, O_2 ... O_m$, calculated by:

$$Accuracy(O_i) = \frac{Value(O_i)}{\sum_{j=1}^{m} Value(O_j)} .100\% \tag{3}$$

We chose $m = 5$ and used our training data sets for testing. The accuracy of the chosen object from attention schema program is more than three times than from space-based program. The average accuracy from attention schema program is 77.1% while from space-based program is 22.3%

Fig. 9(a) and Fig. 10(a) show an example output of HTM-SBN and our attention schema program when we test an image contains object 'Billboard'. The accuracy for each object value from HTM-SBN is showed in Fig. 9(b). The accuracy of chosen object 'Billboard' is 22 percent. The accuracy of object 'Billboard' from attention schema program is 80 percent as showed in Fig. 9(b).

The average prediction accuracy for the chosen object increases when the number of object parts increases as described in Fig. 11. However, in this experience, we only use objects that consist of two to four parts, so that, how many parts for an object this rule still be true it can be further research. We will discuss more about this problem in the next section.

4 Discussion

We rotate the objects at their centrals to get our data sets. It means that it only recognizes the right object when that object appears at the center of input image and it has a fixed size to object that our network had learnt. In order to take advance from this attention schema we should preprocess the input image. The object should be moved to the center and be scaled to fix the size of object has learnt before. We also need more power segmentation algorithms to separate an input image into the right object parts for our schema.

For some angles of objects, the shape of two completely different objects can be alike because our input images for space-based network are in binary format. However, our testing images consist of two to four object parts with different colors, so that, when we segment it into parts, the shapes of these parts from each object are different from the other. So that, the object-based networks can classify these objects correctly and solve this problem. In more advance, the camera should capture our objects in the angles that people often do.

In this paper, we did not use object outline as usual as HTM use in their Vision Project. Using object outline in image, their program is easy to fail if the testing object moves a little away from the image center, so they have to move their object around the original position. The reason is that their object outline in testing image does not fit to the object outline the space-based network has learnt before. So that, instead of using object out-line, in our schema we used object full filled with color (our image data sets is still in binary format). Our program still recognizes the right object even it was moved a little away from the input image center. However, in the future research, we should use grayscale format instead of binary format for our image data sets.

In this attention schema, we also used HTM-SBN to recognize *m* objects at first, after that, we calculate their parts then the object values. So that, our schema may go wrong if the right object is not in the set of *m* objects output from HTM-SBN. This is our attention schema weakness.

Hamilton (1859) first raised the question of the span of attention:

> '... *Suppose that the mind is not limited to the simultaneous consideration of a single object, a question arises, How many objects can it embrace at once?'*

We do not discuss this question in this paper, we just chose maximum five objects into attention field and each object consists of two to four parts. With these conditions, our experience gives us a rule: *'The average prediction accuracy of the chosen object increases when the number of object parts increases'*. However, our attention schema can be developed for each object part *P* if we separate it into sub-parts SPs. At this higher layer, this part P will be the attention field and each sub-part SP will become a sub-object in attention field P. So that, our attention schema can have multi-layers as mentioned in Ref [10].

All HTM networks recently have to train from the beginning if we want it to learn a new object. In our attention scheme, we only has to train (N+1) objects for the HTM-SBN from the beginning and we create one HTM-OBN for that new object. We reuse all HTM-OBNs of N old objects that we have trained before. This is an advantage of this schema. However, in further work we should develop our HTM algorithms for our network that can be retrained when it learn a new object.

This attention schema is applicable to a broad class of problems from machine vision, to fraud detection, to semantic analysis of text. However, we should have more research to put this attention schema to become a real-world object recognition program. In this schema, we use the HTM framework and we hope in the near future the HTM algorithms can be improved and get more powerful to learn and relearn objects in the real-world.

5 Conclusion

The mechanisms of object-based and space-based visual attention have been widely investigated in psycho-physics and neuroscience research, however, modeling visual attention in computer vision is a quickly growing field, especially for building computable models of covert attention. Until now, these are only some computational models for object-based attention has been developed like Koch and Itti's saliency-based attention model [4].

We have presented a computable model for our attention schema. By using object rotation method, we define the object parts and developed object-based network from space-based network. We combined space-based and object-based to make our attention schema. The experimental results showed that the accuracy of the chosen object from our attention schema is higher than the program that only uses space-base network.

In this paper, we borrow ideas from visual attention and the framework of Hierarchical Temporal Memory to make a computational model. Besides the strengths of our model, there are still several limitations that we did not present here a complete theory of goal-driven effects on visual attention, which is necessary for understanding visual attention. This is still an open field for further research.

References

1. Biederman, I.: Recognition by components: a theory of human image understanding. Psychol. Rev., 115–147 (1987)
2. Duncan, J.: Selective attention and the organization of visual information. J. Exp. Psychol. Gen., 501–517 (1984)
3. Hawkins, J., George, D.: Hierarchical temporal memory: Concepts, theory and terminology (2008), http://www.numenta.com/Numenta_HTM_Concepts.pdf (18/12/2008)
4. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for repid scene analysis. IEEE Trans. Pattern Anal. Machine Intel. 20(11), 1254–1259 (1998)
5. Lavie, N., Driver, J.: On the spatial extent of attention in object-based visual selection. Percept. Psycho-phys., 1238–1251 (1996)
6. Numenta, Inc. 'Advanced NuPIC Programming', Ver 1.8.1, p. 94, http://www.numenta.com/for-developers/software/pdf/nupic_prog_guide.pdf (September 2008)
7. Numenta Inc., 'Zeta1 Algorithms Reference', Version 1.5, Patents pending (2008, http://www.numenta.com/ (18/12/2008)
8. Burt, P., Rosenfeld, H.T.: A Segmentation and Estimation of Image Region Properties Through Co-operative Hierarchical Computation. IEEE Tran. on SMC 11(12), 802–809 (1981)
9. Posner, M.I.: Orienting of attention. Q. J. Exp. Psychology, 3–25 (1980)
10. Sun, Y., Fisher, R.: Object-based visual attention for computer vision (2002)

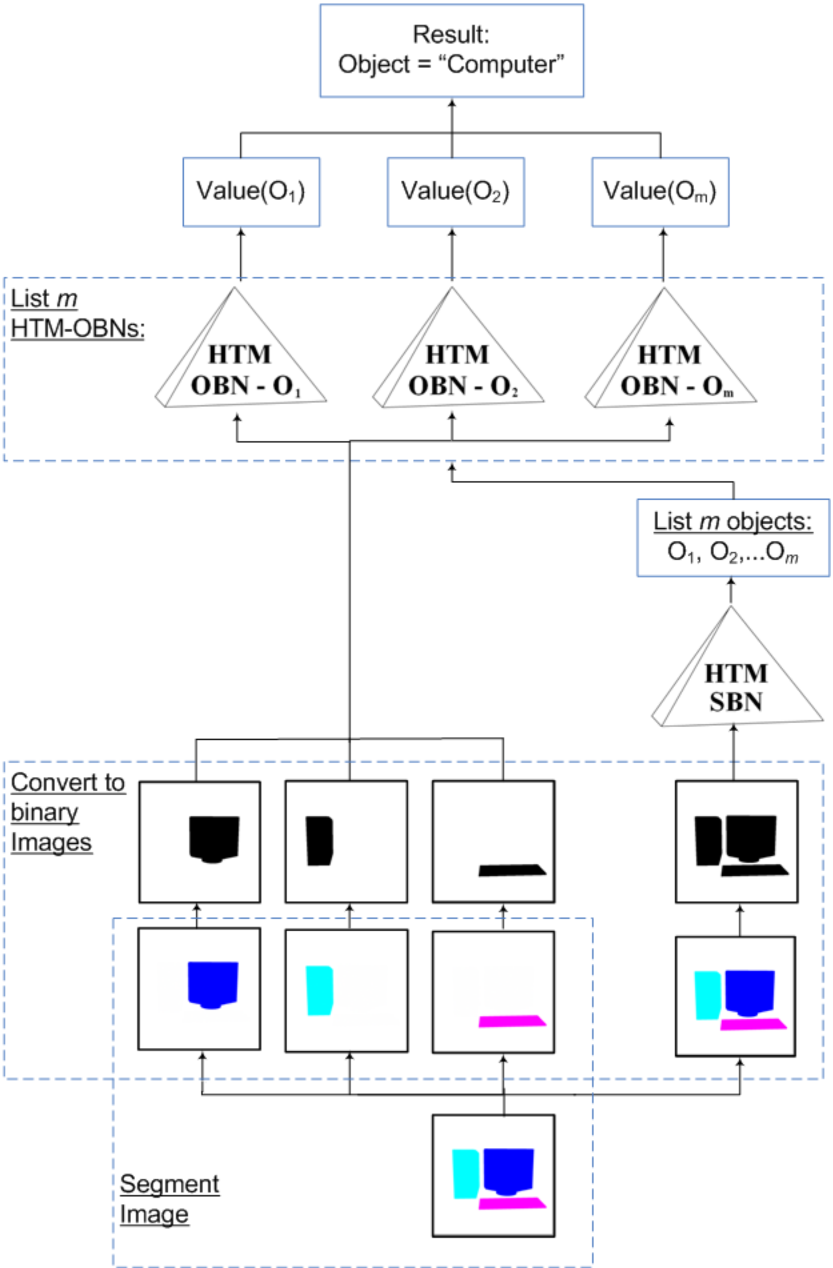


Fig. 8: Diagram of attention schema program.

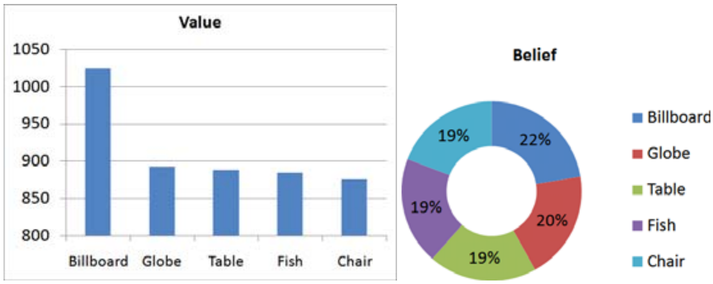


Fig. 9: (a) An example prediction vector output from HTM-SBN for object ‘Billboard’ – (b) An example accuracy values for objects from HTM-SBN with input object ‘Billboard’.

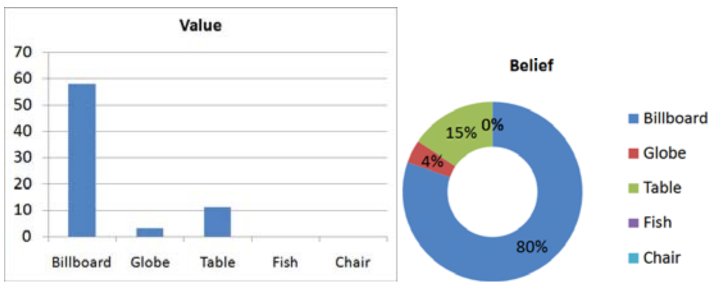


Fig. 10: (a) An example prediction vector output from attention schema for object ‘Billboard’. (b) An example accuracy values for objects from attention schema with input object ‘Billboard’.

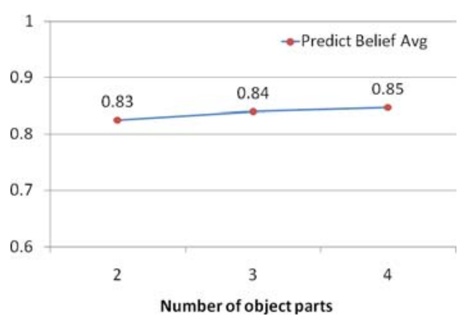


Fig. 11: Average prediction accuracy value for the chosen object and its number of parts.

Towards a Comprehensive Data Fusion Architecture For Cognitive Robotics

Malek Baklouti, Jamil AbouSaleh, Bahador Khaledgi and Fakhreddine Karray

Abstract A generic cognitive robotics framework should integrate multimodalities to preserve consistency, minimize uncertainty, and adopt human like concepts in order to achieve efficient interaction with the operator. Fusion is the process of combining observations, knowledge, and data from multiple sensors into a single and coherent percept. There are several data fusion architectures existing in the literature, nevertheless, a complete and unified architecture for data fusion is not in the picture yet. In this paper, we present a new data fusion architecture pursuing the same goal of realizing such generalized architecture initiated by JDL (Joint Director's of Laboratories). The proposed architecture comprises two degrees of freedom represented by three levels of abstractions, and four layers of situation awareness. We also suggest incorporating a cognitive memory model that best suits our targeted robotics applications.

1 Introduction

Most service robots deploy a multi-modal strategy that combines range of sensors to extract key features from the operator. When dealing with disparate sensors, several fundamental questions arise: what is the utility of sensor fusion and why to integrate multiple modalities? How to effectively integrate multiple modalities and combine the measurements provided by these sensors?

When using a single sensor system, there could be missing observations when the sensor cannot measure all relevant attributes of the percept and ambiguity might occur. This can be defined as perception uncertainty and presents the major disadvantage of a single sensor system in the way that it is unable to reduce uncertainty

Malek Baklouti
Thales Training and Simulation,France , e-mail: malek.baklouti@thalesgroup.com

Jamil AbouSaleh, Bahador Khaledgi and Fakhreddine Karray,
e-mail: {jaabousa,bkhaleghi,karray}@waterloo.ca
PAMI Lab, University of Waterloo

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 145–155.
springerlink.com

in perception. The need to cope with this uncertainty constitutes a major challenge in robotics. One way to tackle this issue is to use active perception techniques. For instance, active vision methods attempt to reduce the uncertainty related to an observed object by adequately moving the camera to have clearer view of that object. However, different views may be insufficient to make up for incomplete or ambiguous observations. Multiple sensors are then needed to reduce uncertainty.

Although, in some cases, it might be enough to deal with only one modality, having more than one mode is a necessity in cognitive robotics since the interaction between humans and their environment is naturally multimodal. For example, if we want to order (in a natural way) a robot to pick up and bring an object, we would look, speak, and point towards that object at the same time. Therefore, a generic cognitive robotics framework should integrate multimodalities to preserve consistency, minimize uncertainty, and adopt human like concepts in order to achieve efficient interaction with the operator.

Fusion is the process of combining observations, knowledge, and data from multiple sensors into a single and coherent percept. The main issue is then how to integrate information gathered by various sensors efficiently so that we overcome incomplete, inconsistent and imprecise knowledge. The initial methods for sensor data fusion approach this problem from a mathematical perspective and deploy evidential/probabilistic models to solve this problem [3, 14, 4, 5, 17]. More recent methodologies adopt a systematic approach and rely on modular architectures [15, 9, 11]. This would in turn extend the range of applicability of the data fusion scheme by basing it on general and interactive principal modules. There are several data fusion architectures existing in the literature [10, 15, 8], nevertheless, a complete and unified architecture for data fusion is not in the picture yet. In this paper, we present a new data fusion architecture pursuing the same goal of realizing such generalized architecture initiated by JDL (Joint Director's of Laboratories).

In the remaining of this paper, we first provide a formal definition of the data fusion process, and later in section 2, we enumerate some of its applications with respect to cognitive robotics. In section 3, the principles, advantages, and shortcomings of the previously proposed high-level architectures are discussed. Then, in section 4, we propose a new data fusion architecture, which is an attempt to integrate the contributions of the major existing architectures into a unified general framework for data fusion. section 2 concludes this paper. An insight about some future work is also presented in this section. It is important to note, however, that although our discussion is geared towards cognitive robotic applications, the architecture itself is general (as intended to be) and might be applied in wide variety of different domains.

2 Data fusion for cognitive robotics

2.1 Data fusion definition

The JDL Data Fusion Subgroup gave a concise definition of data fusion as ”the process of combining data to refine state estimates and predictions” [15]. However, there is a misnomer in the literature when speaking about fusion. In fact, we can distinguish different types of fusion: (1) data fusion where numerical entities are combined to produce more informative data, (2) sensor fusion where different devices are grouped under one model [7], (3) information fusion where multiple data are fused to create the knowledge base [6], and (4) cognitive fusion that integrates the information in the user's mental model [15, 10]. In the area of robotics, the fusion framework is expected to have the ease of implementation, the ability to adapt to new sensing configurations, the flexibility to adjust in response to major changes in the external environment, as well as the ability to determine its own sensing strategy. Sensor data fusion and integration can occur at multiple levels from early input level to intermediate level to decision level. This integration takes into account the nature of sensors fused (homogeneous or heterogeneous) as well as the prior knowledge of the environment and the human input when applicable. The interaction between the sensors can be complementary, competitive or coordinative [2]. In the area of cognitive robotics, the fusion process is usually deployed to support robot actions by refining and reducing the quantity of information that the robot needs to examine to achieve timely, robust, and relevant assessments and projections of the situation. Cognitive fusion is inspired from cognitive information processing; and attempts to fuse data in the same way as humans do. It is defined as the process of merging multi-sensor imagery based on neural models of visual processing and pattern recognition [12].

2.2 Application in cognitive robotics

As mentioned in section 1, our architecture is general in principle but geared towards cognitive robotics. Several applications of data fusion schemes are described in order to highlight the need for different components to be incorporated in our architecture. The applications of data fusion are pervasive and underlie the core problem of sensing, estimation, and perception. In their work [19], the authors overview and analyze the status of multi-sensor information fusion technology applied in robotics field. Multi-sensor information fusion technology can increase the rate of the intelligence and is essential for designing cognitive robots.
Data fusion is very instrumental in human machine interaction where multiple (often complementary) sensor data are combined before making any decision. Song et al. [13] propose an emotion recognition system that extracts visual and auditory features from sensor data and deploys an SVM-based classifier to perform bimodal

information fusion. Another example is the multi-modal object attention system presented by Fritsch et al. [1]. The system's architecture comprises three levels: deliberative, intermediate, and reactive, and is able to identify objects referenced by the user with gestures and verbal instructions.
Deliberative level, also called planner-based control, uses sense-plan-act decomposition to construct an intelligent system. It generates action sequences initiated by the user speech. The architecture is complemented by perceptual components in the reactive layer, and a module for storing internal representations in the intermediate layer. The perceptual components process sensor data to gather information about the environment, which is then, stored as internal representation together with additional information acquired from the process of interacting with the user.
Another interesting application for data fusion takes place in the area of autonomous robot navigation. Moving robots use multi-sensor information fusion technology to construct a model for the unknown environment, and navigate steadily even under unstructured environment. Tan et al. [16] propose a hybrid modular system for sensor fusion in autonomous robot navigation. The modules are implemented as agents in different layers: reactive layer, knowledge layer, and social knowledge layer. The middle level (knowledge layer) contains projective agents that develop plans and make decisions for a long-term purpose. Conversely, the reactive agents are sensor-based, and are responsible for real-time performance.
These are only a few of the applications of data fusion. But generally speaking, one can draw the conclusion that the information provided by the sensors in different states and locations, eliminates the redundancy and contradiction among the multi-sensor information, reduces uncertainty for human sensing, thus making use of information complementary, and provides relatively comprehensive and coherent perceptive description about the surroundings for path planning.
Another general conclusion can be made about the different data fusion applications: Memory/Database is an instrumental component in any data fusion system, as it stores current sensor data, extracted features, and previous knowledge about the domain which might be used in inferring a decision once the fusion process is completed. The stored information can be further modified and adjusted to adapt to new situations and external variations. It can be also concluded that fusion is not restricted to one level or another, as it might occur at one or more layers, and following different schemes as well. Finally, we would not forget the importance of projecting a current situation into the future to predict the consequences of any decision that might be taken, and using that information to refine the produced decisions. Therefore, any proposed high level architecture should highlight such factors before addressing further details.

3 Review of generalized data fusion architectures

Several data fusion architectures have been proposed in literature. However, in this section, we will focus on the most recent, and famous ones, JDL, Endsley as well as Jakobson, as they have widely used in many work, and our proposed architecture in this paper is inspired by such work, as a complete and unified architecture for data fusion is not in the picture yet. In the following subsections, both JDL and Endsley architectures will be presented in details.

3.1 JDL Architecture

Joint Director's of Laboratories (JDL) were the first who proposed an influential model for data fusion. It is a functional model that illustrates the primary functions, relevant information, databases, and interconnectivity needed to perform data fusion. They divided the process into 5 levels as shown in figure 1. In this model, Steinberg and al. [15] presented different levels in a generalized manner, thereby making them applicable in various contexts. Several sensor fusion categories/levels were defined :

- Level 0: Data Alignment
- Level 1: Entity Assessment (e.g. signal/feature/object)
- Level 2: Situation Assessment
- Level 3: Impact Assessment
- Level 4: Process Refinement (i.e. sensor management)
- Level 5: User Refinement

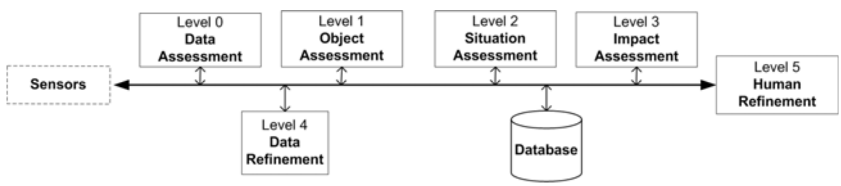


Fig. 1: JDL fusion Model [15]

Level 0 or Data Assessment provides a local access and fuses the low level data (Direct signal, raw bit, pixel values). Level 1, however, deals with objects which are tracked and identified over time. A measure of correlation is then applied to infer a relation between the objects, and build a knowledge base in level 2 (Situation Assessment). The potential outcomes of a planned action are then estimated in level 3 or Impact Assessment. Data Refinement and User Refinement levels provide the system with feedback. At its beginning, the JDL model was more focused on military applications. However, many researchers have successfully attempted to map the model to other applications such as bioinformatics, robotic application etc.
Although the JDL model is general enough to be applicable to problems within different domains, two main drawbacks can be highlighted:

1. It's still hard to see the relation between different layers, as no clear communication protocol between the different levels is presented. This relation could be complementary, hierarchical, or sequential etc.
2. JDL architecture does not explicitly mention the decision and action phases, and how action sequences are generated.

3.2 Endsley architecture

Another important architecture is proposed by Endsley [9]. The author claims that the information processing mechanism is altered by individual's abilities, experience and training. In her architecture, she proposes three hierarchical primary phases that represent the core of "situation awareness". However, although Endsley focused her work on "Situation Awareness", her architecture still fits the problem of data fusion and could be regarded as a general representation of the JDL proposal. Perception is the first stage where the system perceives the status, attributes and

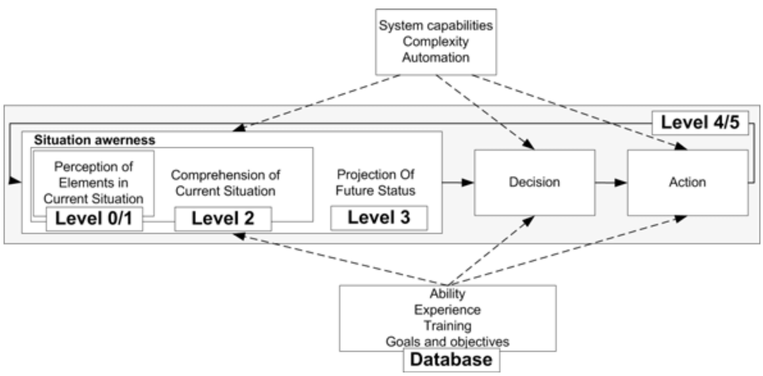


Fig. 2: Simplified Endsleys model and the correspondence of the JDL model. Endsley architecture fits the problem of data fusion and could be regarded as a general representation of the JDL proposal. We can see in the figure a mapping between JDL and Endsley modules.

dynamics of relevant elements in the environment. Then, based on the knowledge acquired in the first level, the system tries to make a holistic view of the environment, comprehending the significance of objects and events. The last step represents the ability of the system to project actions on the elements in the environment. This is achieved through knowledge of the status and dynamics of the elements (Level 1) and comprehension of the situation (Level 2). We note that both of the two architectures (JDL and Endsley) present perception and knowledge as key modules of the model. They both noted the importance of the ability to forecast future events. As shown in figure 2, we can also recognize a clear mapping between JDL and Endsley modules. In addition, Endsley further expands the JDL architecture by including decision and action modules, with clear deterministic interaction scheme among the modules.

3.3 Jakobson Architecture

In another relevant work, Jakobson [10] proposed a generalized two dimensional fusion model (figure 3). The first dimension represents the area (Analysis, Reasoning or Acting) and the second dimension represents the level of fusion (signal, data or cognitive level). The fusion of data implies the ”Analysis” of the data and objects and complex dynamic situations. Then, ”Reasoning” about the objects and situation, and the prediction of the changes in the situation. Finally, ”Acting” corresponds to the implementation of the panned action. This fusion is then done in three levels of abstraction (signal, data and cognition). One can note that the model proposed by Jackobson overlaps with the proposed models by JDL and Endsley, but it is more general in a sense that it combines the notions of fusion within the architecture of reasoning. Furthermore, Jakobson's proposal include some attractive features such as providing a well-defined interaction scheme among the modules and offering high level of flexibility by performing the data fusion at three levels.

Regarding the major data fusion architectures presented in this section, several

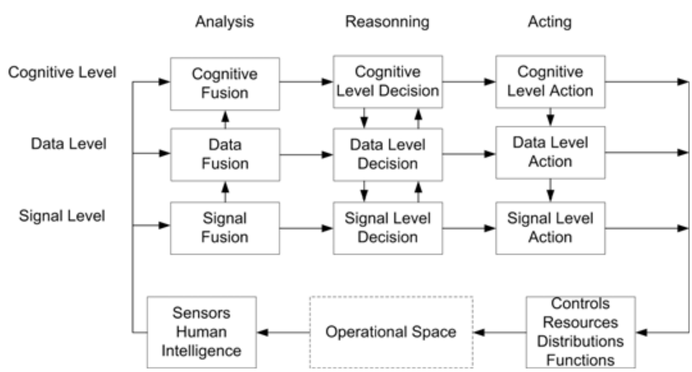


Fig. 3: Generalized two dimensional fusion model proposed by Jakobson [10]

drawbacks can be made about each one of them. JDL model does not specify a clear scheme for interaction among the modules. Endsley's architecture is restricted to a single level and does not offer the flexibility offered by the multi-level data fusion processing found in Jakobson's architecture. On the other hand, Jakobson's proposal neglects the role of memory in the architecture, and does not entails a feedback loop that is required to refine the decision hypothesis.

4 Proposed Architecture

In this section, we propose a new data fusion architecture developed to integrate different characteristics from previous models within a unified framework that is intended to overcome the shortcomings mentioned before. Our architecture is based on the following major considerations:

1. Data fusion should be performed at different levels in order to extend its flexibility.
2. Data fusion and cognitive processes (perception, reasoning, projection, etc) should not be treated independently
3. The system and/or the user are involved in the fusion model through a refinement process (feedback)
4. Different factors influence the fusion process, such as memory/database and experience
5. The fusion process is accomplished within a predefined context/goal

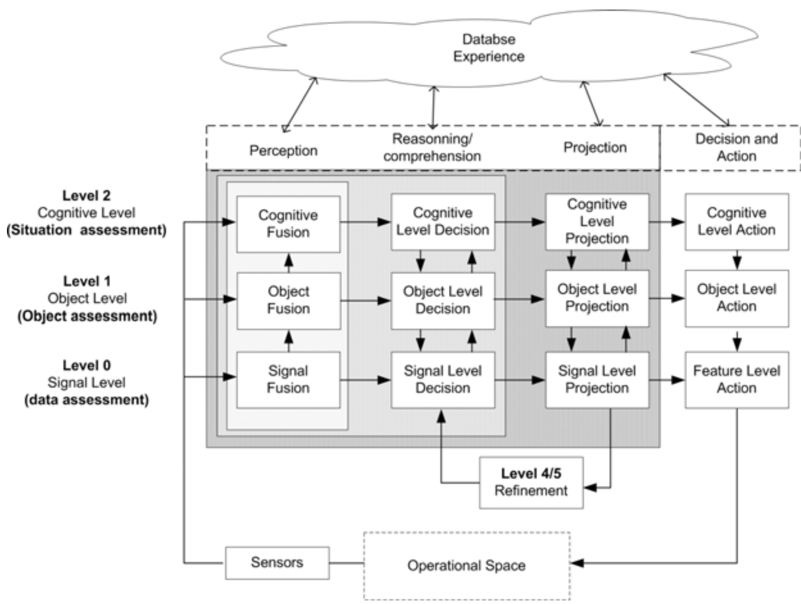


Fig. 4: The proposed architecture

It has been shown that for any fusion system to be successful, it must address the entire process flow: from data acquisition to awareness, prediction and the ability to request elaboration or additional data. Therefore, we considered two dimensional design of our architecture, similarly to Jakobson's work. The vertical axis represents the levels of abstraction, while the horizontal one corresponds to the situation awareness. We have three levels of abstraction in the fusion process: Signal level, Object level, and Cognitive level. Signal level deals with low level data, and estimation of states of sub-object entities. It includes inferences that do not require assumptions about the presence or characteristics of entities; whereas object level estimates the discrete physical objects. It is concerned with the estimation of the identity, classification, attributes, activities, location, and the potential states of entities. Finally cognitive level deals with inferences regarding relationships and contextual implications among entities. It also deals with the estimation and prediction of the utility or cost of an estimated world state to a user objective. The horizontal axis deals with the four layers of situation awareness: perception, reasoning and comprehension, projection, and decision and action. Perception provides information about the status, attributes and dynamics of the relevant elements in the environment. Reasoning level of the situation focuses on mimicking how people combine, interpret, store, and retrieve information. It includes the classification and integration of information according to their relevance to the target goal. Projection level represents the ability to forecast the future, in order to predict the best decision to make. Feedback about the predicted situation are forwarded to the reasoning layer through the refinement module. This reasoning refinement process keeps taking place till an appropriate acceptable decision is obtained. Finally, the decision and action level takes the corresponding measures in order to achieve the user's target in a top down manner. It can be noted that the user can be self involved in the data fusion loop, where he can play a major role in the refinement process, when not being conducted automatically.

Memory/databases play a major role in the data fusion architecture, where previous knowledge can benefit the agent to reason better according to its own experience. Many researchers and neuroscientists tried to build a model that mimics the human memory; however lots of work is still ahead before we reach a better level of understanding of the human memory. In this work, we suggest the cognitive memory model proposed by Bernard Widrow as it fits best cognitive robotics applications. Dr. Widrow formulated a unique approach to understanding memory that is based on memory behavior and performance, instead of neuro and brain anatomy [18]. He states that it is very much probable that the data is stored permanently as patterns (visual, auditory, tactile, etc.) in any available empty storage location. Storing information can occur at any level of abstraction, and stage of awareness as presented in our proposed architecture. Sensory inputs concerning one single object or event are stored together as patterns in single "memory folder". Therefore, when the contents of the folder are retrieved, sights, sounds, tactile feel, smell, are all obtained simultaneously. Retrieval of stored information results from reading the contents of a folder when prompted by a set of sensory inputs. The folder contents could also provide prompt patterns to find additional related folders that were not found in the initial search, thus starting a chain reaction of search.

5 Conclusion

In this work, we aim at proposing a more comprehensive data fusion architecture geared toward cognitive robotics applications, and pursuing the same goal of realizing such generalized architecture initiated by JDL (Joint Director's of Laboratories). The proposed architecture comprises two degrees of freedom represented by three levels of abstractions, and four layers of situation awareness. We also suggested a cognitive memory model that best suits our targeted robotics applications. Future research trend for this work could be to proof the correctness and efficiency of this architecture, through implementing it for real life cognitive robotics applications.

References

1. Fritsch, J., Haasch, A., Hofemann, N., Sagerer, G.: A multi-modal object attention system for a mobile robot. In: International Conference on Intelligent Robots and Systems (2005)
2. Williams, S., Durrant-Whyte, H., Makarenko, A., Brooks, A., Grocholsky, B.: A decentralized architecture for active sensor networks. In: IEEE ICRA (2004)
3. Abidi, M.A., Gonzalez, R.C.: Data fusion in robotics and machine intelligence. Academic Press, Inc., London (1992)
4. Cholvy, L.: Applying theory of evidence in multisensor data fusion: a logical interpretation. In: International Conference of Information Fusion (2000)
5. Fraichard, T., Bessiere, P., Coue, C., Mazer, E.: Multisensor data fusion using bayesian programming: An automotive application. In: IEEE-RSJ International Conference on Intelligent Robots and Systems (IROS) (2002)
6. Dasarathy, B.V.: Information fusion– what, where, why, when, and how? Information Fusion 2, 75–76 (2001)
7. Llinas, J., Hall, D.L.: Handbook of Multisensor Data Fusion. CRC Press, Boca Raton (2001)
8. Endsley, M.R.: Design and evaluation for situation awareness enhancement. In: Human Factors Society (1988)
9. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Human Factors Journal 37, 32–64 (1995)
10. Buford, J., Jakobson, G., Lewis, L.: An approach to integrated cognitive fusion. In: 7th International Conference on Information Fusion (2004)
11. Salerno, J.: Information fusion: A high-level architecture overview. In: Information Fusion (2002)
12. Waxman, D.F.l., Ivey, R.: A multisensor image fusion & mining: from neural systems to cots software. In: International Conference on Integration of Knowledge Intensive Multi-Agent Systems (2003)
13. Song, K.T., Chang, F.Y., Han, M.J., Hsu, J.H.: A new information fusion method for bimodal robotic emotion recognition. Journal of Computers (2008)
14. Shafer, G.: A mathematical theory of evidence. Princeton University Press, Princeton (1976)
15. Steinberg, A.N., Bowman, C.L., White, F.E.: Revisions to the jdl data fusion model. In: Sensor fusion: architectures, algorithms, and applications (1999)

16. Tan, K.C., Chen, Y.J., Wang, L.F., Liu, D.K.: Intelligent sensor fusion and learning for autonomous robot navigation. Applied Artificial Intelligence 19, 433–456 (2005)
17. Thrun, S.: Probabilistic algorithms in robotics. AI Magazine (2000)
18. Widrow, B.: "cognitive" memory and its applications. Stanford University
19. Han, B., Zhao, X., Luo, Q.: Survey on robot multi-sensor information fusion technology. In: Intelligent Control and Automation, pp. 5019–5023 (2008)

Nondestructive Flaw Detection System of Intelligent Robot with Machine Vision

Zhiguo Chen and Wenbo Xu

Abstract In order to improve efficiency and precision of nondestructive testing, a nondestructive flaw detection system based on intelligent robot is described. The developed system on 4-dof industrial robot can perform detection in three-dimensional space and the adaptive-network-based fuzzy inference system has been adopted to improve inspection adaptability. A color image segmentation algorithm and an improved adaptive region growing algorithm were proposed and proved to be effective in defect detection. Actual results indicate that the developed system has advantages of good stability and high precision.

1 Introduction

Nondestructive flaw detection is a settled technique in several applications. It has become a routine tool in goods production and plays a very important role. Its main aim is qualitative and quantitative detection of the possible defects on a given structure [1]. Over the past two decades, machine vision (MV) systems for automated visual inspection applications have consolidated their early promise and have become vital components in the design of advanced manufacturing systems [2]. Researchers continue to seek ways to improve the performance and flexibility of MV systems. Modern industrial robots are true marvels of engineering and they are reprogrammable [3], in many applications they are programmed once and then repeat that exact same task for years.
Our goal is to integrate industrial robot with machine vision so as to improve working efficiency of robot and expand inspection scope of machine vision, and thus it can be seen that nondestructive flaw detection system of intelligent robot with

School of Information Technology, Jiangnan University
No. 1800, Lihudadao Road, Wuxi, Jiangsu 214122, China
e-mail: chenzgwx,xwb_sytu@hotmail.com

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 157–165.
springerlink.com

machine vision will improve production efficiency and enhance degree of automation and it will have broad market prospects.

2 System structure

2.1 Nondestructive flaw detection system

Ultrasonic flaw detection is widely used to detect and characterize hidden internal defects in engineering materials such as metals, plastics and composites [4]. The radiographic inspection for defects, with an assortment of films available for recording the images, is the most popular and widely used nondestructive flaw detection technique even today. The technique used in welding industry can inspect and evaluate defect of the internal structure of material such as casting, forging, and surface welding defect for example crack, gas cavity, slag inclusion, shrinkage cavity, excessive penetration, fish eye, undercut, de-brazing, incompletely filled weld, incomplete penetration etc [5]. In order to solve the low efficiency and poor precision problems of traditional manual operation on nondestructive flaw detection, an integrated nondestructive flaw detection system installed on an industrial robot is developed.

2.2 Principle of control system

The proposed four-axis industrial robot of nondestructive flaw detection system consist of high-performance motion control card and AC servo systems with nondestructive testing unit installing on the bottom of rotation axis. IPC (Industrial PC) is responsible for robot control, image processing, feature extraction and defect identification etc. Rotary motion of motor is converted into linear motion using rack and pinion gearing in robot's axes of X, Y and Z, so positioning and trajectory interpolation in three-dimensional Cartesian coordinate system is achieved. X and Y axes of robot may be selected for coordinated motion consisting of linear and circular segments. In addition, Z axis can be controlled such that it remains tangent to the motion of the selected axes X and Y. Detecting head is mounting on W axis and it can rotate around Z axis.

The main linear motion units of robot consist of high-intensity and high-linearity aluminum extrusion profiles. In order to increase output torque and decrease load moment of inertia, all of servo motors have installed NEUGART precision planetary reducers. System assembly adopts gantry or wall-hanging depending on the environment of the scene. Figure 1 illustrates the system structure of robot and Figure 2 illustrates the elements of control system.

The Amp is motor amplifier and it may be configured in three modes: Voltage Driver, Current Driver and Velocity Loop. The Encoder generates N pulses per revolution. The Dac or D-to-A converter converts a 16-bit number to an analog voltage. The Zoh, or zero-order-hold, represents the effect of the sampling process, where the motor command is updated once per sampling period. More advanced designing methods and adjusting the tuning parameters are available with technical manuals from Galil [6][7][8]. Table 1 illustrates the control system parameters of robot

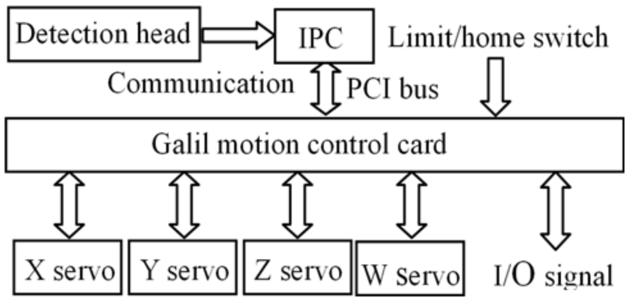


Fig. 1: System structure of robot.

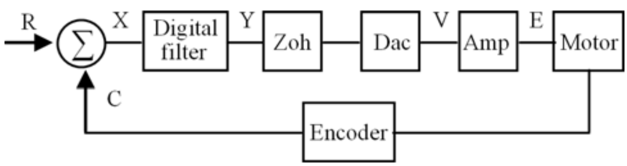


Fig. 2: Functional elements of control system.

2.3 Intelligent control technology

Because of a variety of workpiece shapes and in order to reduce robot's teaching times, the neuro-fuzzy classifier algorithm will be applied to the industrial image data set. The reader is referred to the text by Jang et al [9] for complete details on neuro-fuzzy algorithm, and in particular the adaptive neuro-fuzzy inference system (ANFIS) used in this study.

A fuzzy logic system is composed of three primary stages: fuzzification, rule evaluation, and defuzzification. Fuzzification is the process of changing crisp (numerical) input values into linguistic variables. Each input value can be associated with one or more linguistic variables with varying degrees of association. The degree of association is computed using membership functions $\mu(x_m)$, where x_m is the mth feature attribute input value. Once the inputs have been fuzzified, the fuzzy rules can be evaluated. These take the form of IF…THEN… rules where the parts evaluated between IF and THEN are the antecedents and the parts after the THEN are the consequents. The antecedent part of the rule j is calculated by

$$w_j = MIN(\mu_1(X_1), \mu_2(X_2), \ldots, \mu_p(X_p)) \tag{1}$$

where μ_1 is a membership function from feature attribute x_1, μ_2 is from feature attribute x_2, etc. Defuzzification is the reverse of fuzzification. One can refer to text [10] for details. The workpiece to be detected is matching by ANFIS algorithm at first and the trained image data sets come from those past objects. Some basic feature points are extracted from testing workpiece images and if they are similar to one in the existing database, the robot will inspect it as it has already done. This will significantly improve the efficiency of inspection.

3 Image processing algorithms

3.1 Color image segmentation algorithm

When a monochrome image is represented in the RGB color space and the resulting components are mapped independently, the transformed result is a pseudocolor image in which input image gray levels have been replaced by arbitrary colors [11]. Transformations that do this are useful because the human eye can distinguish between millions of colors-but relatively few shades of gray. Thus, pseudocolor mappings are used frequently to make small changes in gray level visible to the

Parameter name	X axis	Y axis	Z axis	W axis
Servo motor	Panasonic 4KW	Panasonic 4KW	Panasonic 4KW	Panasonic 2KW
Velocity input gain	100 (r/min)/V	100 (r/min)/V	200 (r/min)/V	200 (r/min)/V
Encoder frequency	10000 counts/sec	10000 counts/sec	10000 counts/sec	10000 counts/sec
Velocity gain	120	110	80	50
Velocity integration	1000	1000	1000	1000
Control mode	Velocity	Velocity	Velocity	Velocity
Axis resolution	1000 counts/mm	2000 counts/mm	505 counts/mm	40000 counts/mm
KP	5	20	4.5	4.5
KI	2	6	4	3
KD	10	30	12	10

Table 1: Control system parameters of robot.

human eye or to highlight important gray-scale regions. The method of pseudocolor transformation of gray image is to apply linear transformation independently in RGB color space. Figure 3 illustrates the welding X-ray image and Figure 4 shows the result of pseudocolor transformation.
Now the aim is to segment the color image and extract defect regions. Suppose that the objective is to segment objects of a specified color range in an RGB image. Given a set of sample color points representative of a color of interest, we obtain an estimate of the "average" or "mean" color that we wish to segmentation. Let this average color be denoted by the RGB pixel in an image as having a color in the specified range or not. To perform this comparison, we need a measure of similarity. One of the simplest measures is the Euclidean distance. Let z denote an arbitrary point in the RGB space. We say that z is similar to m if the distance between them is less than a specified threshold, T. The Euclidean distance between z and m is given by

$$\begin{aligned} D(z,m) &= ||z-m|| \\ &= \left[(z-m)^T(z-m)\right]^{1/2} \\ &= \left[(z_R-m_R)^2+(z_G-m_G)^2+(z_B-m_B)^2\right]^{1/2} \end{aligned}$$

where $||\bullet||$ is the norm of the argument, and the subscripts R, G, and B, denote the RGB components of vectors m and z. The locus of points such that $D(z,m) \leq T$ is a solid sphere of radius T. By definition, points contained within, or on the surface of the sphere satisfy the specified color criterion; points outside the sphere do not. Figure 5 illustrates the segmentation result.

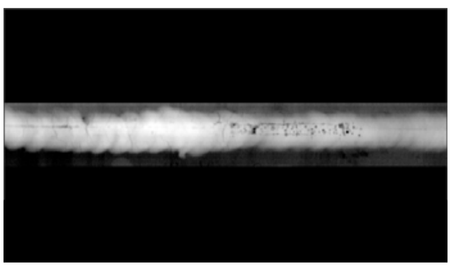


Fig. 3: Welding X-ray image.

3.2 Improved growing segmentation algorithm

Region growing is a procedure that groups pixels or subregions into larger regions based on predefined criteria for growth. The basis approach is to start with a set of

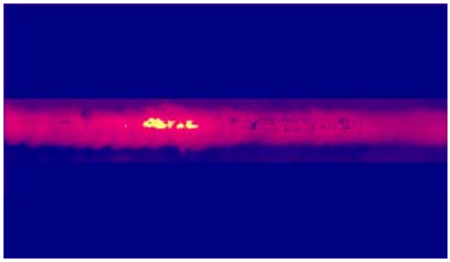


Fig. 4: Result of pseudocolor transformation.

Fig. 5: Segmentation result.

"seed" points and from these grow regions by appending to each seed those neighboring pixels that have predefined properties similar to the seed. Various expansion algorithms have been proposed by researchers based on it. For example, Pohle consider that pixels of region to be segmented are supposed to obey normal distribution. Distributed parameters can be estimated by original region growing and then these parameters are used in the second iteration growing. Better results can be found in [12].

The first order of the proposed algorithm is to determine the initial seed points. In this application, it is known that some pixels in areas of defective welds tend to have the maximum allowable digital value. The extract value T_1 can be found as follow steps:

1. Let f1 denote the negative image of original f;
2. Filtering f_1 with adaptive spatial filters so as to remove pepper and salt noises, and get f_2;
3. Get the index I from f_2 where gray level lower than average;
4. Calculate average digital value T_1 in f_1 using I where gray level higher than average.

The next step is to choose a threshold T_2. This number was based on analysis of the histogram of f and represents the difference between 255 and the location of the first major valley to the left, which is representative of the highest intensity value in the dark weld region. The third step, if a pixel is found to be connected to more than one

region, the regions are automatically merged. Figure 6 illustrates traditional region growing segmentation result and Figure 7 is the result of adaptive region growing segmentation.
It is evident by comparing Figure 7 with Figure 6 that the adaptive region growing procedure did indeed segment the defective welds with a more reasonable degree of accuracy.

Fig. 6: Functional elements of control system.

Fig. 7: Functional elements of control system.

3.3 Software development

System software consists of two parts: control software which is in charge of robot's motion control, and manipulate software whose main duty is image displaying and processing. System architecture divides into some subsystems relatively independent adopting use-case technology so that it is easier to grasp the overall situation. User interface software is developed by Visual C++ 2005 sp1 and database uses SQL Server 2005 sp2.

4 Experiments

In this section, details of our implementation and experimental results are presented. The proposed system described above has been tested on a prototype robot. In order to evaluate the performance of nondestructive flaw detection, we use the time needed and accuracy to recognize a set of predefined flaw points in the object as evaluation metric. The results of the experiments are reported in Table 2.

As it is shown in Table 2, a higher performance is achieved with proposed methods. The major cause might be that conventional method is relying on human being to read X-ray image and give the final decision. Another reason is that traditional method should rotate the turn around table periodically and then readjust it so as to capture better images, and proposed robot can move the detection probe to the right positions automatically and quickly.

5 Conclusions

A nondestructive flaw detection system based on intelligent robot has been described. The system was tested using images generated by X-ray detection equipment. A color image segmentation algorithm and an improved adaptive region growing algorithm were proposed and tested. Results indicate that both algorithms perform well. The developed system is adaptable to other machine vision-based automated inspection applications.

Methods	Proposed method 1	Proposed method 2	Traditional method
Accuracy	95.2 %	93.2 %	90.5 %
Time (s)	652	580	3852

Table 2: Performance comparison.

References

1. Raj, B.: Practical Non-Destructive Testing, 3rd edn. Alpha Science Int'l Ltd., Oxford (2007)
2. Davies, E.R.: Machine Vision Theory, Algorithms, Practicalities, 3rd edn. Morgan Kaufmann, San Francisco (2004)
3. Norberto Pires, J.: Industrial Robots Programming, Building, Applications for the Factories of the Future. Springer, New York (2006)
4. Song, Z., Wang, Q., Du, X., et al.: A High Speed Digital Ultrasonic Flaw Detector Based on PC and USB. In: Proceedings of IEEE Instrumentation and Measurement Technology Conference, vols. 1-3, pp. 1–4 (2007)
5. Polikar, R., Udpa, L., Udpa, S.S., et al.: Frequency invariant classification of ultrasonic weld inspection signals. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 45(3), 614–625 (1998)
6. Galil Motion Control, Inc., C/C++ Programmers Tool Kit Reference Manual Rev. 2.0. Rocklin, California (2002), `http://www.galilmc.com/support/manuals.php` (accessed June 24, 2007)
7. Galil Motion Control, Inc., DMC-18x2 Manual Rev 1.0f. Rocklin, California (2003), `http://www.galilmc.com/support/manuals.php` (accessed June 24, 2007)
8. Galil Motion Control, Inc., 18x2 COMMAND REFERENCE Manual Rev 1.0d. Rocklin, California (2004), `http://www.galilmc.com/support/manuals.php` (accessed June 24, 2007)
9. Jang, J.S.R., Sun, C.T., Mizutani, E.: Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. Prentice-Hall, Upper Saddle River (1997)
10. Ross, T.: Fuzzy Logic with Engineering Applications. Wiley, London (2004)
11. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall, Upper Saddle River (2007)
12. Pohle, R., Tonnies, K.D.: A new approach for model-based adaptive region growing in medical image analysis. In: Proceedings of the 9th International Conference on Computer Analysis and Patterns, Warsaw, pp. 238–246 (2001)

A Novel Online Textual/Graphical Domain Separation Approach for Sketch-Based Interfaces

Danilo Avola, Andrea Del Buono, Pierluigi Del Nostro, and Rui Wang

Abstract Multimodal interfaces can be profitably used to manage the increasingly complex applications and services which support human activities in everyday life. In particular, sketch-based interfaces enable users to effortless and powerful communication way to represent concepts and/or commands on different devices. This kind of information can be expressed by users performing two types of object: freehand drawing (graphical domain) and/or handwriting (textual domain). Usually, current frameworks require that users, somehow, indicate whether they are performing one or the other object. In this way, the frameworks can adopt the suitable recognition process to interpret as expressed by users. Moreover, more complex situations can occur when users perform, on a same schema, both types of object. This paper describes a novel intelligent framework able to automatically distinguish, in online way, freehand drawing from handwriting. The proposed approach works taking into account only the mathematical features belonging to the sketch performed by the user during interaction activity. Moreover, the approach can be used on schemata made up by heterogeneous objects which can also be overlapped.

Danilo Avola
University of Rome "La Sapienza", Department of Computer Science, Via Salaria 113, 00189 Rome, Italy, e-mail: avola@di.uniroma1.it

Andrea Del Buono
CSKLab National Research Center, Department of Advanced Research, Via Savoia 84, 00198 Rome, Italy, e-mail: andrea.delbuono@csklab.it

Pierluigi Del Nostro
University of Rome "Rome 3", Department of Computer Science, Via della Vasca Navale 79, 00146 Rome, Italy, e-mail: pdn@dia.uniroma3.it

Rui Wang
"Saarland" University, Department of Computational Linguistics, Saarbrücken 66041, Saarbrücken Germany, e-mail: wang@coli.uni-sb.de

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 167–176.
springerlink.com

1 Introduction

Multimodal interfaces allow users to interact, in natural way, with any desktop or mobile devices using multiple different modalities (e.g.: sketch, gesture, speech, gaze, etc.). This kind of interaction provides a powerful tool by which users manage the increasingly complex applications and services that surrounding their activities in everyday life. In particular, sketch-based interfaces enable users to express concepts, to provide commands and to represent ideas in an immediate, intuitive and simple way. In order to express this kind of information users can perform two types of object: freehand drawing (graphical domain) and/or handwriting (textual domain).

The approaches and the algorithms able to detect and to recognize the mentioned objects are deeply different. Moreover, the pre-processing steps that have to be carried out on the sketch before it is given to one or the other recognition engine depends only on the type of object. For these reasons, current frameworks require that users, somehow, indicate whether they are performing freehand drawing or handwriting activity. Besides, it has to be taken into account that complex concepts, commands or ideas could be expressed, by users, using both types of object. In this case, the first step that a framework has to accomplish is to distinguish, within the same schema, what are the objects related to the freehand drawing and what those related to the handwriting. The situation can further be complicated considering that a sketch environment includes several *user actions*, such as: deletes, restyling (of whole object or part of it), and so on. Moreover, it has to be considered that different homogeneous and/or heterogeneous objects can have particular *spatial relationships* with one another, such as: inclusion, overlapping, closeness, and so on.

This paper describes a novel intelligent framework able to automatically distinguish, in online way, freehand drawing from handwriting. The framework, working only on the mathematical features belonging to the sketch performed by the user, is able to overcome all the just mentioned critical duties tied to both *user actions* and *spatial relationships* related to homogeneous and/or heterogeneous objects. Compared to the actual adopted methodologies regarding the separation between freehand drawing and handwriting, our developed approach (and related prototype) has three advantageous main aspects that jointly put our results in the vanguard in the sketch-based interfaces area. The first one regards the possibility to perform both types of object, in online way, on a same schema. The second one concerns the possibility to have both types of object with any possible *spatial relationships* with one another. The last one regards the possibility to perform the separation process without the need to consider specific application domains and/or templates and/or libraries for one or the other type of object. The developed prototype, and the related experimental session, shows the effectiveness of the obtained results.

The paper is structured as follows. Section 2 proposes some remarkable related works which discuss about the separation between freehand drawing and handwriting. Section 3 introduces the novel approach and the related developed framework. Section 4 shows experimental results on a wide application domain made up by several technical/non-technical freehand drawings and handwriting strings. Finally, Section 5 concludes the paper.

2 Related Works

There are many works about text/graphics separation on different documents, some of these also regarding the online recognition of handwriting (specially on Japanese, Chinese or Arabic language). But no so much works regarding the online separation between objects belonging to freehand drawing and handwriting are given. Besides, no so much experimental results based on an effective framework are shown. However, our work has been inspired from some remarkable approaches described below. A real interesting work, that has driven some our choices regarding mathematical feature extraction process, is shown in [6]. In this work the authors detail a recognition online process able to identify handwritten characters. In order to perform this task the stokes belonging to the characters are treated as curves which have to be suitably classified. A similar work is shown in [7]. The authors model on-line ink traces for a set of several symbols to suitably fit low-degree polynomial series. In order to accomplish this task they use a collection of mathematical writing samples which provides a succinct way to model the stylus movements of current test users. Another remarkable work is shown in [8]. In this work the authors describe a pattern recognition framework able to distinguish global styles of different Chinese people's handwritings. The approach achieves concrete results by combining wavelet transform and generalized gaussian model. Another interesting approach is shown in [3]. In this work the authors propose an offline method to distinguish, within the document, multiple fonts, math notations, and graphics. The strength of proposed work is the possibility to perform document analysis without considering specific structural models, moreover some principles expressed in this work can be suitably adopted in online framework. In [4] the authors propose a method to separate and recognize the touching/overlapping alphanumeric characters in raster-scanned color cartographic maps. The developed approach performs four main segmentation steps, in the first one the map is segmented to extract all text strings including those that are touching other symbols, strokes and characters. In the second one a OCR-based recognition with Artificial Neural Networks (ANN) is applied to define the coordinates, size and orientation of alphanumeric character strings presented in the map. During the third step four straight lines or a number of curves computed as a function of primarily recognized by ANN characters are extrapolated to separate those symbols that are attached. Finally, in the last step, the separated characters input into ANN again to be finally identified. Also in this case several ideas regarding, for example, the overlapped characters, have been deepen. A similar approach, equally useful, is shown in [5]. Also in this work the authors face the separation of overlapping text and graphics in document image analysis. The proposed approach is based on the observation that the constituent strokes of characters are usually short segments in comparison with those of graphics. More than others this approach can be exploited to support online separation frameworks. Another remarkable work is shown in [9]. In this work the authors present a novel holistic technique for classifying Arabic handwritten text documents. More specifically, the proposed approach exploits both segmentation techniques and feature extraction processes to teach to ANN the classification of styles/fonts need to retrieve Arabic handwritten text documents. From our point of view this work has been interesting “to certificate” some measurement approaches (in particular feature extraction processes) used within of our developed framework.

3 Proposed Approach and Related Framework

This section details the proposed novel approach which works considering only the mathematical features belonging to the sketch performed by the user during interaction activity. In particular, the main algorithmic aspects by which the framework has been developed are shown. In order to explain the section content, next subsection introduces both some basic concepts and a general simplified design of the framework.

3.1 Basic Concepts and Designed Framework

The following definitions and considerations are necessary:

- Stoke Definition: A stroke is the set of pixels obtained during the following per action: pen down, pen drawing, pen up.
- Object Definition: An object is any sketch, belonging to freehand drawing or handwriting area, made up by one or more than one stroke.

In order to generalize the freehand drawing and handwriting concepts, in this work the following definitions are considered:

- Graphical Domain Definition: Belong to this domain all types of freehand drawing. In particular, the drawings can have geometrical and/or non-geometrical characteristics, besides they can represent every object and/or abstraction and/or situation in real life.

Some examples of instances of graphical domains are: dataflow diagram (dfd), entity relationship diagram (erd), electrical schemes, general graphical symbols/icones, geometrical figures, flowcharts diagram (fcd), freehand descriptive drawings, and so on.

- Textual Domain Definition: Belong to this domain all types of cursive or block letters strings performed by handwriting. In particular, the strings have to be written according to the following main rules: horizontally (writing from left to

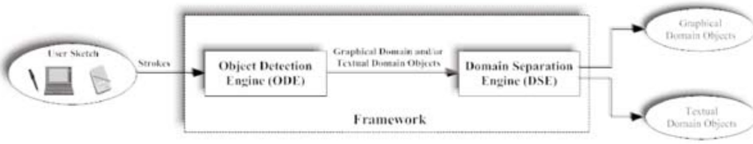


Fig. 1: Design of the Developed Framework.

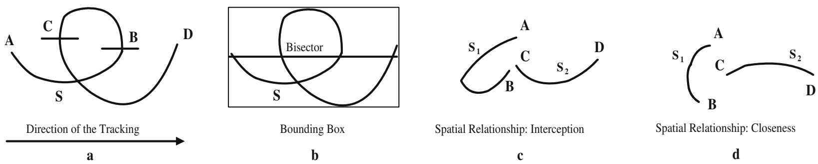


Fig. 2: a) Curvature b) Entropy c) Interception d) Closeness.

right) or vertically (writing from up to bottom). Moreover, the used characters have to belong to the roman alphabet and/or roman numerals.

Some examples of instances of textual domains are: labels on diagrams/schemata, textual description of objecs/situations, and so on. In order to describe the main algorithmic aspects by which the framework has been developed in Figure 1 a general design of the developed framework is shown. Our developed framework is made up by two main sequential elaboration engines: Object Detection Engine (ODE) and Domain Separation Engine (DSE). The first one analyzes the set of strokes performed by user, during interaction activity, and identifies how many and what objects the user has sketched. The second one elaborates each single object determining whether it belongs to the graphical or textual domain. The next two subsections detail the just introduced engines.

3.1.1 Object Detection Engine

The aim of the ODE is to identify how many and what objects the user has sketched. In order to accomplish this task on each single stroke and on each set of strokes having particular spatial relationships (*interception* and *closeness*), two main mathematical features are computed: *curvature* and *entropy*. In Figure 2 the just introduced measures are shown. More specifically, as shown in Figure 2-a, the curvature measure represents the sum of the inner angles according to the direction of the tracking of the stroke and/or set of strokes. The stroke S (starting from point A) has two changes of direction (points B and C) and ends to the point D. The inner angles of three sub-strokes ([A, B], [B, C] and [C, D]) are summed to provide the total curvature measure of the mentioned stroke. In Figure 2-b the entropy measure is shown. It represents the messy level of the pixels that make up the stroke,

and/or set of strokes, related to the “bisector” of the bounding box. In particular, the stochastic distribution of the two set of pixels (above and below the bisector) is computed to evaluate the messy level of each set. These two types of measure are always computed on each single stroke, moreover when two or more strokes have particular spatial relationships (*interception* and *closeness*) the mentioned measures are computed on the joined stroke made up by the related set of strokes.
As shown in Figure 2-c two (or more) strokes can be considered intercepting when the start and/or end point of one coincides (taking into account a distance tolerance) with the start and/or end point of the other. While, as shown in Figure 2-d, two (or more) strokes can be considered close when they are near each other (also in this case, taking into account a distance tolerance). Indeed, two or more intercepted strokes are joined together only if they present similar values of curvature and entropy, otherwise they have to be considered only strokes in closeness spatial relationship. Moreover, two or more strokes in closeness spatial relationship are never joined (independently from the curvature and entropy measures).
By the just introduced concepts it is possible to introduce the following definition of homogeneous object able to detect and to distinguish each object sketched by user during interaction activity:

- Homogeneous Object Definition: An object results homogeneous only if it has one of the following main features:
  - High levels of curvature and entropy on each stroke and on the joined stroke;
  - Low levels of curvature and entropy on each stroke and high levels of curvature on joined stroke;
  - High levels of curvature or low levels of entropy on each stroke and high levels of curvature on joined stroke;
  - Medium levels of curvature and entropy on each stroke and high levels of curvature and entropy on joined stroke.

It is possible to obtain some situations in which the found measures do not ensure a correct discrimination of the objects within the schema, in this case it is also considered that users adopt part-by-part strategy, described in [1], to perform any sketch of an object. This strategy states that users tend to end sketching an object before sketching another. Figure 3 shows an example of object detection process. As it is possible to observe ODE, on given schema, detects six different objects (highlighted by the six rectangles shown in the middle window of the screenshot in Figure 3). It is important to highlight that as ODE ends its activity, it provides only how many and what objects the user has sketched, but it does not detect the domain (graphical or textual) of the objects.
Ultimately, it has to be considered that strokes performed in vertical way are “translated” in horizontal during object detection process.

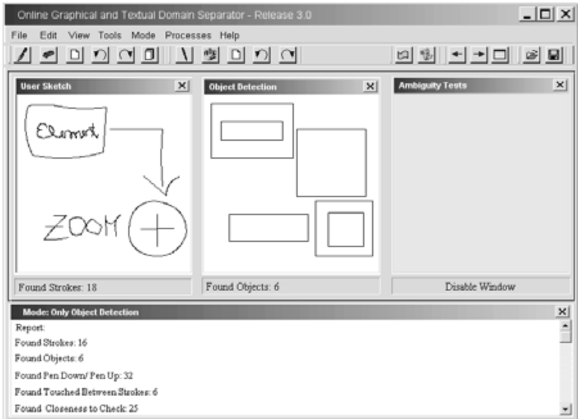


Fig. 3: Example of Six Recognized Heterogeneous Objects on a Schema.

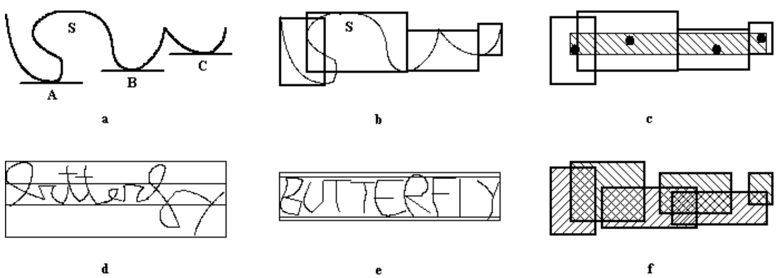


Fig. 4: a) Minimum Points b) Enclosing Rectangles c) Linearity d) and e) Bands Ratio f) Ratio.

3.1.2 Domain Separation Engine

The aim of the DSE is to analyze each object coming from ODE and to detect it as belonging to graphical or textual domain. Indeed, on each object several mathematical features are performed with the purpose to certify the object as belonging to the textual domain. Even if only one measure fails then the object is certified as belonging to graphical domain. In order to explain the measures to analyze each object, it is necessary to introduce the concept of enclosing rectangle built by absolutes and relatives minimum points of the stroke. Indeed, the mentioned concept is a variation of the approach proposed in [2] which considered both maximum and minimum points. Briefly, as shown in Figure 4-a and 4-b, it is possible to build enclosing rectangles that contain segments of the stroke through the absolutes and relatives minimum points belonging to the transposition of the two-dimensional space of the stroke on a fixed one-dimensional axes (e.g.: x axes). More specifically, as shown in Figure 4-a, the points A, B and C identify the minimum points (both absolute and

relatives) of the stroke S. By these points, as shown in Figure 4-b, the related enclosing rectangles are built. These rectangles support the definition of the following three main mathematical measures: *enclosing rectangles linearity*, *bands ratio* and *enclosing rectangles ratio*.
The first measure checks the alignment of the enclosing rectangles. As shown in Figure 4-c, on each enclosing rectangle the barycenter point is computed. The enclosing rectangles result aligned if all the barycenter points are horizontally aligned according to specific threshold. The second measure, obtained by density of the pixels within the enclosing rectangles, computes, on a selected object, three density bounds (top, middle and bottom). These bounds identify the behavior of the statistical distribution of the pixels of an object. As shown in Figure 4-d and 4-e these bounds have a high meaning on cursive or block letters strings, more specifically:

- Cursive or Block Letters Strings are characterized by one of the following main features:
  - High density of pixel within the middle bound and equitably distributed low density on both top and bottom bounds;
  - Density of pixels equitably distributed on all three bounds;
  - Density proportionately growing, on the three bounds, in one of the following orders: top-middle-bottom, bottom-middle-top, middle-top-bottom and middle-bottom-top.

In Figure 4-f an example of the last measure is shown. In particular, it regards the ratio between the area obtained from the union of the enclosing rectangles and the area obtained from the intersection of the enclosing rectangles. In particular, according to this measurement, an object is potentially recognized as belonging to the textual domain if the ratio is no more than fifteen percent.
It is important to highlight that an object sketched vertically is, first of all, brought back to the horizontal case according to the tracking verse. As just mentioned, if all three main measures are positively checked on an object, then it can be considered belonging to the textual domain, else it belongs to the graphical domain. Obviously, each measure considers different levels of exceptions depending on: length of strokes, initials and/or finals strokes, strokes made up by small number of pixels (i.e.: points), and so on. Figure 5 shows the domain separation result performed on the example introduced during previous subsection. It might be useful to add another "engine" that merges objects of a same domain having inclusion spatial relationship (in this case the circle with the cruise could be considered a single object). Indeed, this further step depends only on the specific semantic recognition engine used for interpreting the meaning of symbols and/or strings.

4 Experimental Results

Our experimental session has considered sketch activity of two types of schema: simple schema and complex schema. A simple schema was a sketch made up by

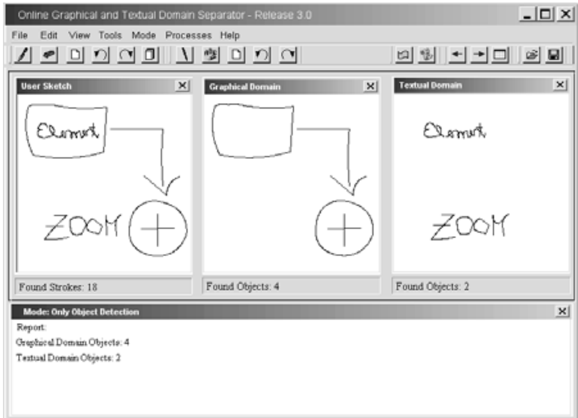
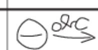
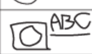
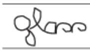


Fig. 5: Four Graphical Domain Objects and Two Textual Domain Objects.

Table 1: Experimental Session.

N° of Simple Schemata (315)	N° of Strokes	Recognized Domain	N° of Objects	N° of Complex Schemata (145)	N° of Strokes	Recognized Domain	N° of Objects
	3	Graphical	2		6	Graphical/Textual	4
	3	Graphical	2		7	Graphical/Textual	4
-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------
	9	Textual	1		5	Graphical/Textual	3
	4	Textual	1		5	Graphical/Textual	2
-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------
315	2634	Percentage of Recognition 96%	741	145	1513	Percentage of Recognition 89%	912

homogenous objects, which could be either any simple/complex graphical symbol or any simple/complex string. A complex schema was a sketch made up by heterogeneous objects. In this way has been tested each aspect of the two proposed engine (ODE and DSE). In both schemata it was possible to use the desired number of strokes. In Table 1 summarized results of our experimental session are given. As it is possible to observe, test users have sketched 460 schemata subdivided as follows: 315 simple schemata and 145 complex schemata. On simple schemata have been achieved almost optimal results, while on complex schemata the success rate has been lower. This event depends on the error level that occurs during multi-objects separation process in ODE. More specifically, when a user performs several object, on a same schema, can occur that the spatial relationship measures become ambiguous. Moreover, the separation of complex freehand drawings from strings made up by block letters is a hard task.

It is important to highlight following considerations. The schemata have been performed by different users (about 25), this is an important factor because, as well known, each user has its own tracking style that can make the behavior of sketch-based interface oriented algorithms unreliable. The different elements belonging to the tested schemata cover most application contexts, such as: technical diagrams, generic freehand drawings, modeling diagrams, and so on.

5 Conclusion

Sketch-based interfaces enable users to effortless and powerful communication way to represent concepts and/or commands on different devices. This kind of information can be expressed by users performing objects belonging to graphical and/or textual domain. Before using recognition process on each object, it is necessary to detect their related domain. This paper describes a novel intelligent framework able to automatically distinguish, in online way, the related domain of an object. The experimental results, based on developed framework, show the effectiveness of the proposed approach.

References

1. Avola, D., et al.: Ambiguities in Sketch-Based Interfaces. In: The Proceedings of the 40th Hawaii International Conference on System Sciences (HICSS 2007), p. 290b. IEEE Press, Los Alamitos (2007)
2. Avola, D., et al.: Formalizing Recognition of Sketching Styles in Human Centered Systems. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part II. LNCS, vol. 4693, pp. 369–376. Springer, Heidelberg (2007)
3. Changhua, W., Gady, A.: Document Image De-warping for Text/Graphics Recognition. In: Caelli, T.M., Amin, A., Duin, R.P.W., Kamel, M.S., de Ridder, D. (eds.) SPR 2002 and SSPR 2002. LNCS, vol. 2396, pp. 348–357. Springer, Heidelberg (2002)
4. Velázquez, A., Levachkine, S.: Text/Graphics Separation and Recognition in Raster-Scanned Color Cartographic Maps. In: Lladós, J., Kwon, Y.-B. (eds.) GREC 2003. LNCS, vol. 3088, pp. 63–74. Springer, Heidelberg (2004)
5. Cao, R., Tan, C.L.: Text/Graphics Separation in Maps. In: Blostein, D., Kwon, Y.-B. (eds.) GREC 2001. LNCS, vol. 2390, pp. 167–177. Springer, Heidelberg (2002)
6. Golubitsky, O., Watt, S.M.: Online Stroke Modeling for Handwriting Recognition. In: Proceedings of the 2008 Conference of the Center for Advanced Studies on Collaborative Research: Meeting of Minds (CASCON 2008), pp. 72–80. ACM Press, New York (2008)
7. Char, B.W., Watt, S.M.: Representing and Characterizing Handwritten Mathematical Symbols through Succinct Functional Approximation. In: Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), vol. 02, pp. 1198–1202. IEEE Computer Society, Washington (2007)
8. He, Z., You, X., Tang, Y.Y.: Writer Identification Using Global Wavelet-Based Features. International Journal on Neurocomputing 71(10-12), 1832–1841 (2008)
9. Brook, S., Al Aghbari, Z.: Classification of Personal Arabic Handwritten Documents. Transactions on Information Science and Applications WSEAS 5(6), 1021–1030 (2008)

Effective Region-based Relevance Feedback for Interactive Content-based Image Retrieval

Walid Barhoumi, Abir Gallas, and Ezzeddine Zagrouba

Abstract This paper proposes an effective framework for interactive region-based image retrieval. By utilizing fuzzy coarse segmentation and the graph structure for representing each image, the retrieval process was performed by measuring the image similarity according to the graph similarity. To assess the similarity between two graphs, fuzzy inter relations among regions feature vectors and spatial dispositions as well as fuzzy regions weights are explored. A region-based relevance feedback scheme was also incorporated into the retrieval process, by updating the importance of query image regions based on the user feedbacks, leading to a further performance improvement. Experimental study proves that the proposed region-based relevance feedback mechanism tailors the system semantic behavior relatively to each user personal preferences through the accumulation of the useful semantic information from the feedback information.

1 Introduction

The exponentially grown amount of visual information available in image databases imposes the urgent need of efficient tools for automated indexing and smart retrieval engines. Early indexing and retrieval systems use text to describe image content requiring excessive manual annotation of large image collections. To overcome the subjectivity and the ineffectiveness of the text-based image retrieval systems [1], most current engines use visual features such as color, texture and shape in order to index and retrieve digital images. However, most of the various proposed features, which can be either global or block-based, cannot adequately capture local variations in the image [2] and are thereby incapable to capture efficiently high-level

Equipe de Recherche "Systmes Intelligents en Imagerie et Vision Artificielle" (SIIVA),
Institut Supérieur d'Informatique,
2 Rue Abou Rayhane Bayrouni, 2080 Ariana, Tunisia.
e-mail: valid.barhoumi@laposte.net, abirgallas@yahoo.fr,
ezzeddine.zagrouba@fsm.rnu.tn

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 177–187.
springerlink.com

semantic concepts in user's mind. Thus, in order to narrow down this unsolved semantic gap [3], recent region-based engines intend to represent images in the object level in order to be close to the user's visual perception of an image [4]. Recently, learning techniques, notably relevance feedback (RF), have also shown remarkable performance improvement when applied to the context of image retrieval [5]. The main idea of the relevance feedback is to allow the interaction of the user with the retrieval system in order to learn the user's perceptions and preferences. Nevertheless, although many relevance feedback methods were successfully introduced in retrieval engines using global representation, it is seldom applied to the region-based image retrieval (RBIR) [6]. This is mainly due to the hard representation of different images in a uniform feature space since these images have often different number of regions [3]. As mentioned by [7], the incorporation of relevance feedback in image retrieval consists to developing a weighting scheme to explicitly "guide" the retrieval, or to applying machine learning techniques to reduce the problem to a standard classification problem.

The main purpose of our work is addressing the issue of integrating region-based image retrieval with relevance feedback learning while allowing them to benefit from each other in order to improve the retrieval performance. Indeed, this paper presents an interactive RBIR system based on RF. The first level of online interaction between the user and the system consists to the presentation of the query image. We avoided regions-as-query which would add complexity in user interface and users' interaction. Next, since RBIR performance is substantially based on the region segmentation quality, which is never perfect in practice [8], images are coarsely segmented into fuzzy regions and features are extracted from each region while using fuzzy logic. The suggested system represents then an image as a weighted graph such that each node is labeled with low-level features of a particular region and it is weighted according to the correspondent region's visual importance. Besides, since spatial information is considerably related to the semantics of image content, the spatial information is incorporated into the graph structure. Our graph modeling is totally based on fuzzy logic what should optimize the proposed RBIR effectiveness and robustness [7]. Then, the online extracted graph of the query-image is compared against the stored graphs representing images in the database using a cost-efficient graph matching technique. The relevance feedback scheme is also combined to dynamically adjust the node weights of the image graph. Our image RF method is designed according to the characteristic of the region-based representation and it is based on heuristic technique with empiric parameter adaptation similarly like in weighting methods from text-based retrieval [9]. Indeed, we proposed herein a simple, but effective, interactive mechanism which looks to adapt the query regions weights towards user's positive feedbacks. These positive feedbacks try to approximate the ideal target query while trying to learn the common regions (objects) from relevant images. The rest of this paper is organized as follows. The proposed schemes of graph-based RBIR, and region-based RF are respectively introduced in sections 2 and 3. The objective assessment study of the suggested interactive retrieval system will be presented in section 4. Lastly, conclusions and ideas for further works are summarized in section 5.

2 Graph-based RBIR

We proposed herein an efficient graph-based RBIR approach allowing to model semantic-level objects and inter-objects spatial relationship. The suggested graph modeling incorporates as much expressive and discriminating information as possible and, at the same time, achieves a relatively low computational cost. It offers a fully region-based indexing method based on fuzzy logic to incorporate region quantification, in terms of low-level features and visual importance, and inter-regions spatial disposition information, in order to improve the retrieval effectiveness and robustness. Based on the basic assumption that any region could be useful in the retrieval process [10], all the regions of each image are considered. Then, a complete graph is defined relatively to each image, where each node represents a salient region and is quantified in terms of two fuzzy properties: the low-level features of the region and its visual importance weight. Besides, inter-regions spatial disposition is also quantified in terms of five fuzzy properties. Thus, the used graph structure summarizes as maximum as possible the visual content of an image with real value fuzzy features, so more information can be provided to retrieve images.

2.1 Fuzzy Region Quantification

The principle of RBIR is based on the inclusion of object notion into the similarity evaluation between images. Thus, images are coarsely segmented into a dynamic number of regions that roughly correspond to objects. These salient regions are the processing units in our RBIR and RF approaches. In order to ensure robustness against inaccurate segmentation, instead of ignoring the ambiguous boundary pixels during the features extraction, we used a fuzzy segmentation technique. It consists on applying the basic watershed algorithm on each color component of the image. Besides, in order to overcome over-segmentation effects, a region-growing post treatment merges seeded watershed regions with lower values, according to their color histogram similarities. The segmentation performance is fairly good, since in most situations, each extracted region sketches approximately a semantic object inside of an image (Fig. 1). Then, each salient region is described in terms of two fuzzy properties. On one hand, visual content of a region is approximated by low-level features. Here, we used only color features in order to avoid the common weighted summation of different features, since it is difficult to decide the weights corresponding to the features and it is time consuming. Thus, to quantify the visual content of a salient region, we are limited to the intensity mean of this region in each component in the RGB color space. While computing the three color means, each pixel inside the considered region was weighted with its fuzzy degree of membership within this region, in order to guarantee more robustness against inaccurate segmentation. On the other hand, the visual importance of each region is expressed with fuzzy weight. Here, we used two fuzzy factors for determining relative importance of each region in the query-image, such that the only requirement is that the sum of importance weights of an image should be equal to 1. In fact, a spatial location factor looks to reflect the process of human visual perception while enhancing the weighting importance assigned to the region whose pixels are closer to the attention center. Besides, each region is weighted according to its area ratio to the corresponding image, modeling its importance in the query image relatively to the human visual system [6] (Fig. 1). Weights of the query-image regions will be then adaptively updated in the user relevance feedback scheme.

2.2 Fuzzy Inter-region Spatial Disposition

Each edge between tow region-nodes is labeled with a quintuplet modeling the inter spatial positioning between these regions. Indeed, the disposition of a region relatively to another one is quantified in terms of a binary value of inclusion and four fuzzy values estimating the degree of positioning on the left, on the right, in the top and in the bottom [11]. Note that we applied fuzzy logic to define region features and inter-regions spatial disposition in order to address the typical feature representation impreciseness. Indeed, fuzzy logic allows a certain degree of variations of the feature values, what improves the robustness and the effectiveness of the indexing scheme, and thereby the retrieval effectiveness.

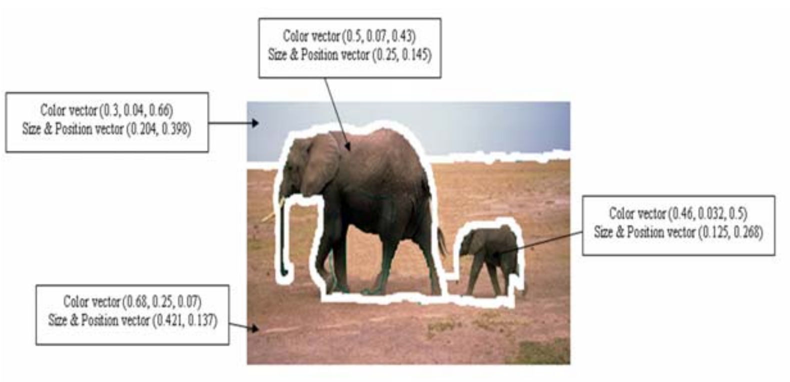


Fig. 1: Fuzzy region quantification. The set of segmented regions, such that each salient region is quantified in terms of two fuzzy properties: the low-level features of the region (color vector) and its visual importance (size ratio, spatial position).

2.3 Cost-efficient Graph-based Image Similarity Measurement

In order to profit of as much image information as possible, the suggested definition of the overall distance between two image graphs is a balanced scheme in similarity measure between nodes and edges matching. It formulates the matching problem in a probabilistic framework and simultaneously models both first-order region properties and second-order spatial relationships for all the regions in the image [12]. In fact, optimal node matching is first done in order to define, relatively to each region $R_j (1 \leq j \leq n)$ of the query image, the most similar region $R_j^{'} (1 \leq j \leq t)$ in the studied image of the database, in terms of low-level color features. This is iteratively done while starting by matching the more important regions in the query-image, according to the predefined weights $w_j (1 \leq j \leq n)$. Let's note that the used region matching process is partially robust against inaccurate segmentation effects, especially those caused by occlusion, since it allows that many query regions can be matched with the same region of the tested image. The node-based matching is then iteratively improved while considering second-order inter-regions spatial relationships. Indeed, the initial region-based matches are used as evidence for a regression model, based on dynamic programming [13], which estimates visual and topological correspondence across the entire image. These estimates allow further matches to be discovered and refined. We estimate and utilize probabilistic weights for each correspondence, allowing the algorithm to detect and discard mismatches (Fig. 2). These weights are mainly defined according to likelihood of the n nodes and of the $(n^2 - n)/2$ edges relating them. Finally, the overall similarity measurement $S(1)$ between the query-graph $G(|G| = n)$ and a tested graph $G^{'} (|G^{'}| = t)$ is defined by the inverse of the weighted average distance between the matched regions as well as between the matched edges $(E_j, E_j^{'}) (1 \leq j \leq (n^2 - n)/2)$. The proposed graph-based image similarity evaluation preserves the following desired characteristic of distance metric: "the distance between the same images equals to 0" [14].

$$S(G, G^{'}) = \left[\frac{1}{n} \sum_{j=1}^{n} w_j d(R_j, R_j^{'}) + \frac{2}{n(n-1)} \sum_{j=1}^{n} w_j d(E_j, E_j^{'}) \right]^{-1} \quad (1)$$

Where d denotes the Euclidean distance between two feature vectors, such that R_j is a 3-uplet ($\in [0,1]^3$) composed of each color-component average within the j^{th} region and E_j is a 5-uplet ($\in [0,1]^5$) describing the spatial disposition of the j^{th} region relatively to another region inside the same image.

3 Relevance Feedback Based on Dynamic Region-weighting

The principle objective of the suggested region-based relevance feedback scheme is to adjust query-region visual importance from the users' positive feedbacks. Indeed, if the user is dissatisfied by the retrieved images, the proposed relevance feedback scheme asks him to indicate a set of relevant images within the already produced retrieval. In our case, each iterative interaction with the user consists of one

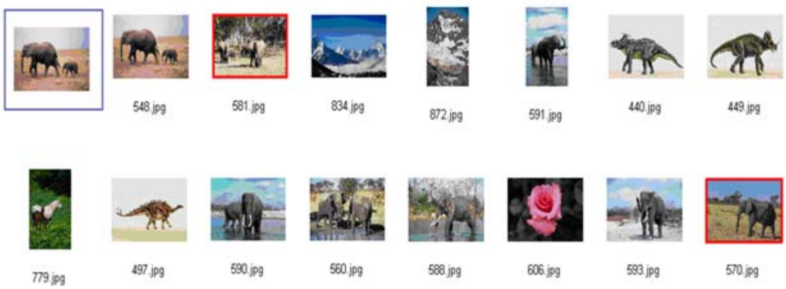


Fig. 2: The preliminary retrieved images sorted according to the similarity measurement. The blue-framed image is the query and the following ones are the 15 most similar images.

unordered collective positive feedbacks selection. While adopting the “most similar, highest priority” paradigm [15], query-regions weights are heuristically updated based on the feedback data. Considering the used graph-based similarity measurement scheme, the basic assumption behind our region-based RF is that each of the already retrieved images should have at least one region which is highly similar, in terms of low-level features, to one (or more) region(s) of the request image. Nevertheless, it is usual that an image was retrieved because it has some regions which are highly similar to ”unimportant” query-regions, relatively to user’s objectives of the retrieval. These undesired regions, which can be significant but do not represent the focus of the user’s perceptions of the image content, corresponds in most of cases to background regions. Thus, after selecting relevant images by the user, the proposed RF scheme identifies among the query-image regions, the important ones in order to focus on them during the next iteration of retrieval. In fact, in most of the studied cases, all positive feedbacks share same visual objects, what maximizes the possibility that many regions inside user’s relevant images are associated with the same regions in the query graph. Then, that element is a crucial region (object) according to the user’s expectations, and thereby its weight should be empirically increased ($+\alpha$). Otherwise, each query-region which is not highly similar to at least one region of a relevant image should be accordingly penalized. The basic assumption is that important regions should appear more times in the positive images and fewer times in all the images of the database [16].
Given the query-image I and the subset ξ of relevant images $I_k^{\prime}(1 \leq k \leq p)$, interactively selected by the user, the proposed region-based relevance feedback scheme (Fig: 3) can be summarized as follows:

The used weight updating respects the requirement that the sum of query-regions weights should be equal to 1. However, it is found that after a relatively high number (≥ 4) of successive RF iterations, some regions are weighted with negative values. These regions are above all those with low initial weights, which generally

```
for each relevant image I'_k in ξ do
    cpt ← 0;
    for each region R_j in the query-image I(1 ≤ j ≤ n) do
        Ω_j ← {R'_j ∈ I_k such that d(R_j, R'_j) ≈ 1}
        if Ω_j ≠ ∅ then
            w_j ← w_j + α;
            cpt ← cpt+1;
        else
            w_j ← w_j − ((α * cpt)/(n − cpt));
        end if
    end for
end for
```

correspond to insignificant request-image regions. Thus, the weight updating of these regions permits to compensate inaccurate segmentation, particularly over-segmentation effects, inevitably caused by watershed-based segmentation techniques. Besides, many regions in a relevant image, selected by the user, can be assigned to the same request region what ensure further robustness against inaccurate segmentation affected by occlusion. Thus, a relatively reduced small number ($1 \sim 3$) of selected relevant images can considerably improve the performance in subsequent iterations (Fig. 4), and further expedite the subsequent retrieval at the same time [14]. In fact, since the used similarity measurement is highly based on query-regions weights, much more semantically similar images, relatively to the user personal expectations, will be efficiently retrieved. Fig.3. illustrates how the proposed region-based relevance feedback discovers, given two selected relevant images (the red-frames images in Fig. 2), the more appropriate query-regions for explaining the user's expectations, by using weight updating. Based on our tests, the optimal value of α was empirically fixed at 0.2.

4 Retrieval Effectiveness Assessment

To evaluate the retrieval effectiveness of the proposed interactive system, a subset consisting of 265 images is arbitrarily chosen from Corel Photos and FeiFei databases. These images contain a wide range of contents such as scenery, animals, vegetation, people and cars. The used image database was manually classified into the following seven categories: buses, cars, elephants, dinosaurs, flowers, horses and mountains. Each category contains about 38 images, each of which can be a query-image candidate. As evaluation criterion, we used the parametric precision-recall curve which is standard measure to identify the performance of CBIR or RF schemes [17]. Indeed, precision-recall curves of Fig. 5 evaluate the contribution of the spatial disposition while applying the proposed graph-based RBIR, with and without edges consideration during the graph matching process. The produced curves evaluate as well the contribution of multiple RF rounds. To be fair, the same features as well

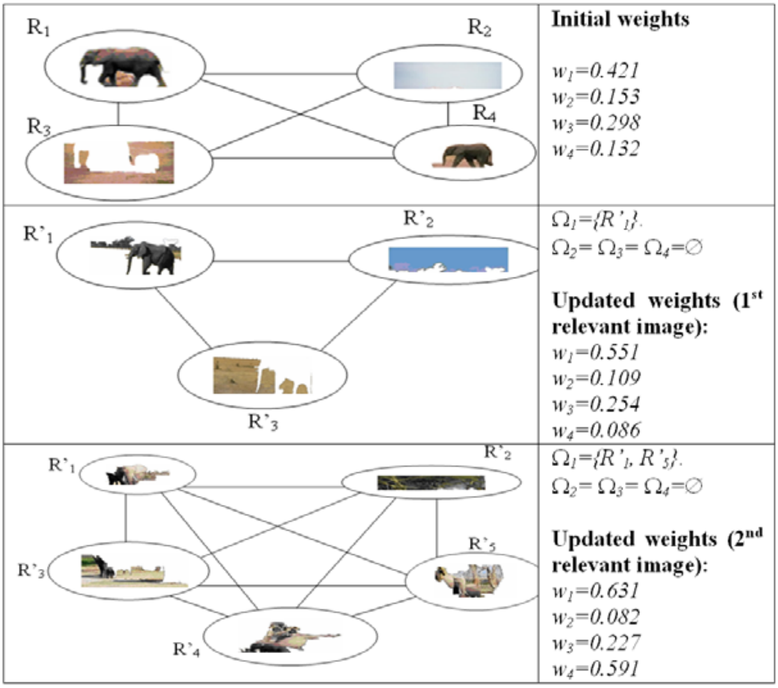


Fig. 3: An example iteration of the proposed region-based RF scheme.

as the same number (*two*) of relevant images in a RF round are used for all representations. We note also that only images recording ”high” similarity measurement (upper a predefined threshold) are returned by the retrieval system, such that a retrieved image is considered as a true match in the evaluation only if it belongs to the same category of the query image. Experimental results show that the suggested retrieval system allows recording encouraging results. In particular, the accuracy of the retrieval is higher for the categories of horses and elephants, since the background of images belonging to these categories are further simple. However, the lower retrieval accuracy was recorded for the buses category. This is mainly due to the fact that buses images contain complicated background with diverse features under different poses. Lastly, the proposed region-based RF scheme improves substantially the retrieval performance of our region-based image retrieval system. It tailors the system semantic behavior relatively to each user personal preferences.

5 Conclusion and Perspectives

This paper presents an interactive CBIR system by using a graph-based similarity measure. There are two novel points in this work, including (1) a fuzzy graph-based representation is proposed to encode the regional spatial relationship for an image; (2) adopting the relevance feedback to update the region weights so that the retrieval results are refined by the further graph matching process.

During offline preliminary preparations, region-based graph of all images in the database are extracted automatically by using the fuzzy logic. During online process, when a query image is supplied by the user, all of the images are sorted according to their similarity to this query. To evaluate similarity between two graphs, fuzzy inter relations among regions feature vectors and spatial dispositions as well as fuzzy regions weights are explored. Moreover, we designed a simple relevance feedback for region weighting by use of user's positive feedbacks. Indeed, if the user is not satisfied with the retrieval results, he can specify a subset of relevant images to refine the retrieval results in the following iteration. The incorporation of the RF process on the set of returned images minimizes the multiple interpretations for the same query image, which is a usual disadvantage of graph CBIR systems [18].

In fact, the proposed region-based RF scheme allows considerably removing the irrelevant outputs and enhancing the relevant ones. We note that the proposed weighting-region-based image matching and relevance feedback algorithms are applicable for any RBIR system. Actually, we try to improve the proposed region reweighting technique, while adapting a probabilistic framework. The main idea consists to update each "important" region in the query-image relatively to the degree of similarity between this region and the correspondent one(s) in each relevant image. Indeed, this degree simulates the probability that each couple of similar region (in the query-image and in a relevant image) corresponds to the same real object. We plain also to extend the proposed region-based RF framework for the general case of positive and negative feedbacks, where users can cast relevant and irrelevant images, in order to better express user's particular interests. Thus, all preliminary retrieved images can be labeled, in accord with their relevance to user's expectations, into many feedbacks groups.

References

1. El-Kwae, E.A., Kabukai, M.R.: Efficient Content-Based Indexing of Large Image Databases. ACM Trans. Information Systems 18, 171–210 (2000)
2. Ko, B.C., Byun, H.: Query-by-Gesture: An Alternative Content-Based Image Retrieval Query Scheme. J. Vis. Lang. Comput. 13, 375–390 (2002)
3. Jiang, W., Er, G., Dai, O., Zhong, L., Hou, Y.: Relevance Feedback Learning with Feature Selection in Region-based Image Retrieval. In: IEEE International Conference on Acoustics, speech, and Signal, pp. 509–512. IEEE Press, Pennsylvania (2005)
4. Jing, F., Li, M., Zhang, H.J., Zhang, B.: An Efficient and Effective Region-based Image Retrieval Framework. IEEE Trans. Image Processing 13, 699–709 (2004)

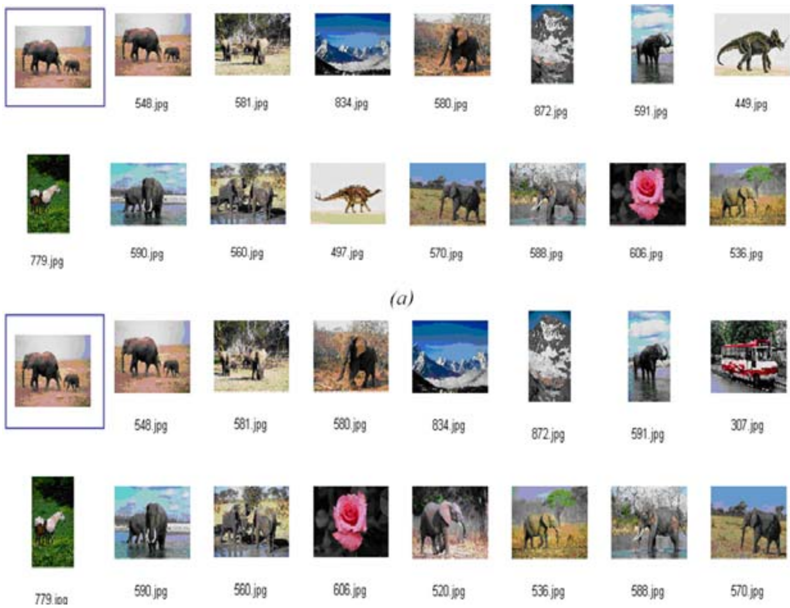


Fig. 4: Retrieved images after applying the RF scheme. a- after one iteration, b- after three iterations (two relevant images are selected in each RF iteration).

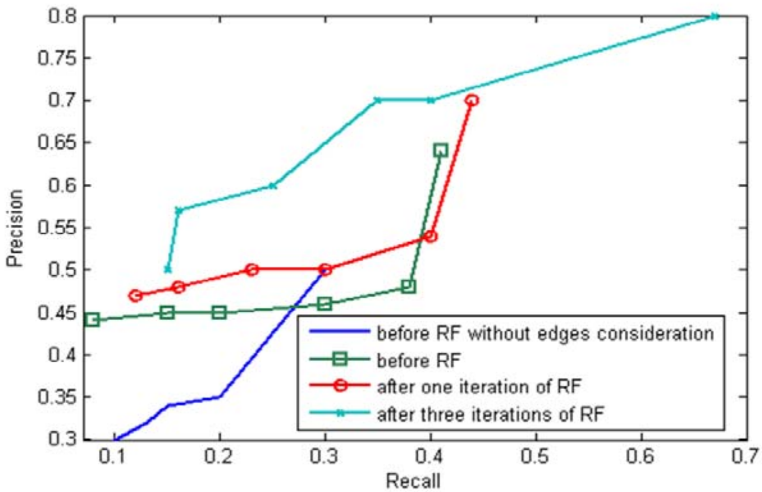


Fig. 5: APrecision-recall curves.

5. Jiang, F., Li, M., Zhang, H.J.: Entropy-based Active Learning with Support Vector Machines for Content-based Image Retrieval. In: IEEE International Conference on Multimedia and Expo., pp. 85–88. IEEE Press, Taipei (2004)
6. Su, W.T., Chu, W.S., James-Lien, J.J.: Heuristic Pre-clustering Relevance Feedback in Region-based Image Retrieval. In: Narayanan, P.J., Nayar, S.K., Shum, H.-Y. (eds.) ACCV 2006. LNCS, vol. 3852, pp. 294–304. Springer, Heidelberg (2006)
7. Zhang, R., Zhang, Z.: FAST: Fast and Semantics-Tailored Image Retrieval. In: Petrushin, V.A., Khan, L. (eds.) Multimedia Data Mining Knowledge Discovery. Springer, Heidelberg (2007)
8. Ko, B., Peng, J., Byun, H.: Region-based Image Retrieval using Feature Relevance Learning. J. Pattern Anal. 4, 174–184 (2002)
9. Dorado, A., Djordjevic, E., Izquierdo, E., Pedrycz, W.: Supervised Semantic Scene Classification based on Low-level Clustering and Relevance Feedback. In: European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, London, pp. 181–188 (2004)
10. Jing, F., Li, M., Zhang, L., Zhang, H.J., Zhang, B.: Learning in Region-based Image Retrieval. In: IEEE International Symposium on Circuits and Systems, pp. 206–215. IEEE Press, Cairo (2003)
11. Zagrouba, E., Ouni, S., Barhoumi, W.: A Reliable Image Retrieval System Based on Spatial Disposition Graph Matching. International Review on Computers and Software 2, 108–117 (2007)
12. Wang, T., Rui, Y., Sun, J.G.: Constraint Based Region Matching for Image Retrieval. J. Computer Vision 65, 37–45 (2004)
13. Karouia, I., Zagrouba, E., Barhoumi, W.: Video Similarity Measurement Based on Attributed Relational Graph Matching. In: Nguyen, N.T., Katarzyniak, R. (eds.) New Challenges in Applied Intelligence Technologies, pp. 173–182. Springer, Heidelberg (2008)
14. Zhang, R., Zhang, Z.: Addressing CBIR Efficiency, Effectiveness, and Retrieval Subjectivity Simultaneously. In: ACM International Workshop on Multimedia Information Retrieval, pp. 71–78. ACM Press, New York (2003)
15. Li, C.Y., Hsu, C.T.: Image Retrieval with Relevance Feedback based on Graph-theoretic Region Correspondence Estimation. IEEE Trans. Multimedia 10, 447–456 (2008)
16. Jing, F., Li, M., Zhang, H.J., Zhang, B.: Learning Region Weighting from Relevance Feedback in Image Retrieval. In: IEEE International Conference on Acoustics. IEEE Press, Orlando (2002)
17. Muller, H., Muller, W., Squire, D.M., Maillet, M.S., Pun, T.: Performance-Evaluation in Content-based Image Retrieval: Overview and Proposals. Pattern Recognition Letters 22, 593–601 (2001)
18. Jung, S.H.: Content-based Image Retrieval using Fuzzy Multiple Attribute Relational Graph. In: IEEE International Symposium on Industrial Electronics, pp. 1508–1513. IEEE Press, Aquila (2002)

Workshop and Invited Session

International Workshop on Human-Computer Interaction in Knowledge-based Environments

Prof. Maria Virvou and Prof. George A. Tsihrintzis
University of Piraeus, Greece
e-mail: {mvirvou,geoatsi}@unipi.gr

Multimedia Techniques for Device and Ambient Intelligence MTDAI

Dr. Marco Anisetti and Dr. Valerio Bellandi
Università degli Studi di Milano, Italy
e-mail: {marco.anisetti,valerio.bellandi}@unimi.it

Intelligent Systems for Healthcare

Dr. Luigi Gallo and Dr. Giuseppe De Pietro
National Research Council (ICAR), Italy
e-mail: {luigi.gallo,giuseppe.depietro}@na.icar.cnr.it

Design of Intelligent Environment for Supporting Human Interaction and Creation

Prof. Toyohide Watanabe, Prof. Shigeki Matsubara, Prof. Fuminori Akiba and Prof. Takami Yasuda
Nagoya University, Japan
e-mail: {watanabe,matubara,akibaf,yasuda}@is.nagoya-u.ac.jp

The Feasibility of Auto Profiling Online

Manfred Pienemann, Bi-Jar Lin, and Long-Yeu Chung

Abstract In this paper we will demonstrate the feasibility of using Auto Profiling (AP) online to analyse an informant's linguistic data and his/her language distributions. AP is an extension of the linguistic profiling approach that was originally carried out manually. This procedure was further developed into a computer-based procedure (Rapid Profile) (RP) that assists the analyst, although this is still essentially a manual observation-based procedure. The automatic profiling approach presented in this work is the first fully automatic procedure for linguistic profiling. It accepts free-style written interlanguage input and automatically assigns a C-structure and aspects of the L2 morphology to the interlanguage input. On this basis the system is capable of reliably assessing the stage of acquisition of the input sample. In addition, it gives feedback to the specific errors contained in the input, and can also be used for automatically tagging L2 corpora. At the end of this paper, we will provide an analysis of real data to demonstrate that the Auto Profiling algorithm is not only based on the application of Processability Theory (PT) to Specific L2s and on the logic of RP, but also can be used online to profile an informant's data in real time under time constraints.

Manfred Pienemann
University of Paderborn and University of Newcastle upon Tyne

Bi-Jar Lin
Tung Fang Institute of Technology and University of Newcastle upon Tyne

Long-Yeu Chung
Chia Nan University of Pharmacy and Science

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 191–200.
springerlink.com

1 Introduction

The ideal situation, in linguistics and other fields, is to be able to analyse real data in real time. However, attempts to achieve this have not been very successful, especially in the context of second language learning situations. This is mainly due to the approaches that have been applied in tackling this problem, such as probability methods. Although using these methods can undertake quite a lot of data analysis, they will eventually come to a point where run-time errors will be produced, such as with garden path sentences. In this work, we actually expand on a paper presented at the 8^{th} International Symposium on Processability Approaches to Linguistic Acquisition (PALA) at Verona, Italy, in 2008, and provide evidence to show the feasibility of using Auto Profiling online. The paper first discusses how computers can be used in analysing data. We will then explain why current computer programs cannot be properly used in analysing linguistic data, especially that from second language learners. After that, we will demonstrate the importance of linguistic knowledge in analysing data. We will explain how Auto Profiling based on Processability Theory (PT) (Pienemann 1998, 2005) [1][2] can be used to analyse data and solve the current computer garden path problems. In addition, we will demonstrate how the algorithms from PT and AP can be used online. Finally, to prove that it is feasible to use AP online to analyse data, we will list the methods used and the results of applying them to real data.

2 Literature Review:

Imagine that we are in a situation where a computer will automatically collect data from an informant, which will then be automatically analysed, with the results appearing in real time online. Moreover, the correct and incorrect sentences are automatically and immediately annotated, without any run-time errors. Unfortunately, however, this vision cannot currently be realized, due to the following problems:

2.1 First problem: tractability

The issue of tractability is very important in determining whether a parsing method is going to work or not. Currently, most parsers take advantage of corpora data and use lexical frequency to design their algorithms. By doing this, they can save a lot of time in determining the structure of a sentence. For example, the words that are likely to appear after the term ‘eat’ can be predicted by examining a large amount of data, and thus a parser can determine the structure of a sentence. These method work well with data which have similar sentences as the original corpora, however, when other types of data are used which do not have similar patterns, this method will not work because the parsers cannot predict what a second language learner’s text structure is. For example, the possible items after the word ‘eat’ can be anything from ‘ate’, ‘take’, ‘we’, ‘to’, and so on, meaning there is no way for the probability methods to be used in the algorithm because almost anything is possible. Therefore, while probability methods can be a possible approach for a native speaker or an advanced second language user, they are far less suitable for use with a second language learner.

2.1.1 Computers are very good at analyzing data

Pennington and Stevens (1992, p4) [3] pointed out that Doughty successfully established several reasons for employing computers in research on second language acquisition. In fact, they (p5) [3] further stated that the pre-condition for successful implementations of computer-based projects in applied linguistics depends on whether there are powerful and flexible software tools which are capable of assisting the applied linguist in their research and language study.

2.1.2 Current parsing methods cannot analyse a second language input in real time

Pennington and Stevens further noted that Sampson (2005) [6] also agreed that it is appropriate to use computers in applied linguistics, and that they can be of benefit to linguistic researchers with regard to increasing the availability of databases to be analysed. Sampson even argued that manual analysis can provide the statistical data necessary for improving techniques in automatic parsing, which could in turn be applied to the automatic parsing of an even greater range of machine-readable corpora. However, this approach is not applicable for second language learners’ texts, as a statistical approach will encounter a lot of garden path sentences due to the unpredictable structures produced by such learners. Pienemann also pointed out this phenomenon. Meanwhile, Pienemann also noted this phenomenon, and also stated (p202) [2] that one crucial feature of all interlanguage varieties of a given target language is that they share a lexicon, an as yet undetermined subset of the rule system of that target language and a set of rules which are neither part of the source nor target language. Thus, many formal features of these developing linguistic systems are unknown to the researchers. Therefore, Pienemann argued that it appeared inappropriate to attempt to write a single parser for the analysis of such data, as can be done for the analysis of that produced by native speakers, because the construction of a parser presupposes knowledge of the rule system of the language to be parsed.

2.1.3 Linguistics knowledge is essential in parsing natural languages

The way to know whether a problem is tractable or not is by comparing it to others whose tractability is already known. In a related vein, Barton et al. (1987, p2)

[4] state that 'complexity theory studies the computational resources - usually time and memory space-needed to solve particular problems, abstracting away from the details of the algorithm and machine used to solve them.' This theory also provides robust classification schemes - complexity classes - which tell us whether certain problems are likely to be computationally tractable or intractable. Note that here the word 'tractable' roughly means always solvable in a reasonable amount of time on an ordinary computer, and thus some problems that are currently intractable can be expected to become tractable at a later date, when computer hardware and software have improved sufficiently. Barton et al. (p2) [4] then clarify that this classification holds regardless of any other factors, such as what algorithm we use or how many highly competent programmers are hired to work on the task. In short, they state that a hard problem cannot be switched into an easier complexity class by using a more ingenious algorithm. They therefore conclude that since the approaches based on computers and algorithms cannot currently deal with problems which originate from natural languages, we should instead rely on the help of linguists to undertake the language processing.

2.1.4 PT is the linguistic knowledge for parsing natural languages

Over the last twenty years, linguistic theories have developed significantly, especially those related to the processing mechanisms required for abstracting natural language problems. The specific theories examined here are Processability Theory (PT) and Auto Profiling. PT is used to tackle problems that are otherwise intractable, through the features of different levels of developmental stages of English. This is because the structural options available to the learner are constrained by the processing procedures and architecture of the target language (p.169) [1] - and it is on these that the predictive power of Processability Theory is based.

2.2 Second problem: annotating an incorrect sentence

2.2.1 Constituent structure

There are many ways to label a constituent structure, or C-structure, in the field of linguistics. In fact, Langacker (1973, p 108) [5] noted that one should not conclude that surface structures are fully understood or that it is always an easy matter to ascertain the precise tree structure to be associated with a given string of morphemes In addition, some conceptual structures of sentences have no direct surface realization. Indeed, Langacker (p 110) [5] pointed out that some components in the conceptual structures of sentences have no direct surface realization at all. Consequently, it is impossible for researchers to manually label all C-structures. To address this problem, many automatic C-structure parsers have been designed, with the aim of solving the problems of inconsistency that arise from human errors. Unfortunately, most automatic parsers still operate at a preliminary level of data annotations, and specifically are either too detailed or not detailed enough to be used to examine material produced by second language learners. When too detailed, the parser either does not provide enough essential information to show the whole picture of a learner's language knowledge, or produces results that do not relate at all to such knowledge. However, when an automatic parser is not detailed enough, it tends to produce too simple a picture of a learner's language knowledge. In addition, more often than not, such parsers are based on corpora which are a set of aggregated data, and thus they cannot directly describe what really happens to any specific second language learner. Finally, real time annotations are currently beyond the capabilities of all these systems.

2.2.2 What is the purpose of parsing?

Sampson (2005) [6] defined "parsing" as detecting the structure implicit in superficially-unstructured sequences of items. He also stated that linguists represent such structures by using tree diagrams, or by inserting labeled brackets among the words. For example, with the SUSANNE parsing scheme a sentence is represented as: [S [Ns:s The quick brown fox] [Vz jumps] [P:q over [Ns the lazy dog]]] In addition, Sagae et al. (2004) [7] write that parsers use a computational model of natural language to analyze a sentence, producing its syntactic structure as the output. This structure may be represented as a constituent tree (C-structure or parse tree), a syntactic feature structure (or f-structure), or a dependency structure. These differences may make the outputs appear unrelated, but in reality these various representations all show how words are combined together to form a sentence.

2.2.3 Problems with some current parsers with regard to annotation

2.2.3.1 The processed form does not meet the researchers' needs

Sagae et al. [7] noted that although there are now several hundred studies that have used the CHILDES database to study the development of morphosyntax, most of these have been forced to use the database in an almost raw and unprocessed lexical form - without tags for parts-of-speech and without syntactic parses. Even if parsers can annotate the texts, they are not designed for use by researchers. Sagae et al. [7] observed that over the past decade a number of annotation efforts have resulted in large amounts of text being annotated with syntactic parse trees, known as "treebanks" (Marcus et al. 1993) [8]. However, this kind of annotation is designed to facilitate the training of statistical language analysis tools, instead of the study of language acquisition.

2.2.3.2 Current parsing methods cannot parse the texts of second language learners

Sagae et al. [7] stated that LCFlex uses the most probable succession of rules as a way to solve the statistical syntactic disambiguation in each analysis of a particular utterance obtained through an ordered succession of grammar rule applications. The probability of each competing analysis is determined based on a statistical model of bigrams and rule applications obtained from a training corpus (Sagae et al. 2004, cited in Rose & Lavie, 2001) [9], with this corpus required to disambiguate different situations so as to find the correct analyses. For example, instead of having a rule VP → V NP with probability 0.8 based on training data, the probability of rule VP→V NP is 0.6 if its parent rule is S→NP VP, and 0.2 if its parent rule is VP→VP PP. Therefore, these rule probabilities not only disambiguate the parsing problems, but also are sensitive to context. Another setback is that using statistical language analysis is limited to the texts from native speakers. When it comes to second language learners, their production of language cannot be predicted as well as that of native speakers. Consequently, probability methods cannot be used because these learners do not follow the predicted probability. While other approaches, such as using stack methods, will solve some local problems, they are not useful for whole sentences. Moreover, these stack methods also use the native speakers' corpora, which are quite different to those of second language learners. This is why in this paper we try to demonstrate the feasibility of using another approach to annotate and parse the language knowledge of a second language learner, thus avoiding the problems that arise with both probability and stack methods. In addition, the parsing method we propose here can annotate the language production in real time. Furthermore, it will not have the problems commonly found in most of the current parsing methods, such as going through a lot of unnecessary routes or garden paths. The method used here is very similar to using rule comparison as a base with a parsing concept in the algorithm. This can make annotation easier and more reliable, with no manual correction required. Specifically, while other parsers try to find out the most likely combination of syntactical structures for a sentence such as 'I can fly', AP based on PT can use the developmental stages in English to decide which combination is more likely without resorting to a probability analysis, nor by using POST, which takes a long time to find the most likely combination (Sagae et al., 2004, cited in Parisse & Le Normand, 2000) [10]. For example, in any sentence, AP can find several combinations of syntactical structures, and then relies on the different developmental stages to discover the most appropriate one. To achieve this goal, the results of different developmental stages will be compared and then the highest stage will be chosen as the structure of a sentence. For example, the sentence "I can fly" can have the structures - pronoun auxiliary-verb or pronoun noun-verb. According to the algorithm in the Correct Ranking Bank data, which classifies pronoun auxiliary-verb as stage 2 and pronoun noun-verb as stage 1, the former structure will be used because it has the higher developmental stage. Therefore, with the help of PT, AP can easily identify the most likely structure without having to use probability.

3 Method:

AP will be tested online under time-constrained conditions as well as in real time to see whether it can really achieve what it is designed to do. The following features of the system will be demonstrated.

3.1 How informants' data is analysed

Pienemann (p162) [1] pointed out that although forms may emerge in the same shape as in the target language, learners may intend them to relate to different functions. Therefore, Pienemann argued that only through a careful and unbiased distributional analysis can the genuine systematicity of the interlanguage be found. Consequently, distributional analysis is required to check whether or not learners have really acquired what they seem to have learned. AP is equipped with several functions to find out the interlanguage distributions of a learner. First, when the data have been input into AP, it will check each word type in real time, and each word may have more than one type. After coding all the words in a sentence, AP will produce one or more strings of word types. For example, if a word like 'work' has two types - a noun and a verb, there will be at least two strings of word types in this sentence. In addition, if there are two words - each with two word types, there will be four strings of word types - 2 * 2 = 4. When each word in a sentence has been classified according to its word type, AP will match these strings with the Correct Ranking Bank data based on the Processability Theory. The Correct Ranking Bank contains different types of strings along with the sentence's temporary developmental levels of English, based on Processability Theory. After being matched, each sentence will have at least one temporary developmental level of English. In cases where there is more than one level, the highest one will be chosen. When all the sentences are coordinated, the AP interfaces will show all the highest temporary levels of the sentences. After analysing them, the AP will use emergence criteria to further analyse an informant's data so as to know his/her final developmental stage and other interlanguage distributions.

3.2 How the results are displayed

There are two main parts in the display function - Sentences from Practice and Sentences from Tests, and both are the same with regard to how each sentence is coded, analysed, displayed and, corrected,. Sentences from Practice contains the material produced by an informant during practicing, and once this is input each sentence will be coded, analysed and displayed - showing its temporary level, C-structure and ID number. This number is very useful, because it can be referred to later on anywhere in this system. Researchers in particular can use the ID number to trace related results in the system, such as its C-structure, temporary highest level of developmental stage, suggested sentence, and syntactical variation information, as well as other sentences produced by the same informant. One of the key displays in AP is the feedback. When there are no major mistakes, the feedback will show that no errors were found. However, when there are some significant mistakes, the corrected sentences will be shown, and there may be more than one corrected sentence for each one input. In the Sentences from Tests part, the informants' real test data are analysed and displayed in the same way as in the Sentences from Practice, with all of the sentences from both parts shown at the same time. Informants can thus click on any sentence they have produced and check its temporary level, find out if there are any errors, and then see the related corrections.

3.3 How some of the features can be modified

Although AP uses data banks to code and analyse informants' data, it is very likely that the material provided by these users will not be in these, especially if they are second language learners. There are several ways to tackle this problem, such as using MySQL data banks to modify the AP programs. Specifically, for data which has nothing to do with the fundamental concept of the analysis, AP provides MySQL data banks for researchers to change or add more data to in order to meet their research needs. However, if it is related to the fundamental concept, the AP can also be easily modified. This is mainly because all the AP programs are modularized, such as the Correct_ranking_bank module, C_structure module, and Distribution_rules module. Consequently, whenever there is a fundamental problem, only the modules which are related to it need to be changed, while the others can still be used. However, for problems that do not require fundamental modifications, researchers can just add more entries to the data banks to expand their capacity. To achieve this purpose, researchers simply click on the MySQL browse function and look for the AP config file, where they can click on the specific data bank to add or delete entries. After adding or deleting material, researchers can go back and analyse informants' data by restarting AP to load the new data bank, without going through compiling process-a process which most researchers are not familiar with. Other data bank files can be modified in a similar fashion. For example, the Error Tracer module can be altered by finding what error message needs to be suggested and then adding it to the existing set.

3.4 How the modifications are evaluated

To evaluate the modifications to AP, researchers can simply check the implication scale of all the data longitudinally collected from a specific informant, as well as their interlangauge distributions, to see whether the results of these data meet the construct validity. For instance, when researchers find out that an informant has apparently acquired a certain developmental stage without having acquired the lower level or levels, the validity of these modifications or the overall AP system will be called into question. Another way to evaluate the modifications is by checking whatever data is misplaced. For example, if a sentence is wrongly coded, the faulty data should be corrected, and then the sentence should be input again to check the result of the new modifications. If the problem still exists, then more modifications to the system must be undertaken.

3.5 Results

3.6 General findings

The results of the online and on-site investigations are exactly the same, showing that Auto Profiling can be used in any kind of situation, and is thus a very adaptive system. Moreover, the results show that AP is also very reliable, always providing the same feedback for the same data in real time, including interlanguage distribution and annotated sentences. In addition, when a word or a structure is not in the data bank, they can easily be added, and the system then works as it is supposed to. This shows that modifications can be successfully carried out without affecting the whole system. In fact, in these tests, all the data was processed without any garden path problems being encountered. Finally, the most important finding is that AP can provide the developmental stage of each informant online and in real time. Furthermore, in our tests we did not find any informant who had not achieved at least the lower level developmental stages, and this means that Processability Theory (PT) - the basic theory of AP - is appropriate for use in profiling the language knowledge of learners.

4 Conclusions

With the results outlined above, we are quite confident of the feasibility of using AP online. Due to the use of Processability Theory, especially the time-constrained function, the results are not only reliable, but also valid. In addition to oral input, AP can be also used for transcribed texts, which can help researchers find the differences between speaking and writing. Furthermore, AP can be modified and checked easily, especially the word and structure banks, which means that it is a flexible system. Finally, AP can elicit, analyse and provide the results of data online in real time, including the developmental stages of English and inter-language distributions.

References

1. Pienemann, M.: Language Processing and Second Language Development: Processability Theory. Benjamins, Amsterdam (1998)
2. Pienemann, M.: Cross-linguistic Aspects of Processability Theory. Benjamins, Amsterdam (2005)
3. Pennington, M.C., Stevens, V.: Computers in Applied Linguistics. Multilingual Matters, Clevedon (1992)
4. Barton, G.E., Berwick, R.C., Ristad, E.S.: Computational Complexity and Natural Language. The MIT Press, London (1987)
5. Langacker, R.: Language and its Structure: Some Fundamental Linguistic Concepts, 2nd edn. Harcourt Brace Jovanovich, USA (1973)
6. Sampson, G. (2005), http://www.grsampson.net/RLeafAnc.html
7. Sagae, K., MacWhinney, B., Lavie, A.: Behavior Research Methods. Instruments, & Computers 36(1), 113–126 (2004)
8. Marcus, M.P., Santorini, B., Marcinkiewics, M.: Building a large annotated corpus of English: the Penn Treebank. Computational Linguistics 19, 313–330 (1993)
9. Rosé, C.P., Lavie, A.: Balancing robustness and efficiency in unification-augmented context-free parsers for large practical applications. In: Junqua, G.-C., van Noord, G. (eds.) Robustness in Language and Speech Technology. Kluwer Academic Press, Holland (2001)

Progress Evaluation Based on Fuzzy Relationships

Sylvia Encheva and Sharil Tumin

Abstract In this paper we present a method for early identification of students who would have serious difficulties in studying a particular subject due to lack of preliminary knowledge. Methods from formal concept analysis are applied for finding dependencies among results from a test on students' preliminary knowledge and the outcome of learning new knowledge within the same subject. Fuzzy logic rules are further incorporated into the process of establishing level of difficulties a student would have while learning new concepts and skills within a particular topic in that subject.

1 Introduction

Calculus is a compulsory subject for first year engineering students on bachelor level at many universities. One of the most difficult topics to learn appears to be integration. Our goal is to find a way to identify students who are exposed to a serious danger of not being able to obtain sufficient knowledge and skills in integration due to lack of preliminary knowledge in mathematics. Once these students are identified they will be suggested to take additional classes in mathematics. This will decrease the amount of students failing exams in mathematics as well as the amount of students failing exams in other subjects that require mathematical skills. In addition it will facilitate the process of allocating teaching resources at the beginning of a semester.

Sylvia Encheva
Stord/Haugesund University College, Bjørnsonsg. 45, 5528 Haugesund, Norway
e-mail: sbe@hsh.no

Sharil Tumin
University of Bergen, IT-Dept., P. O. Box 7800, 5020 Bergen, Norway
e-mail: edpst@it.uib.no

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 201–210.
springerlink.com

Fuzzy reasoning methods are often applied in intelligent systems, decision making and fuzzy control. Some of them present a reasoning result as a real number, while others use fuzzy sets. Fuzzy reasoning methods involving various fuzzy implications and compositions are discussed by many authors, f. ex. [1], [2], [4], [5], [8], [9], [10], [11], and [13].

An interesting prediction method presented in [6] applies formal concept analysis and fuzzy inference. In particular it shows how to calculate the value of a membership function of an object if the object belongs to a known concept. The authors however do not state whether classical set theory or fuzzy set rules are applied in establishing an object's belonging to a concept. It seems that the method can be applied only to objects with predetermined concept membership. Subsequently the method does not show how to handle objects that partly belong to more than one concept. In this paper we propose solutions to these questions.

The rest of the paper is organized as follows. Section 2 contains definitions of terms used later on. Section 3 describes how we propose to select students that need extra classes and Section 4 is devoted to a system description. Section 5 contains the conclusion of this work.

2 Background

The mathematical preparedness of students embarking upon science and engineering degree programmes has been the subject of close scrutiny over recent years, with disheartening conclusions [14]. Since mathematics is a key facet of all engineering degree courses it appears also to be one of the reasons for increased dropout rates and poor student progression.

Let P be a non-empty ordered set. If $sup\{x,y\}$ and $inf\{x,y\}$ exist for all $x,y \in P$, then P is called a *lattice* [3]. In a lattice illustrating partial ordering of knowledge values, the logical conjunction is identified with the meet operation and the logical disjunction with the join operation.

A *context* is a triple (G,M,I) where G and M are sets and $I \subset G \times M$. The elements of G and M are called *objects* and *attributes* respectively [3], [12].

For $A \subseteq G$ and $B \subseteq M$, define

$$A' = \{m \in M \mid (\forall g \in A) \;\; gIm\},$$

$$B' = \{g \in G \mid (\forall m \in B) \;\; gIm\}$$

where A' is the set of attributes common to all the objects in A and B' is the set of objects possessing the attributes in B.

A *concept* of the context (G,M,I) is defined to be a pair (A,B) where $A \subseteq G$, $B \subseteq M$, $A' = B$ and $B' = A$. The *extent* of the concept (A,B) is A while its *intent* is B. A subset A of G is the extent of some concept if and only if $A'' = A$ in which case the unique concept of the which A is an extent is (A,A'). The corresponding statement applies to those subsets $B \in M$ which is the intent of some concepts.

The set of all concepts of the context (G,M,I) is denoted by $\mathfrak{B}(G,M,I)$. $\langle \mathfrak{B}(G,M,I); \leq \rangle$ is a complete lattice and it is known as the *concept lattice* of the context (G,M,I).

The *sum-of-1-criterion* [7] states that

$$\Sigma_{i \in M_i} m_i(x) = 1, \quad \forall x \in \chi$$

where $M_i, i = 1,...,k$ denotes all possible membership terms $\{m_i, i = 1,...,k\}$ of a fuzzy variable in some universe of discourse χ.

An affiliation value α to a concept (A,B) is defined as

$$\alpha(A,B) = \frac{\Sigma_{o \in A, f \in B} m_{of}}{|A| \cdot |B|}$$

The affiliation value represents the relative extent to which an object belongs to this concept or an attribute is common to all objects in the concept.

By θ we denote the threshold for membership values above which an entry is regarded as significant, as in [6]. This is achieved by computing the arithmetic mean of all entries within a column and take it as a threshold.

3 Correlations

Based on real data from previous years we first prepare Table 1 that illustrates correlations between students' preliminary knowledge and already obtained knowledge and skills in integration. Using formal concept analysis we can extract all concepts based on the data from Table 1 and build a corresponding concept lattice, Fig 1. In Table 2 the significant entries, i.e. those with values not smaller than the column's θ, are denoted by '$\star$'. We pay particular intention on the concepts that contain failure in integration (If).

At the beginning of a semester students take a test with mathematical problems. The test results show a student's tendency to fail, possession of good knowledge or very good knowledge related to operations with fractions, logarithmic functions and trigonometrical functions. Our further work aims to establish whether a student belongs to a concept containing the attribute failure in integration. If affirmative the student is strongly advised to attend additional mathematical classes. The suggested approach is not limited to concepts containing the attribute failure in integration only. On the contrary, it can show to which concept any particular student belongs to.

Below we concentrate on concepts that have the attribute failure in integration.

Table 1: Context for students groups

	Fractions			Logarithms			Trigonometrical functions			Integration		
	fail	good	very good	fail	good	very good	fail	good	very good	fail	good	very good
	Ff	Fg	Fvg	Lf	Lg	Lvg	Tf	Tg	Tvg	If	Ig	Ivg
G1			$\star$			$\star$			$\star$			$\star$
G2			$\star$		$\star$			$\star$				$\star$
G3		$\star$			$\star$				$\star$		$\star$	
G4	$\star$				$\star$			$\star$			$\star$	
G5		$\star$		$\star$				$\star$		$\star$		
G6		$\star$			$\star$		$\star$			$\star$		

3.1 Single Concept

In this case student's results belong to one concept having If as an attribute or to the set of concepts having If as an attribute.

Suppose there is only one concept having If as an attribute. A student belongs to that concept if the attributes' values, related to his/her test result, belong to the corresponding attributes' intervals values of the concept (all attributes' values except If because the student has not been tested on integration).

In case there are several concepts having If as an attribute we apply the same rule like the one with a single concept but we work with attributes' intervals values of the set of all concepts having If as an attribute (again all attributes' values except If because the student has not been tested on integration).

Example 0.1. The test results of a new student ($St1$, Table 2) belong to the intervals determined by the concept

$$I = \{Fg, If\}, \;\; E = \{G5, G6\}.$$

Therefore, the If value of $ST1$ belongs to the interval $[0.7, 0.8]$ where $\theta(If) = 0.4$. The student is therefore advised to take additional classes since $\theta(If) = 0.4$ is smaller than any value in the interval $[0.7, 0.8]$.

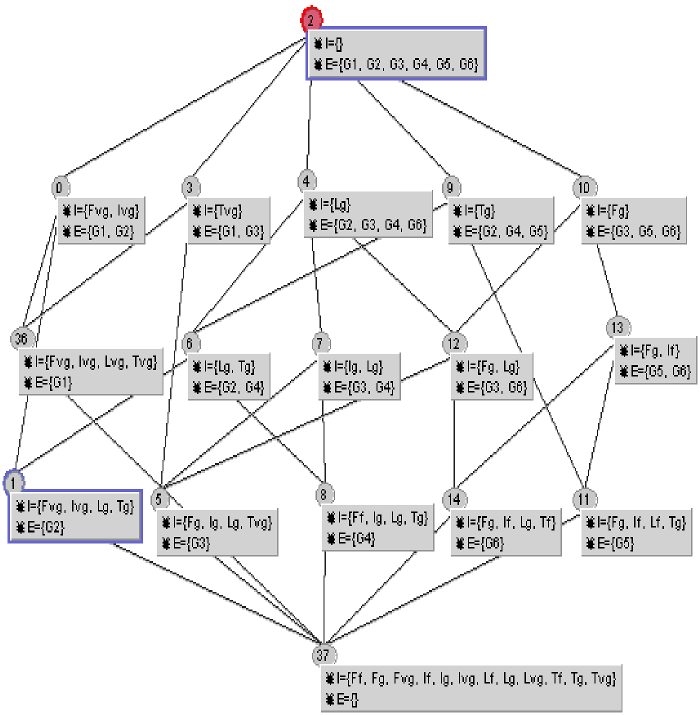


Fig. 1: Concept lattice

Table 2: Context for students groups with numerical values

	Fractions			Logarithms			Trigonometrical functions			Integration		
	fail Ff	good Fg	very good Fvg	fail Lf	good Lg	very good Lvg	fail Tf	good Tg	very good Tvg	fail If	good Ig	very good Ivg
G1	0	0	1⋆	0.3	0.1	0.6⋆	0.1	0.1	0.8⋆	0.1	0.2	0.7⋆
G2	0	0.1	0.9⋆	0.1	0.7⋆	0.2	0.1	0.7⋆	0.2	0.2	0.3	0.5⋆
G3	0.2	0.7⋆	0.1	0.2	0.6⋆	0.2	0.1	0.4	0.5⋆	0.3	0.5⋆	0.2
G4	0.7⋆	0.2	0.1	0.3	0.6⋆	0.1	0.3	0.5⋆	0.2	0.4	0.5⋆	0.1
G5	0.2	0.6⋆	0.2	0.8⋆	0.1	0.1	0.4	0.5⋆	0.1	0.7⋆	0.3	0.1
G6	0.3	0.6⋆	0.1	0.3	0.6⋆	0.1	0.8⋆	0.2	0	0.8⋆	0.2	0
θ	0.23	0.38	0.4	0.33	0.45	0.22	0.3	0.4	0.3	0.41	0.33	0.26
St 1	0.3	0.5	0.2	0.5	0.5	0	0.7	0.3	0			
St 2	0.2	0.7	0.1	0.6	0.2	0.2	0.1	0.8	0.1	0.33		

3.2 Several Concepts

Suppose test results of a student belong to more than one concept where at least one concept does not have If as an attribute. The value of the attribute $If(s)$ for that student is calculated according to the function

$$If(s) = \frac{\Sigma If(C_i) \cdot \alpha_{C_i}}{l}$$

where C_i are the involved concepts, α_{C_i} are the corresponding affiliation values and l is the number of the involved concepts.

If the student's results are not equal to any of the known attributes' we take the concept that has the closet attribute value. In case there are several concepts with attribute values equal to the one we are interested in, we consider all of these concepts.

Example 0.2. We apply the suggested approach to the tests' results of another new student $(St2)$. The test results of the student $St2$, (Table 2) belong to the concepts with objects $G2, G3, G5$, and $G6$. The affiliation values for concepts with objects $G2, G3, G5$, and $G6$ are

$$\alpha_{G2} = \frac{0.9+0.7+0.7+0.5}{4} = 0.7$$

$$\alpha_{G3} = \frac{0.7+0.6+0.5+0.5}{4} = 0.58$$

$$\alpha_{G5} = \frac{0.6+0.8+0.5+0.7}{4} = 0.65$$

$$\alpha_{G6} = \frac{0.6+0.6+0.8+0.7}{4} = 0.7$$

The sum-of-1-criterion is applied while values of membership functions in Table 2 are determined. The order of the corresponding $If(St2)$ value is then calculated

$$\begin{aligned} If(St2) &= \frac{If(G2)\alpha_{G2}+If(G3)\alpha_{G3}+If(G5)\alpha_{G5}+If(G6)\alpha_{G6}}{4} \\ &= \frac{0.2\cdot 0.7+0.3\cdot 0.58+0.7\cdot 0.65+0.8\cdot 0.7}{4} \\ &= 0.33 \end{aligned}$$

Since $0.33 < 0.41 = \theta(If)$, the second student is not advised to take additional classes.

4 System Description

A system prototype is build as a Web-based application using Apache HTTP server [15], mod_python module [16] and SQLite database [17]. The mod_python module provides programmable runtime support to the HTTP server using Python programming language. The whole application components are

1. Web-based users interface,
2. application logic, and
3. database interaction were written in Python.

Initially the system database was created to be able to store data with the structure as shown in Table 1 and Table 2. Subsidiary data tables for users' and courses' information were also needed.

The users, i.e. expert tutors, teachers, and students interact with the system using Web forms. Before any interaction with the system can take place, a user needs to be authenticated first. Experts and teachers can submit and update data, while students can only view information.

For a particular subject, an expert tutor will first submit data that will be used to construct data table as shown in Table 2. Taking information as shown in Table 2 as an example, the expert will provide the system with context data as in Fig. 3.

The system will then check that there are no duplicate attribute combinations and insert the context data in to the database. A corresponding concept lattice is then built with a proper objects' and attributes' labelling. All relationships among extents and intents are represented in an internal abstract data structure. The concept lattice structure is then saved in the database in a format that can be loaded into a program without any additional recalculations needed.

Next, the expert will provide the system with numerical values that correspond to the context data as the example shown in Table 2. The significant (maximum) values took the '1' positions of the context, while the lesser two values took the '0' positions, (see Fig. 3). During the process of populating the data table, the system checks that these conditions and the sum-of-1-criterion hold for each group data. Values for each of the attributes are calculated and saved at their proper places.

The system is now initialized and ready to be used by teaches and students in providing recommendations on whether or not a student needs to take additional classes (courses).

A teacher can then consult the system on the amount of knowledge of a particular student by submitting the student's pre-test results as illustrated in Example 1 and Example 2. In Example 1, the teacher submits the student's result data in the following text-string format:

$$St1 : 0.4, 0.5, 0.1\#0.5, 0.5, 0\#0.7, 0.3, 0$$

The data string is parsed and checked for the sum-of-1-criterion for each of the three blocks and stored into the database. The data string is also translated into the context format where within a block the maximum value gets the 1 value and last two get 0 values. If there are ties in the maximum values then each one will contribute to different translations. For the result data of the student $St1$, the translated result data are shown below, where the last three values 100 are added since the translations concern the case for If:

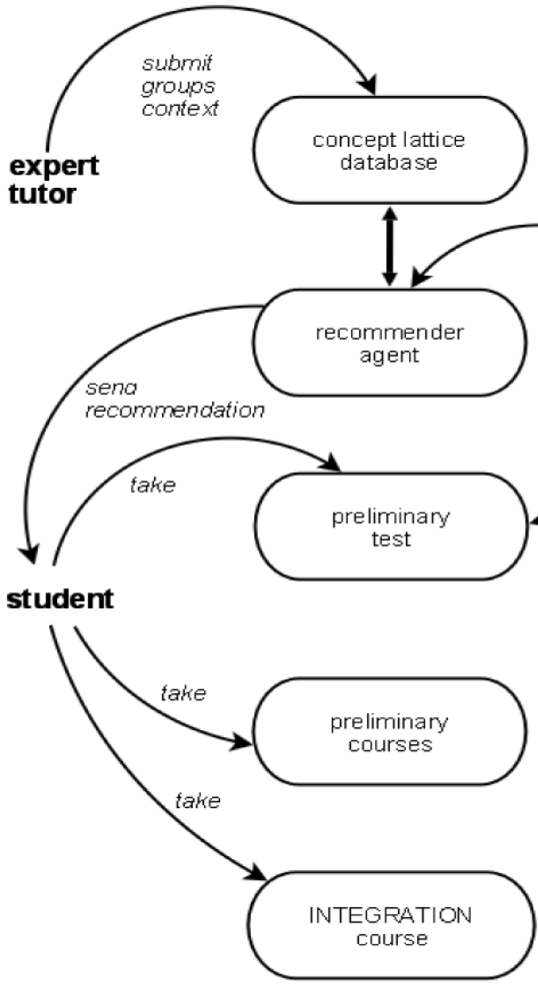


Fig. 2: System conceptual diagram

$G1$: 001001001001
$G2$: 001010010001
$G3$: 010010001010
$G4$: 100010010010
$G5$: 010100010100
$G6$: 010010100100

Fig. 3: Binary context data

$$St1 : 010100100100$$
$$St1 : 010010100100$$

The first three numbers 010 are a code for Fg. In the concept lattice shown in Fig. 1, there is a node representing the concept $I = \{Fg, If\}, E = \{G5, G6\}$, and it so happens that $St1$ has translations equivalent to $G5$ and $G6$. The $If(St1)$ for the student $St1$ is within the interval that is greater then (If), thus the student will be recommended by the system to take additional courses.

The submitted results of the student $St2$:

$$St2 : 0.2, 0.6, 0.2\#0.6, 0.2, 0.2\#0.1, 0.8, 0.1$$

will be translated by the system into:

$$St2 : 010100010100$$

Within this translation pattern:

1. the first 010 corresponds to Fgwhich relates $St2$ to $G3, G5, G6$.
2. the second 100 corresponds to Lf which relates $St2$to $G5$.
3. the third 010 corresponds to Tg which relates $St2$ to $G2$ and $G3$.

With affiliation values for $G2, G3, G5, G6$ the corresponding $If(St2)$ value was calculated to be 0.33, which is lower then (If), thus the student will not be advised to take additional courses.

Each time the calculation of $If(< student - identification >)$ value is done, the system will send an email to that particular student concerning recommendations and information about additional courses.

5 Conclusion

The suggested approach turns out to be quite useful for providing timely recommendations to students who could have serious problems studying a particular subject due to lack of sufficient preliminary knowledge. Even though the approach has been discussed in relation to studying calculus only, we believe that it can be applied to other subjects that require preliminary knowledge and/or skills.

References

1. Bellman, R.E., Zadeh, L.A.: Decision making in a fuzzy environment. Management Sciences, Series B 17, 141–164 (1970)
2. Carlsson, C., Fuller, R.: Optimization under the fuzzy if-then rules. Fuzzy Sets and Systems (1998)
3. Davey, B.A., Priestley, H.A.: Introduction to lattices and order. Cambridge University Press, Cambridge (2005)

4. Felix, R.: Relationships between goals in multiple attribute decision making. Fuzzy sets and Systems 67, 47–52 (1994)
5. Fuller, R., Zimmermann, H.-J.: Fuzzy reasoning for solving fuzzy mathematical programming problems. Fuzzy Sets and Systems 60, 121–133 (1993)
6. Herrmann, C.S.: Fuzzy logic as inferencing techniques in hybrid AI-Systems. In: Martin, T., L. Ralescu, A. (eds.) IJCAI-WS 1995. LNCS, vol. 1188, pp. 69–80. Springer, Heidelberg (1997)
7. Herrmann, C.S., Holldobler, S., Strohmaier, A.: Fuzzy conceptual nowledge processing. In: 1996 ACM symposium on applied computing, pp. 628–632 (1996)
8. Mamdani, E.H.: Applications of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Transactions on Computers 26(12), 1182–1191 (1977)
9. Sugeno, M.: Fuzzy measures and fuzzy integrals: a survey. In: Gupta, M.M., Saridis, G.N., Gaines, B.R. (eds.) Fuzzy Automata and Decision Processes, pp. 89–102. North-Holland, NY (1977)
10. Sugeno, M.: Industrial applications of fuzzy control. Elsevier Science Pub. Co., Amsterdam (1985)
11. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man and Cybernetics, 116–132 (1985)
12. Wille, R.: Concept lattices and conceptual knowledge systems. Computers and Mathematics with Applications 23(6-9), 493–515 (1992)
13. Zadeh, L.A.: The concept of linguistic variable and its applications to approximate reasoning, Parts I, II, III. Information Sciences, 8, 199–251; 8, 301–357; 9, 43–80 (1975)
14. http://www.engsc.ac.uk/journal/index.php/ee/article/view/62/97
15. Apache HTTP Server Project, http://httpd.apache.org/
16. Python Programming Language, http://www.python.org/
17. SQLite, http://www.sqlite.org/

A Study of the Effect of Pupils' Learning Achievement on Project-based Learning with Context Awareness Technology

Kuo-Kuang Chu, Chien-I Lee, and Po-Cheng Lu

Abstract This study would like to find out the pupils' learning achievement on the context awareness PBL (Project-Based Learning) activity to identify the different herb plants in the campus. We implemented a context awareness mobile learning environment with RFID and Wi-Fi technologies. The result of the study showed that the five dimensions scores of experiment group are higher than the controlled group, because the experiment group can access learning materials on the web site immediately, and search more relevant information about the herbs on the internet, the works of the experiment group emerge more creative ideas and interesting learning. The pupils' PBL artifacts showed that there are no significant differences in gender and performance factors. The low achievement students could get the assistance from the context awareness mechanisms, and after the members of the team discussed their opinions, they also get a great performance in identifying the herbs and their artifacts.

1 Introduction

Traditional teacher-center pedagogy is a useful teaching method for numerous students in a school that they can acquire a lot of knowledge in a short term, but students usually do not have enough time to comprehend and absorb all knowledge. In recent years the constructivism has became the education reform theme, social constructivist scholars think that every learner is a unique individual, learning process must be based on learner-center, and the role of the instructors should be changed as the facilitators who assist the learner's driving problem, active discovering, understanding and applying knowledge to solve their problem, finally to construct knowledge themselves. Project-Based Learning (PBL) is starting at authentic inquiring

Department of Information and Learning Technology,
National University of Tainan, Taiwan
e-mail: chukk@tn.edu.tw, leeci@mail.nutn.edu.tw,
lbu@mail2000.com.tw

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 211–221.
springerlink.com

problems, and then the learners can use any resources from web or library, through the team work to achieve self-direct learning.

Mobile learning (m-learning) is integrated e-learning and distance education that the learners can learn across contexts anywhere, anytime with any portable devices. Several studies (Sharples, 2000; Liu 2003; Laroussi, 2004; Liu, 2008; Zurita, 2008) have focused on mobile learning and PBL with mobile devices. However the m-learning realizes that the learner can access context with mobile devices via wireless network, but these technologies cannot get the learner's exact location, without these learner's status information, the agent cannot make an intelligent decision to provide help to learners.

To know the learner's location, we used active RFID technology, and we established the Wi-Fi network to provide the learner's mobile device accessing resources on the internet that it realizes the seamless learning when the learner changes different devices. Furthermore, it also enables communicating between instructors and peers that it made the learners able to cooperative learning. The context awareness service can detect and collect the surrounding data from the learner, according to this information and the learner's profile in the database, the agent will decide what the useful information for the learner is.

In this study, we implemented a context awareness PBL system to provide the learning content for the users, and then we would like to find out what is the user's learning effects and attitudes on context awareness PBL, and if the learner's location, profile, portfolio and surrounding factors will influence the learning or not. Furthermore, we would analyze if the gender and achievement have significant differences to influence the learning effects, and we would collect the teacher and students opinions to improve future research after this experiment.

2 Context Awareness Project-Based Learning

2.1 Project-Based Learning

Marx (1994) defined the PBL has "driving question' encompassing worthwhile content that is anchored in a real-world problem; investigations and artifacts that allow students to learn concepts, apply information, and represent knowledge in a variety of ways; collaboration among students, teachers, and others in the community so that participants can learn from one another; and use of cognitive tools that help learners represent ideas by using technology...". Thomas (2000) explains that PBL require "complex tasks, based on challenging questions or problems, that involve students in design, problem-solving, decision making, or investigative activities; give students the opportunity to work relatively autonomously over extended periods of time; and culminate in realistic products or presentations." The Buck Institute for Education (2008) defines PBL as "a systematic teaching method that engages students in learning knowledge and skills through an extended inquiry process structured around complex, authentic questions and carefully designed products and tasks."

Blumenfeld (1991) states that PBL activity should include asking and refining questions, debating ideas, making predictions, designing plans and/or experiments, collecting and analyzing data, drawing conclusions, communicating ideas and findings to others, asking new questions and creating artifacts.To put it briefly, PBL is starting an inquiry project, the learners can use the internet and library resources to achieve self-direct learning, through team-work cooperative learning, driving question, identifying concept, searching information, inquiring, modifying the problem and project, designing and finishing the experiment, analyzing the data and completing their artifacts.

Many evidence shows that PBL enhances the quality of learning and leads to higher-level cognitive development through students' engagement with complex, novel problems, the learners not only respond by feeding back information, but they also actively use what they know to explore, negotiate, interpret, and create. They construct solutions, thus shifting the emphasis toward the process of learning (Buck Institute for Education, 2009).

2.2 Ubiquitous Learning

Ogata (2004) defines ubiquitous learning "has integrated high mobility with pervasive learning environments. While the learner is moving with his/her mobile device, the system dynamically supports his/her learning by communicating with embedded computers in the environment. As for the broad definition of ubiquitous learning, both pervasive learning and mobile learning would be in the category of ubiquitous learning. RFID tags are often used to realize pervasive computing."

2.3 Context Awareness

Dey and Abowd (1999) define the Context is "any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." Guanling & David(2000) figures out the context awareness should include computing, user, time and physical context.
There are some context awareness learning applications, but using context awareness technology to observe the herb plants with PBL has been little research. This study is focus on the outdoor herb plants instruction, engaging context awareness service to provide a ubiquitous learning environment for students.

2.4 Learning Effect

Gagné, Briggs & Wagger (1992) introduce the categories of learning process, such as motor skills, verbal information, intellectual skills, cognitive strategies and attitudes. This research would use attitudes, the score of the learner's artifacts and opinions after cooperative teamwork to evaluate the students' learning effect.

3 Method

3.1 Framework

This study's research framework is showed on Table 1, the detail explains as below:

Table 1: Research Framework.

Control Variable	Independent Variable (Instruction model)	Dependent Variable (Learning effect)
Teacher Learner characteristics Grouping Teaching hours PBL stages Initial behavior	Experiment Group (Context awareness PBL) Controlled group (PBL) Attribute Variable Learning Achievement Gender	Artifacts score Cooperative Teamwork Learning attitude

Independent variable: Instruction model
Experiment group: The learners hold PDA via Wi-Fi network to login into the context awareness PBL system that would get the user's profile, location and accessing time, and provide appropriate learning materials to the user. On the other hand, the user could go into the chat room to discuss with the teacher and the peers.
Controlled group: The learners could only use desktop computer via LAN to login into same PBL system as the experiment group.
Control variable:

1. Learner characteristics: there are no significant differences in science and living technology between the experiment group and the controlled group.
2. Grouping: the experiment group and the controlled group are grouping by heterogeneous for cooperative learning.
3. Teaching hours: the experiment group and the controlled group have same teaching hours.

4. PBL stages: these two groups have the same PBL training practice, same stages, same teaching content and schedule.
5. Initial behavior: there are no significant differences on learner's performance between the experiment group and the controlled group.

Dependent variable: Learning effect
The artifact score: after the PBL activity, the artifact will be evaluated by the PBL artifact scaling and five expert's assessment.

3.2 Participants

This context awareness PBL activity was implemented on an urban elementary school in southern Taiwan, there are two classes that are fifth grade students under normal distribution, and every class has 30 students, so total has 60 students. One class that has 14 male and 16 female is in an experiment group, and the other class that has 15 male and 15 female is in a controlled group, all students of the class was divided into 10 groups, so each group has 3 persons. All students have one lesson of computer science per week for 2 years, so they have basic ICT literacy, including browsing web pages, searching data on internet, word processing, but they do not have yet the PBL experience. In order to verify the student's achievement differences in these two classes, we randomly picked up the student's last semester score of science and living technology to perform the T-Test, the result shows $t = 0.146$, $p = .851 > .05$ as Table 2, it means that there are no significant differences on these two classes.

Table 2: the T-Test for students' score.

Subject	Group	Numbers of student	mean	standard deviation	variation coefficient	t-Test
Science and living technology	experiment	30	87.4	9.18	1.68	0.146
	controlled	30	87.2	10.27	1.87	

3.3 Design

PBL activity contains follow stages:

1. Preparing the activity: the teacher introduces the PBL concept, system and PDA operation.
2. Indentifying the concept: the teacher causes motivation, and organize the concepts for students.
3. Searching for the information: students search relevant information.
4. Inquiring and modifying the problem: the students discuss and inquire to decide the issue.
5. Planning and designing the experiment: every group presents their proposal.
6. Experimenting: the students observe the herbs, note the records, discuss with other peers.
7. Analyzing data: the students integrate different resources to make tables or figures for the artifact.
8. Presenting the artifact: every group presents their artifact, and evaluates the PBL effects by themselves, peers and experts.

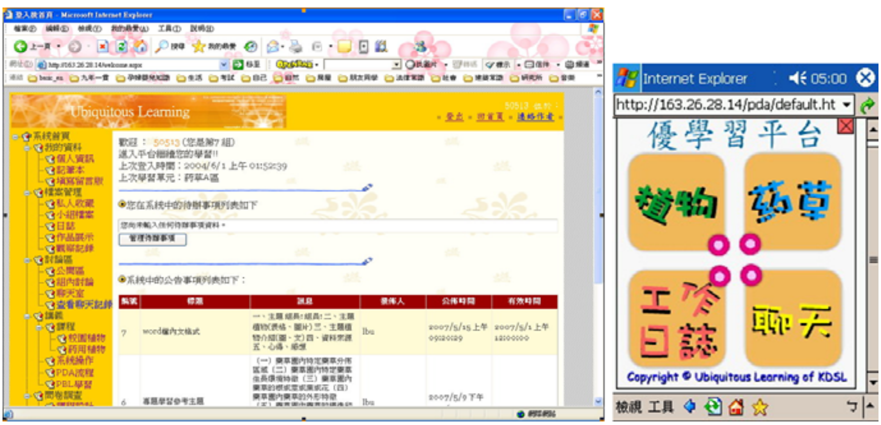


Fig. 1: (a) the context awareness PBL system (PC) (b) the menu of system (PDA).

The context awareness PBL system has a desktop personal computer (PC) and a PDA version, the snapshot shows on Figure 1, it provides experiment and controlled group to use, there are introductions every herb plants as Figure 2(a), and the chat room support learners to discuss as Figure 2(b), it also provides the observation record area to note the appearance of the herb as Figure 2(c).

3.4 Data Analysis

The research covers several statistical results that is maintained by SPSS software, the means, standard deviations are obtained and t-Test is applied to finalize the hypotheses test.

Fig. 2: (a) the herb introduction, (b) the chat room snapshot, (c) note the observation.

4 Results

We take the PBL average score of the experiment group and the controlled group to depict the result as Figure 3.

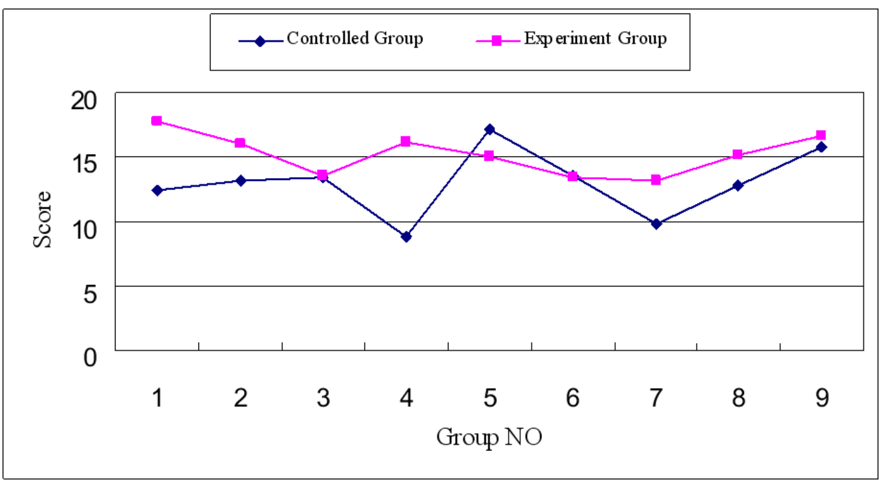


Fig. 3: The score of experiment and controlled group.

This study uses the cooperative learning assessment scale to measure the learner's cooperative learning effect. The scale contains five dimensions:

1. Organization: every concept is consistent and is continuous between method to method.

2. Validity: all content is correct.
3. Research: student can add personal idea and use different data source to improve the quality of the project.
4. Creativity: the artifact has creative idea, and it makes the project become be better.
5. Presentation: the artifact uses multimedia to present, it is more interesting.

Table 3: the t-Test for the cooperative learning assessment scale.

Dimensions	Experiment Group	Controlled Group	p
Organization	3.07	2.82	0.189>0.05
Validity	3.29	3.20	0.384>0.05
Research	3.11	2.47	0.00<0.05 *
Creativity	2.76	2.04	0.00<0.05 *
Presentation	3.00	2.47	0.0001<0.05 *

The t-Test result is showed on Table 3, there are no significant differences between the experiment group and controlled group on organization and validity; but there are significant differences on the research, creativity and presentation. The students' score of the experiment group and the controlled group are 15.22 and 13.00, $p = 0.0001 < 0.05$, it means that the score on the experiment group and the controlled group have no significant differences. If we take a look on the t-Test of gender of the experiment group, the male's score is 15.040, the female's score is 15.212, $p = 0.747 > 0.05$, though the female's score is higher than males, but it still has no significant differences. On the other hand, the male and the female's score on the controlled group are 13.500, 13.025, $p = 0.593 > 0.05$, it also has no significant differences between the males and the females.
We put the personal score of the project and achievement to perform the Pearson's product moment correlation coefficient analysis and t-Test, $r = 0.335$, $p = 0.071 > 0.05$, it means that the personal project score exists positive correlation with the personal achievement score, but they still have no significant differences. On the other hand, the Pearson's product moment correlation coefficient on the controlled group is 0.550 and $p = 0.001 < 0.05*$, it has positive correlation and also has a significant differences between the personal project score and the achievement score.

In order to make sure that the scores from the five experts evaluated are reliable, we performed the Chi-Square Test as shown in Table 4, the result shows that the threshold is 9.49, the five experts evaluated scores have reliability to 18 artifacts

Table 4: the experts' reliability test

Kendall's Coefficient of Concordance	0.241
Chi-Square	86.60
df	4

5 Discussion

A partial explanation for the experiment group is better than the controlled group in research, creativity and presentation that the context awareness PBL system supports the experiment group engages active learning, they could open their mind to find the better idea easier than the controlled group. The gender would not influence the context awareness PBL activity, no matter weather the experiment or the controlled group has the same result. As the correlation coefficient shows that the pupil's project is positive correlation to his achievement, but they still are no significant differences, we interpret this to mean that the low achievement pupils have enough progress after the cooperative learning teamwork process, because they could get more context awareness service and peer's help.

Most pupils are interested using PDA to observe the herb plants. Meanwhile, they could note the records and discuss the result with others in the chat room. This activity facilitates the classmate friendship, and helps them to realize how to discuss and share with others.

6 Conclusion

This study shows that the pupils could easy carry out PBL in context awareness environment with PDA, and they discussed the project in herb plant garden with their team members very well. The context awareness PBL activity enables to promote the quality of observing and intercommunicating. Because the pupils' PBL final artifact should be presented, though they had been practiced by themselves, but they had some troubles in the formal presentation, because they still did not use this assessment approach.

The context awareness service provides the appropriate information according to the user's location, profile and time, not only can the pupils compare and validate the real world plants with collecting relevant data, but also it increases the validity of identifying herb plants and decreases the user's loading using mobile devices. The pupils recorded their detailed authentic observation by PDA, it enriches the observation records, and in the advance designing treasure hunting activity promotes the pupils' learning motivation.

Most pupils enjoy the context awareness PBL environment using PDA, RFID tags, and Wi-Fi wireless network; these technologies promote the pupils' learning motivation and facilitate problem solving, active learning ability, resources sharing and cooperative learning, and the low achievement students also got the assistance from the context awareness mechanism and their peer's help. But the wireless network stability and hand writing recognize accuracy would inference student's learning deeply.

This study finds out that the student's gender and learning performance have not inference to the PBL learning effects. The context awareness mechanisms promote pupils' identification ability of the herb plants, raising the learning motivation to active learning and thinking, meanwhile the students' ICT literacy, problem solving, and social communication ability have also quite progressed.

References

Abowd, 1999. Abowd, G.D., et al.: Towards a Better Understanding of Context and Context awareness. In: Gellersen, H.-W. (ed.) HUC 1999. LNCS, vol. 1707, pp. 304–307. Springer, Heidelberg (1999)

Blumenfeld, 1991. Blumenfeld, P.C., et al.: Motivating project-based learning: Sustaining the doing, supporting the learning. Educational Psychologist 26(3-4), 369–398 (1991)

Buck Institute for Education, 2009. Buck Institute for Education, Handbook: Introduction to Project Based Learning (2009), http://www.bie.org/index.php/site/PBL/pbl_handbook_introduction/#benefits

Buck Institute for Education, 2008. Buck Institute for Education, Project-based learning (2008), http://www.bie.org/index.php/site/PBL/pbl_handbook_introduction/#history

(Gagné, 1992). Gagné, R.M., et al.: Principle of Instructional Design. Harcourt Brace Jovanovich Publishers, Orlando (1992)

Guanling & David, 2000. Guanling, David: A survey of context-aware mobile computing research (Dartmouth Computer Science Technical Report TR2000-381) (2000)

Laroussi, 2004. Laroussi, M.: New e-learning services based on mobile and ubiquitous computing UBI-Learn Project (2004)

Liu, 2003. Liu, T.C., et al.: Wireless and mobile technologies to enhance teaching and learning. Journal of Computer Assisted Learning 19(3), 371–382 (2003)

Marx, 1994. Marx, R.W., et al.: Enacting project-based science: Experiences of four middle grade teachers. Elementary School Journal 94(5), 518 (1994)

Ogata, 2004. Ogata, H., Yano, Y.: Context-aware support for computer-supported ubiquitous learning. In: Proceedings of Wireless and Mobile Technologies in Education (2004)

Ravitz, 2004. Ravitz, J., et al.: Online professional development for project based learning: Pathways to systematic improvement. In: Association for Educational Communications and Technology Annual Meeting, Chicago (2004)

Sharples, 2000. Sharples, M.: The design of personal mobile technologies for lifelong learning. Computers & Education 34(3-4), 177–193 (2000)

Thomas, 2000. Thomas, J.: A Review of the Research on Project-Based Learning. The Autodesk Foundation (2000)

Liu, 2008. Liu, Y.H., et al.: A case study of online project-based mobile learning: Beer King. In: Proceedings of the 8th WSEAS International Conference on Multimedia systems and signal processing, Hangzhou, China. World Scientific and Engineering Academy and Society (WSEAS) (2008)

Zurita, 2008. Zurita, G., Baloian, N., et al.: Using Mobile Devices to Foster Social Interactions in the Classroom. In: 12th International Conference on Computer Supported Cooperative Work in Design. CSCWD 2008, Xián (2008)

Improving interactions in knowledge intensive communities of practices for SMEs

Ileana Hamburg

Abstract Knowledge is a key to success also for small and medium sized companies (SMEs) but many of them lack of understanding what they want to accomplish with a knowledge management (KM) environment and are not ready to implement/use it. In this paper we first present the readiness of SMEs for KM, then we discuss if Communities of Practice (CoPs) are a suitable environment for KM and learning within SMEs. Finally we offer an example for development of knowledge intensive CoPs with efficient interactions both between the actors and with the corresponding supporting IT platforms.

1 Introduction

It is known that a priority theme on many national government agendas and, in general, at the European level is the improvement of the competitiveness of business and industry. Knowledge is a key to all organisations and the success of many of them depends on the effective deployment and continual enhancement of their knowledge base to be competitive. The problem is getting more acute, particularly in the next period, when great challenges loom large for the world economy. Small and medium sized companies (SMEs) have particular needs in facing the challenges of their daily operations [3]. Many European SMEs are not ready for significant international social and economic changes [8]. Some of them have focused on knowledge management (KM) and used it as an enabler for innovation capability, but many of the practiced KM approaches failed. For a SME to manage and sustain business whilst engaging in KM and training for it can be very difficult. Their priority is survival, leading to just-in-time activities; the benefits of KM and learning to the business have to be very clear and measurable [1, 2]. To be effective and acceptable to staff, learning and knowledge management have to be directly related to activities on the job [9, 10]. These requirements can be met best by an intelligent use

Institut Arbeit und Technik, FH Gelsenkirchen, Germany

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 223–230.
springerlink.com

of IT based learning and KM, particularly Web 2.0 methods and applications [17, 18]. They are flexible, and they support the combination of formal and informal learning, collaboration and individualized work. Specifically, they enable a mode of work and interaction beneficial for SMEs and SME networks: the Community of Practice (CoP) [22] with participation of SMEs and other “practitioners”. Say [20] described how researchers try to develop now so-called “third generation” of KM systems which will be focused on enabling the sharing of tacit knowledge. Knowledge intensive and strong interactive CoPs fulfill such requirements. Results of studies, projects and discussions with SME experts and representatives show that one of the most critical but important aspect to be considered when developing/improving KM environments in companies is an evaluation of KM readiness of them. Many companies lack of understanding what they want to accomplish with a KM environment. They do not know if the organization, the management, the culture, the staff and technical infrastructures are “ready” for this or not and how to take advantages of new IT to improve interactions in such KM environments. In this paper we first present shortly a method to determine the readiness of SMEs for KM, then we discuss if CoPs are a suitable environment for KM and learning within SMEs. Finally we offer an example for development of knowledge intensive CoPs with efficient interactions both among the actors and with the corresponding supporting IT platforms.

2 Readiness of SMEs to develop KM environments

A wide variety of approaches have been proposed in conducting assessments of KM readiness but there is a lack of systematic approach and the practice varies with different industries and companies. In this part a systematic KM readiness model is presented. Some proposed categories for the model are Organization, Strategy, Human resources/Users, Culture, IT used. Experience shows that a suitable assessment of KM readiness in SMEs (having limited resources) can be realized in form of a simple questionnaire survey for managers and individual employees to fill in. The answers will be analyzed and strategies for KM should be planned. Some main and derived questions can be the followings:

Organisation/Management: Is your enterprise organized to use efficiently its knowledge, to acquire and create new one?

- Is the adequate understanding that KM means much more that implementing an IT tool or solution?
- There is a well structured knowledge base of the enterprise that can be used by the staff when they need?
- Are informal and/or knowledge intensive networks like CoPs across different units of the enterprise encouraged?
- Are informal and/or knowledge intensive networks like CoPs across the enterprise encouraged?

- Are partners identified for developing of knowledge intensive networks like CoPs?
- There is knowledge to be used across different units available in standable formats?
- Are the interaction facilities with the knowledge base and other knowledge sources user friendly and reliable?
- Is the dissemination of best practice supported in your enterprise?
- Is the enterprise connected with external networks and knowledge sources?

Strategy: Has your enterprise a sustainable program for improvement of KM?

- There is a track record in the enterprise in successful implementation of innovative approaches?
- Has your enterprise a vision how KM can support the company business and staff work tasks?
- There is a strategy for doing this in the next 3 years?
- Are the tasks and responsibilities for this task clear?
- Are there some measures for improving KM in the next 3 years?
- There is a strategy to protect the key information and shared knowledge in the enterprise?
- There exist a complete IT security procedure for information?
- Are key performance indicators for KM in work place?
- There is a regular measurement of the impact KM has on the ways the staff do business and work in the enterprise?

Human resources/Users: Do the enterprise staff understand the concept of KM and commit the seniors managers to support it?

- Do the staff understand the term KM and how to use existing knowledge for their business and work?
- There is a board responsible with KM concepts?
- Do senior managers support knowledge sharing (i.a. during meetings)?
- Are the interactions among staff within and across company units favorable for knowledge sharing?
- There is a senior level systematical review of the effectiveness the staff use KM?
- Are new ideas of the staff encouraged for creating new knowledge?

Culture: Do the behaviors and interactions in the company enable effective KM?

- Is the recording and sharing of knowledge a routine in the company?
- Time is allowed for creative thinking and reflections?
- Are best practices systematically selected and the corresponding knowledge used whenever staff need it?
- Is everyone willing to give advice and to help on request to anyone else in the company?
- Are individuals rewarded for team work and knowledge sharing?
- There is a strong belief that the work can do best only by sharing ideas?

IT: Is there a suitable IT support for KM in the enterprise and is this used efficiently to support KM?

- Is the availability of technology suitable for knowledge sharing and using a main point when new IT acquisitions are discussed?
- Does the company IT unit/team check constantly if the existing IT platform for KM supports knowledge needs of the staff?
- Does the existing IT support effective communication across boundary and even time zones?
- Do the interactions of staff with the IT at work place constitute a normal working practice?

3 Knowledge intensive Communities of Practice

CoPs are groups of people working together at solving openended questions, learning in social and physical contexts of realworld problems and using collaboration and cognitive tools for KM and learning. Some main characteristics of CoPs are the following:

- a shared domain of interest of its members, their commitment to this domain and a shared competence,
- common ideas, joint activities. Members engage in pursuing their interest for the domain and build relationships that enable them to learn from each other,
- common practice because members of a community are practitioners with different levels of expertise. They develop a shared repertoire of resources e.g. experiences, tools, ways to solve problems, a knowledge base of best practices.

CoPs offer new opportunities for KM and learning processes by using new forms of interaction within the teamwork and loose contact between the actors. CoPs show differences to theme-specific cooperation and/or temporary networks. They are lasting for a longer period. Its members are ready to share knowledge and to create new one together and to deal with strategic fields of knowledge in business. Nonaka and Konno [16] have described a knowledge development cycle showing how tacit or implicit knowledge can be made explicit in learning processes. This work and others pointed out, that knowledge developed in CoPs is important for understanding how knowledge develops in different contexts. These distinctions are important when processes of learning and knowledge development in SMEs are analyzed. Important is also the design of CoPs as KM environments and of interactions among members of the CoP and with supporting IT. Some principles of "designing for aliveness", which can guide organizations wishing to start a CoP are explained here. These we have followed in our current project SIMPEL (see below):

- Design for evolution e.g. design elements should be combined in a way acting as catalyst for a natural evolution to a knowledge intensive and life-long learning oriented CoP,

- Keep an open dialog between inside and outside perspectives of the CoP because the last one can help community members to see new possibilities to act effectively and to use them,
- Consider different levels of participation for the members of the CoP (leadership roles, core active group, rare participants, etc.) and different kinds of interactions,
- Develop public and private community spaces,
- Create a rhythm and rules of interacting within the community.

Internet technologies [6] extend the interactions within communities of practice beyond geographical limitations and make possible the building of virtual CoPs (VCoP). These communities free their members from constraints of time and space. In comparison with technical solutions for knowledge management, VCoPs can mark a change from "managing knowledge" to "enabling knowledge".
"Effective knowledge creation depends on an enabling context. What we mean by enabling context is a shared space that fosters emerging relationships" [14]. In order to assure an optimal interaction between users and the IT platforms supporting VCoPs with SME participation, methodologies and processes should be used for the interfaces taking into consideration the IT competences and learning abilities of the learning staff. Interfaces should have a basic level of usability ("really just means making sure that something works well [13]. If it is possible the interfaces should be tested with staff from different SMEs but also with some experts. Heuristic Evaluation can be done by using Nielsens ten Heuristics. The current generation Web 2.0 [15, 19], has a vast potential to create prospering environments for emerging CoPs. It very well can support activities within a community and for staff of SMEs to collaborate as well as the idea of connectivism developed by Siemens [21] where information is constantly changing, learning takes place in distributed networks of people and is based on diversity of opinions; content and services are adaptable and responsive for example to specific needs and goals of SMEs. For example: writing in public blogs encourages the writer to think about the issues in question. In communities, an individual will receive help from a network of peers, so unnecessary searching activity and time can be saved. Castro [5] underlines how the virtual environment helps the feedback mechanism by reducing costs of communication and of storing and effectively retrieving informal feedback. Referring tacit knowledge, interviewed CoP actors from Germany show that one of the useful ways for their work is by asking for and receiving opinions about their written work or about made decisions. A quick and easy system in blogs and the interactive environment of online forums provide a clear mean by which tacit knowledge can be shared. So knowledge intensive VCoP would have a positive impact on the sharing of tacit knowledge, i.e. tacit knowledge which would be shared spontaneously in an office, about new technologies or about company news founds a natural home in a VCoP.
Often a transition takes place from a face-to-face to a virtual CoP, in order to reach more continuous levels of information sharing. The lack of face-to-face contact within a CoP can often be an advantage, because it helps to suppress traditional group norm behavior. On the other hand, it remains open if a community of practice where face-to-face contact is entirely excluded can be sustained over a long period. Despite the great potential, there are also limitations of current technologies in

relation to virtual communities of practice: because virtual community infrastructure can be set up across cultures via Web, cultural and language differences can change the interactions and hinder the flux of activities in the CoP. Such aspects have been considered in the example presented next.

4 Example

We applied the above ideas within the activities of the EU project SIMPEL (SME Improving Practice in eLearning) tracking the suitability and our usage of Web 2.0 and utilizing the CoP structure as an intensive KM environment [4]. We developed strategies to enable SMEs to take full advantage of the eLearning in their training. We involved SMEs and eLearning experts in two communities of practice (one European and one German) [12] to share learning and knowledge and to develop continuous vocational education strategies based on Web 2.0 leading towards the creation of dynamic personalized learning environments. The European CoP is a loosely coupled (weakly framed) CoP, the German one strongly framed where the transmission of knowledge occurred closely between its members.

In the European CoP an "optimal vocational training model" based on eLearning in SMEs was developed. Best practice for capturing and sharing of knowledge and for using eLearning have been collected and guidelines for using them written. This CoP attracted sectors engaged in support, training, design/development, use, in consulting and in policy formulation concerning eLearning in SMEs in the European Union.

The German CoP focused particularly on analysis and testing how informal, workplace oriented learning can be used efficiently in SMEs by working and acting in CoPs. The topic was chosen because analysis shows that individual SME staff show more interest in achieving of competences based on intensive KM [11] for things they can do, rather than for certification. The framework of the CoP is useful for informal learning, knowledge sharing; the social participation of the members is the key for informal learning being embedded into practices and relationships of the workplace. For example keeping up-to-date with administrative and technical changes necessary to solve the daily tasks efficient, and strategies to help solve problems and communicate with colleagues and co-workers. This CoP has permanent members who make regular contributions but also occasional members who use the information and knowledge needed for their work and business and sometimes contribute. For the future it is intended to encourage more SME to participate and to use the CoPs knowledge and resources developed.

In looking for a suitable software to support communities of practice and to facilitate the processes of knowledge sharing and learning, the SIMPEL consortium decided on Moodle [7]. The choice of Moodle was first based on an analysis of some open source virtual learning environments (VLEs) referring sustainability and viability (that influence the costs for adoption and further developments of the system with personalized - PLE services) and of the pedagogical rationale of the environment (how the VLE fits the pedagogical aims of the organisations which uses it).

Some of the key points for evaluating sustainability and viability refer to implementation, maintainability and further developments and are: activity of the community, level of usability, requirements in hardware and software, reliability of the system, support, modular system architecture, compatibility with existing systems within SMEs. Moodle is used also because some project partners have experience with it. The platforms for supporting the two CoPs provide members with tools to capture and share knowledge like blogs, Wikis. The platform for the German CoP offers sites around typical work tasks and roles particularly in a virtual enterprise. Because the project partners have experience with scenarios a scenario was built showing how a knowledge intensive virtual enterprise model can support eLearning at work experience. The scenario was based on the idea that several organizations which CoP members belong put together their knowledge bases, technologies, competences, practice experience of the members. A wide range of media (music, video, animation) have been used to communicate the message of the scenario. Based on the scenario, knowledge-enabling services improved eLearning services, document sharing services and eCommerce services are in the development.

References

1. Attwell, G., Dirckinck-Holmfeld, D., Fabian, P., Kárpáti, A., Littig, P.: Thematic Monitoring under the LEONARDO DA VINCI Programme, Bonn, Germany. Report, Impuls 010 (2003)
2. Averill, S., Hall, T.: An observatory of eLearning in Small Medium Enterprises (SMEs) in Europe - The Promise versus the Reality. In: Richards, G. (ed.) Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare and Higher (2005)
3. Beer, D., Busse, T., Hamburg, I., Mill, U., Paul, H. (eds.): e-learning in European SMEs: observations, analyses & forecasting. Münster, Waxmann (2006)
4. Beer, D., Busse, T., Hamburg, I., Oehler, C. (eds.): Improving eLearning practices in SMEs. Brussels. In: Proceedings of the SIMPEL final conference, April 14. Universitas-Györ (2008)
5. Castro, M.C.: Revisiting Communities of Practice: from fishermen guilds to the global village. Posted on 28 July 2006 by jrobes (2006)
6. Diemers, D.: Virtual Knowledge Communities. Erfolgreicher Umgang mit Wissen in digitalen Zeitalter. Dissertation der Universität, St. Gallen (2001)
7. Dougiamas, M.: Moodle: Using Learning Communities to Create an Open Source Course Management System, Perth, Australia (2004)
8. European Commission, Observatory of European SMEs (2003)
9. Hamburg, I., Lindecke, C., Terstriep, J.: Analysing interaction with computer of SME staff in eLearning processes. In: HCI International 2005: 11th International Conference on Human-Computer-Interaction, July 22-27, p. 10. MIRA Digital Publ., Las Vegas (2005)
10. Hamburg, I.: Shifting eLearning in SMEs to a Work-based and Business Oriented Topic. In: European Distance and ELearning Network: New learning 2.0? Emerging digital territories - developing continuities - new divides; THE EDEN, Annual Conference, June 13-16, p. 4 (2007); Naples: CD-ROM. Budapest: EDEN

11. Hamburg, I., Engert, S.: Competency-based Training in SMEs: The Role of ELearning and E-Competence. In: Proceedings of the 6th IASTED International Conference "Web-based Education", March 14-16, pp. 189–193. Acta Press, Chamonix (2007)
12. Hamburg, I., Engert, S., Petschenka, A., Marin, M.: Improving e-learning 2.0-based training strategies on SMEs through communities of practice. In: The International Association of Science and Technology for Development: The Seventh IASTED International Conference on Web-Based Education, Innsbruck, Austria, March 17-19, pp. 200–205 (2008)
13. Krug, S.: Dont Make Me Think! A common sense approach to web usability. New Riders, Indianapolis (2000)
14. Krogh, G., Ichijo, K., Nonaka, I.: Enabling Knowledge Creation. In: How to Unlock the Mystery of Tacit Knowledge and Release the Power of Innovation. Oxford University Press, New York (2000)
15. Nielsen, J.: Finding usability problems through heuristic evaluation. In: Proceedings ACM CHI 1992 Conference, Monterey, pp. 373–380 (1992)
16. Nonaka, I., Konno, N.: The concept of 'ba': building a foundation for knowledge creation. California Management Review 40(3), 40–54 (1998)
17. OECD, ELearning in tertiary Education: Where do we stand? OECD. Paris (2005)
18. O'Brien, E., Hall, T., Johnson, K.: The Importance of Training Needs Analysis in Authoring Technology Enhanced Learning for Companies. In: Technology Enhanced Learning Tools. IGI Global publication (2008)
19. O'Reilly, T.: What is Web 2.0. Design patterns and Business models for the next generation of Software (2005), http://www.oreillynet.com/lp/a/6228
20. Say, M.: Government Computing Magazine (2006)
21. Siemens, G.: Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning (2005)
22. Wenger, E., McDermott, R., Sydner, W.: Cultivating communities of practice: a guide to managing knowledge. Harvard Business School Press, Boston (2002)

Visual Interface for Content-based Query in Image Databases

Cosmin Stoica Spahiu and Liana Stanescu, Dan Burdescu, and Marius Brezovan

Abstract The paper presents a relational database management system for managing and querying visual information. In order to accomplish this, the DBMS has implemented the Image data type. In a record of this type, there are stored the binary image and the automatically extracted color and texture characteristics. The color histogram with 166 colors in HSV space represents the image color information. A vector with 12 values represents the texture information obtained by applying Gabor filters. The two characteristics vectors are used for content-based visual query process. Beside this original way for visual information storage, the DBMS has a visual interface for building content-based queries using color, texture or both. Adapted for this type of queries, a Select command is built and executed by the DBMS. This system might by easily used in areas, where medium sized image collections are collected, in an efficient way and with low cost.

1 Introduction

Nowadays there is an explosion of multimedia information and the visual information has an important place. The visual data as the other types of multimedia information is very complex, it needs a lot of storage space, update operations, and it has to permit querying. To solve all these demands in an efficient way, a multimedia database management system (MMDBMS) is needed. The MMDBMS must have a series of important characteristics: multimedia data types support; possibility to manage a large number of multimedia objects; hierarchical storage management;

University of Craiova,
Faculty of Automation, Computers and Electronics,
Craiova, Romania
e-mail: `Stoica.Cosmin, Stanescu, Burdescu@software.ucv.ro`
`Marius.Brezovan@software.ucv.ro`

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 231–240.
springerlink.com

conventional database capabilities; possibility to find multimedia information [17]. Retrieval of multimedia information might be implemented using [17]:

- Browsing, this technique is used in CAD/CAM/CASE applications or in hypermedia documents databases and implies finding objects, starting from the root.
- Taking into consideration attributes or characteristics extracted from multimedia information and finding multimedia objects that have in common these types of characteristics. This technique is called content-based retrieval.

For the visual information, this last type of search is called content-based visual query and can be implemented at image or region level. It is necessary to establish an image (region) as the image (region) query and to find all the images in the databases that are similar with it. It can be done using automatically extracted characteristics: color, texture, shape. Such a query can be expressed in a query language or using a visual interface [13]. One of the areas that use mostly the multimedia databases and content-based retrieval is the medical domain. That is because the diagnosis process produces a large number of images. In the medical domain, the content-based visual query is very useful for: diagnostic aid, medical teaching and medical research [14]. The paper presents a DBMS, which has an Image data type, a visual interface for building content-based image query and a modified Select command. The image data type has the possibility to binary store the images and the characteristics that were automatically extracted. The color information is represented by a color histogram in the HSV color space, quantified to 166 colors [8]. A vector with 12 values resulted from Gabor filters method represents the information about texture [7]. This DBMS has the following original elements and advantages:

- The way that images are managed
- Visual interface for content-based image query using color and texture characteristics
- Possibility to see the images when the records are viewed

This is a reason why it is recommended for managing medium sized image collections: for example the ones from personal medical offices. In these cases it is a good alternative compared to other products of high quality, but much more expensive. The paper has the following structure: section 2 presents the related work in this field, section 3 presents the general architecture of the MMDBMS, section 4 presents the Image data type, section 5 presents the proposed content-based query visual interface, and section 6 presents the conclusions and future work in this project.

2 Related Work

In order to manage content-based retrieval for medical image collections there have been implemented a series of applications. Most of them use traditional database management systems as MS SQL Server, My SQL or Interbase. MYSQL offers only the BLOB data type that can be used to store images. A BLOB is a binary large object that can hold a variable amount of data. BLOB columns have no character set, and sorting and comparison are based on the numeric values of the bytes in column values [1]. MS SQL Server offers a data Type called "image", but with no other support. It is considered a variable-length binary data from 0 through 231-1 (2,147,483,647) bytes. There are no pre-defined functions that can be used for extracting characteristics or building content-based queries. More than that, in the MS SQL Server 2008 it is specified that "ntext", "text" and "image" data types will be removed in a future version of Microsoft SQL Server. The recommendation is to avoid using these data types in new development work, and plan to modify applications that currently use them. It should be used nvarchar(max), varchar(max), and varbinary(max) instead [2]. The complete solution is provided by Oracle - the Oracle 10g database server and Intermedia tool that can manage all kind of multimedia data, including DICOM files. In addition to the image support offered via the ORDImage object type, in Oracle 10g version interMedia provides support for the first edition of the ISO/IEC 13249-5:2001 SQL/MM Part 5: Still Image Standard. This standard defines object relational types for images and image characteristics. Each object type includes attributes, methods, and associated SQL functions and procedures. Use of the SQL standard interface may make some applications more portable across various vendor databases. The following object types are defined by SQL/MM Still Image:

- SI_StillImage - holds digital images with inherent image characteristics such as height, width, format, etc.
- SI_AverageColor - describes the average color feature of an image.
- SI_Color - encapsulates color values of a digitized image.
- SI_ColorHistogram - describes the relative frequencies of the colors exhibited by samples of an image.
- SI_PositionalColor - describes the positional color feature of an image.
- SI_Texture - describes the texture feature of the image.
- SI_FeatureList - describes an image that is represented by a composite feature based upon up to four basic image features and their associated feature weights [3, 4].

This kind of solution is not always justified and cannot be implemented by small private doctor's cabinets. The reason is that involves high costs for buying the database server and for designing and implementing complex applications for content-based visual query [5, 6]. None of these systems offers Query-by-Example (QBE) facilities to build content-base visual queries. QBE represents another language for querying. It is different from SQL, and from most other database query languages, in having a graphical user interface that allows users to write queries by creating example tables on the screen. This way a user needs minimal information to create simple text based queries or content-based retrieval queries. The whole language contains relatively few concepts. QBE is especially suited for queries that are not too complex and can be expressed in terms of a few tables [11, 12].

The application presented by us includes both a DBMS kernel and a client interface that can be used to build text based queries or content based retrieval queries. This interface is totally independent from the server thus can be adapted to users' needs without being afraid that you have to modify the server.

3 The Architecture of the MMDBMS

This dedicated MMDBMS allows database creation, table and constraints adding (primary key, foreign keys), inserting images and alphanumerical information, simple text based query and content-based query using color and texture characteristics. The software tool is easy to be used because it respects the SQL standard. It does not need advanced informatics knowledge and has the advantage of low cost. It is a good alternative for a classical database management system (MS Access, MS SQL Server, Oracle10g Server and Intermedia), which would need higher costs for database server and for designing applications for content-based retrieval. Figure 1 presents the general architecture of the MMDBMS.
In the first step the application that uses the server must connect to the database. This way it will be created a communication channel between them. All commands and responses will use this channel to send queries requests and receive answers. The server has two main modules: kernel engine and database files manager. The kernel engine includes all functions implemented in the server. It is composed from several sub-modules each of them with specific tasks:

- The main module. It is the module, which manages all communications with the client. It is the one that receives all queries requests, check what is the type of query requested, extracts the parameters of the query and calls the specific module to execute it.
- Queries response module. After the query is executed, the results will be sent to the Queries Response Module. It will compact the result using a standard format and then return it to the client. The client will receive it on the same communication channel used to send the request.
- Select/Select/Image Processing module. If the main module concludes that is a SELECT SQL command, it will call the Select Processing module. This module extracts the parameters from the query and then search in the database files for specific information. If the query is a SELECT IMAGE query, it will use for comparison the similitude of characteristics instead equality of parameters.
- Characteristics Extraction Module. When the main module receives a SELECT IMAGE or a UPDATE query which uses an image that is not already in the database it is needed first to process it. This module is called to extract the color and texture characteristics of the image. The data of the results will be used to initialize a variable of IMAGE data type.
- Update Processing Module. When the query received from the user is an UPDATE command, it will be called to execute it.
- Delete Processing Module. It is called when the user executes a DELETE command. The kernel executes only logic deletes. It never executes physical deletes.

The physical deletes are executed only when a “Compact Database” command is sent by the user.

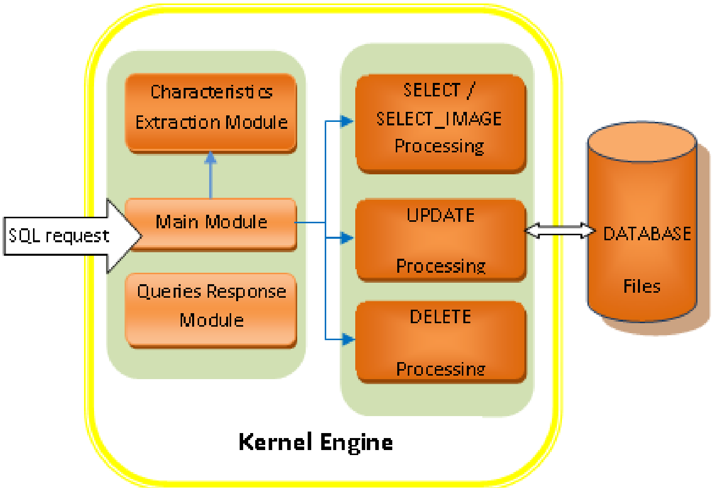


Fig. 1 General architecture of the system

Fig. 1: General architecture of the system.

The second main module is the Database Files Manager. It is the only module that has access for reads and writes to the files in the database. It is his job to search for information in the files, to read and write into files and to manage locks over databases. When a client module request a read form a file it is enabled a read lock for the specific file (that represents a table in the database). All other read requests will be permitted but no writes will be allowed. If the client module request a write to file, it will be enabled a write lock. No other requests will be allowed until the lock is canceled.
The results will always be returned to the module, which made the request. The data read or wrote to files is not structured in any way. This module does not modify the structure in any way. All the results will be raw data, as it is read from the files or received from client modules.

4 The IMAGE Data Type

Along with the classical functions of a DBMS, an element of originality of this system is that offers the possibility to insert new images in the database, together with their relevant information, namely: path and name of the image file, the diagnosis, treatment, evolution, etc. The images are stored directly in the database in a special data type called IMAGE. This type can be used to store all the information regarding the image: color characteristics, texture characteristics, width, height, etc. For implementing the content-based visual query, all the images loaded in the database are automatically processed in a two step process and stored as Image data:

1. The extraction of color feature
2. The extraction of texture feature

For color feature extraction, the images are pre-processed, namely they are transformed from the RGB color space to HSV color space and quantized to 166 colors, being thus prepared for a future query [8, 9]. The color information is stored in the IMAGE variable as a vector with 166 values and it is used furthermore in the content-based image query and content-based region query [8, 9].
In content-based visual query on colour feature (the colour is the visual feature immediately perceived on an image) the used colour space and the level of quantization, meaning the maximum number of colours are of great importance. The colour histograms represent the traditional method of describing the colour properties of the images. They have the advantages of easy computation and up to certain point are insensitive to camera rotating, zooming, and changes in image resolution [7, 8]. The operation of colour quantization is needed in order to reduce the number of colours used in content-based visual query from millions to tens. J. R. Smith proposed the chosen solution - the quantization of the HSV color space to 166 colors [8]. There are many techniques used for texture extraction, but there is not any certain method that can be considered the most appropriate, this depending on the application and the type of images taken into account. We optimized the kernel for medical images where the experiments on medical images indicated that among the most representative methods of texture detection are the Gabor representations. This is the reason for which it was chosen for extracting the colour texture feature from medical images in the database [13]. After the algorithm execution [10], the resulted image texture feature will be a 12-dimension vector that is also added to the IMAGE variable. The results of the content-based visual query process on medical imagery, using the colour histograms with 166 values in HSV colour space and Gabor Filters were presented in [10].

5 The Content-Based Visual Query Interface

The client interface for the Multimedia Database Management System offers the possibility to build the content-based visual query using colour characteristic, texture characteristic or a combination of them, in an easy manner, at the image level.

The elements of this visual interface for content-based retrieval are:

- Similar With - opens the window for choosing the query image
- Select - permits to choose the field (or fields) that will be presented in the results of the query

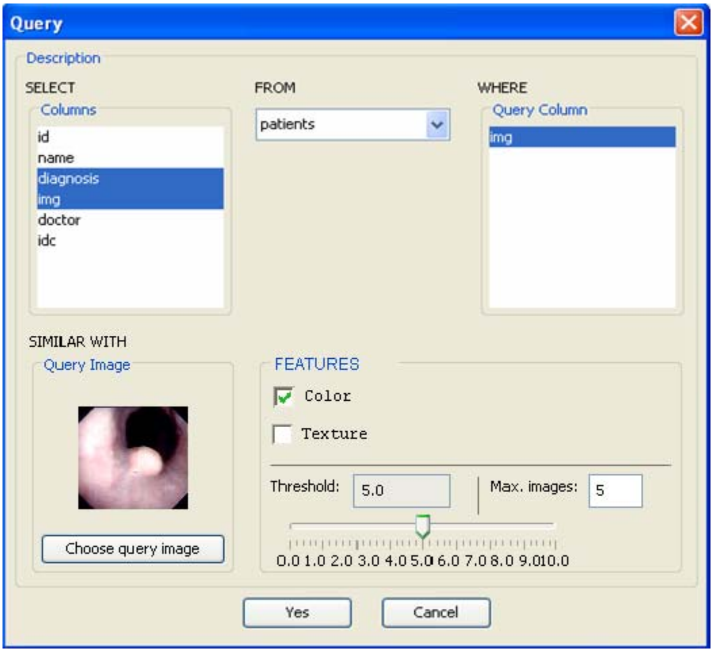


Fig. 2: The interface for building the content-based image query.

- From - it is one of the tables in database, that will be used for the query
- Where - the image type column used for content-based image query
- Features - the characteristic used for content based visual query is chosen - colour, texture or a combination of them
- Threshold - a threshold of accepted similarity between query image and target image is chosen. An image with a similarity under this threshold will not be added into the resulted query images
- Maximum images - specify the maximum number of images returned by the query.

When building the query, a modified SQL Select command is actually performed, which is adapted for content-based image query. An example of such modified command is: Select patients.diagnosis, patients.img From Patients where Patients.img Similar with Query Image (method : color, max.images 5). This modified Select command specifies that the results are obtained from Patients table, taking into consideration the values from diagnosis field, the images similar with the query image for colour characteristic, and there will be 5 resulting images. The query created in this manner is sent to the main module of the kernel. It parses the query and concludes it is a SELECT_IMAGE command. The execution of this special SELECT will be a three-step process:

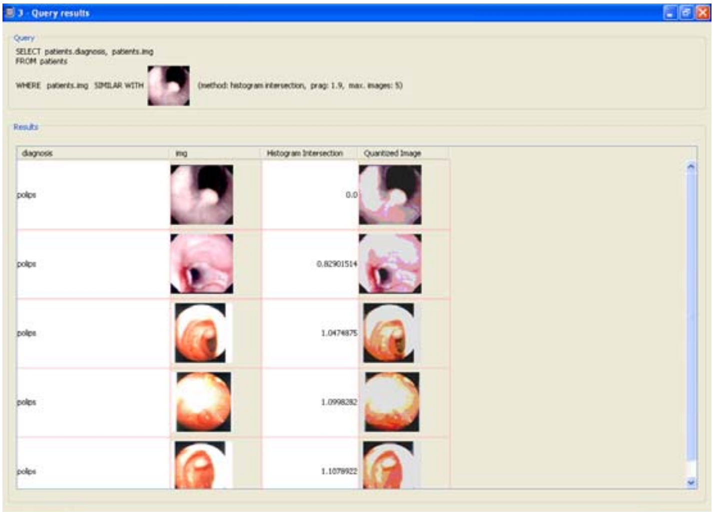


Fig. 3: The results of the content-based image query.

- It is called the Characteristics Extraction Module that will extract the color and texture characteristics of the query image. The result will be put in an Image Type variable.
- It is called the Select Image Processing Module that will search in the database similar images. It will compare the query image characteristics with the characteristics of the target images from the database.
- The Queries Response Module will send the images with similar characteristics back to the client interface. Before sending the results it will format the result to a standard type in order to be correctly interpreted by all types of clients. In the resulting set it is also presented the dissimilarity between the query and target images.

6 Conclusions

The paper presents the organization of an implemented multimedia, relational, database management system that has an original data type, called IMAGE. It is created for managing and querying medium sized personal digital collections that contain both alphanumerical information and digital images (for examples the ones used in private medical consulting rooms). The software tool allows creating and deleting databases, creating and deleting tables in databases, updating data in tables and querying. The user can use several types of data as integer, char, double and image. There are also implemented the two constraints used in relational model: primary key and referential integrity. For the content-based retrieval the MMDBMS contains a query-by-example visual interface that allows the user to execute content-based visual query at image level, taking into consideration the color and texture characteristics. These characteristics are automatically extracted when the images are inserted in the database. The QBE type visual interface allows users to built queries in a simple manner, by creating example tables on the screen. This way a user needs minimal information to create necessary queries.

The advantages of using this intelligent content-based query visual interface are that the specialist can see images from the medical database that are similar with the query image taking into consideration the color and texture characteristics. In this way the specialist can establish a correct medical diagnosis based on imagistic investigation frequently used nowadays. This software can be extended in the following directions:

- Adding new types of traditional and multimedia data types (for example video type or DICOM type - because the main area where this multimedia DBMS is used it is the medical domain and the DICOM type of data is for storing alphanumerical information and images existing in a standard DICOM file provided by a medical device)
- Studying and implementing indexing algorithms for data inserted in the tables

References

1. MySQL 5.0 Reference Manual (2009), http://dev.mysql.com/doc/refman/5.0/en/blob.html
2. SQL Server 2008 Books Online (January 2009), http://msdn.microsoft.com/en-us/library/ms187993.aspx
3. Oracle Database SQL Reference 10g Release 2 (10.2), http://download.oracle.com/docs/cd/B19306_01/server.102/b14200/sql_elements001.htm
4. New interMedia Features in Oracle10g, http://www.oracle.com/technology/products/intermedia/htdocs/imedia_new_features_in_10g.html
5. Chigrik, A.: SQL Server 2000 vs Oracle 9i (2007), http://www.mssqlcity.com/Articles/Compare/sql_server_vs_oracle.htm
6. Kratochvil, M.: The Move to Store Image in the Database (2005), http://www.oracle.com/technology/products/intermedia/pdf/why_images_in_database.pdf
7. Del Bimbo, A.: Visual Information Retrieval. Morgan Kaufmann Publishers, San Francisco (2001)
8. Smith, J.R.: Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis. Ph.D. thesis, Graduate School of Arts and Sciences. Columbia University (1997)
9. Gevers, T.: Image Search Engines: An Overview. Emerging Topics in Computer Vision. Prentice Hall, Englewood Cliffs (2004)

10. Stanescu, L., Burdescu, D.D., Brezovan, M., Stoica Spahiu, C., Ion, A.: A New Software Tool For Managing and Querying the Personal Medical Digital Imagery. In: Proceedings of the International Conference on Health Informatics, Porto, Portugal, pp. 199–204 (2009)
11. Samuel, J.: Query by example (QBE), http://pages.cs.wisc.edu/~dbbook/openAccess/thirdEdition/qbe.pdf
12. Query by Example, http://en.wikipedia.org/wiki/Query_by_Example
13. Muller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A Review of Content_based Image Retrieval Systems in Medical Application - Clinical Benefits and Future Directions. Int. J. Med. Inform. 73 (2004)
14. Khoshafian, S., Baker, A.B.: Multimedia and Imaging Databases. Morgan Kaufmann Publishers, Inc., San Francisco (1996)

Enhancement of computer character animation utilizing fuzzy rules

Piotr Szczuko and Bożena Kostek

Abstract A new method for processing of character animation is presented. It involves fuzzy inference with both rules and membership functions derived from results of subjective evaluation tests. During processing a new motion phases are added to an animation increasing its quality and changing fluidity and stylization of motion. Animation parameterization is presented, new parameters are designed, and the relation between coefficients proposed and subjective features of motion are established. Quality and fluidity increase are verified during subjective evaluation of animations processed by the created animation enhancement system.

1 Introduction

Animated computer characters, often used for virtual reality applications, computer games, educational software, serve as actors in animated movies. Currently the most advanced method for acquiring animated movement is Motion Capture, alas it has very high technical requirements [10]. Other drawbacks of this method are motion representation being unfriendly for animator and hard to edit by hand, yielding correction of mistakes almost unachievable. Capturing motion of a real actor does not enable to achieve exaggerated animation, typical for the animated movies and cartoons, and moreover changing the animation style is practically impossible. Other animation methods are developed, aiming at and achieving interesting and high quality motion without motion capture. For example utilizing a single controller movement as an input [2], digitalizing hand drawn animations [4], a physical simulation [5], a simulation of the human muscular system [6], prototyping of the motions with sketches [7], and finally an automatic assessment of the motion realism [14].

Multimedia Systems Department,
Electronics, Telecommunication and Informatics Faculty,
Gdańsk University of Technology, Poland

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 241–251.
springerlink.com

In the paper a new method for creating animated movement is proposed, combining computer animation with traditional cartoon one [1][13]. It is a continuation of work described in the earlier work by the authors [11][12] now with a relation between coefficients being established and quality and fluidity increase verified. As an effect of processing a fluid, high subjective quality animated movement is achieved, with a style spanning between natural and exaggerated depending on the user input. Advanced computer animation techniques used in the authors' approach guarantee effectiveness in designing an animated movement while utilizing keyframe principles well-known to animators [3]. The traditional animation approach serves as a source of practical rules for correct utilization of motion phases, assuring fluency of motion, readability of message, correct display of character personality. That knowledge is available in an animation literature, formed during last hundred years of animation evolution [1][13].
A new method proposed and verified in our work combines both domains of animation techniques utilizing at the same time fuzzy processing. Fuzzy logic allows for operating on non-crisp values and imprecise terms, and for processing linguistic variables related to subjective descriptors originating in human perception of motion features. Using fuzzy logic within this scope is, according to our knowledge, a unique approach to the computer animation.

2 Aim of the work

The main goal of our work is to verify a possibility of enhancing computer animation using computer intelligence methods, namely fuzzy logic, in such a way that the automatically acquired animated motion will be fluid, stylized and subjectively of high quality. Fuzzy logic is used for modeling the animation knowledge represented as relations between various motion parameters. That knowledge is obtained from the literature sources and also gathered from the data mining of parameters extracted from animation examples. Knowledge processing enables to indicate those parameters that have the strongest influence on subjectively evaluated motion features, namely its *quality*, *fluidity* and *style*. The paper is organized as follows: first in sec. 3 an animation creation model is formalized. In sec. 4 motion parameters are introduced, and relations between them are being discussed in sec. 4.1. Analysis of subjective tests results leads to formulation of evaluation function, presented in sec. 4.2, which associates parameters of motion. Next upon the function rules are built, utilized in fuzzy inference, discussed in sec. 4.3. The fuzzy system is used for enhancement of animations, and the results are presented in sec. 5, concluded in sec. 6.

3 Animation creation model

For further analysis we proposed a model of animation creation process, presented in Fig. 1. The input data for the model are: director/animator requirements related to action, and timing, placing of animated objects (parameters in vector A) and requirements related to quality (parameters in vector Q). The mechanism of the character movements comprises mainly limbs, therefore the animated objects are character's limbs. These parameters describe sequence of the character's poses defining the action. A working version of the animation is prepared by the animator, fulfilling technical requirements defined in A, but also accompanied with additional motion phases not defined by director (parameters in vector B). Additional phases do not change the main action, but add subtle variations to transitions between poses, influencing character personality, and motion fluidity, quality, and style. The working version is subjectively evaluated, and the resulted quality features are stored in vector $Q*$ for comparison with the given Q being expected to be fulfilled by the final animation. If the comparison results are not satisfactory, the animation is corrected, i.e. motion features related to given requirements stored in A are not modified, but additional phases proposed by animator stored in B are changed. The animator should decide which parameters of B should be changed for achieving best results. The process is repeated until the worked-out animation has features close to the assumed ones, then it is provided as a final version at the output.

- motion parameters stored in vector A are not changed during the preparation of output animation, i.e. sequence of poses in time and space remain intact,
- depending on the assumed action described in vector A, parameters stored in B should be changed in such a way that the quality requirements stored in Q are fulfilled as precisely as possible.

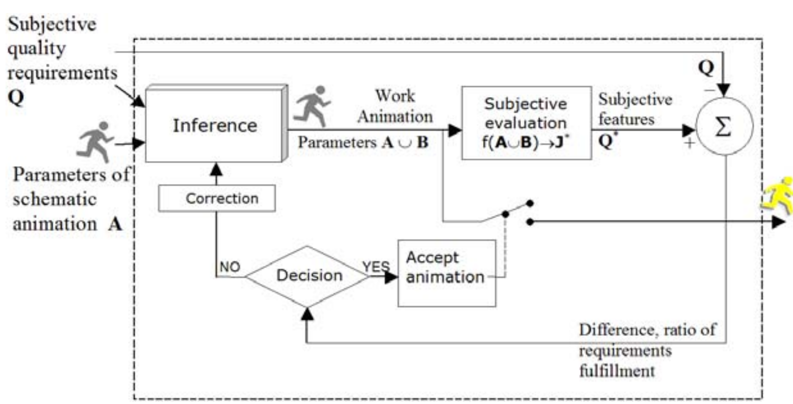


Fig. 1: Model of the animation creation process.

Adding new keyframes does not change the objective action of the character, but influences subjective nature of the motion. This is used by the animator, as he adds phases of *anticipation* (a motion preceding main transition between two poses adjacent in time), *overshoot* (a motion after transition, when motion stops slowly not abruptly), and *moving hold* (subtle changes of a pose while actions stops). Anticipation displays preparation for the action, e.g. a squat before jump, an overshoot portraying inertia, and moving hold is responsible for showing balancing and maintaining aliveness. A correct utilization of these additional phases influence naturalness and fluency of motion, related to high subjective quality of animation.
For simplification of the problem the following assumptions have to be made:

1. *anticipation* and *overshoot* are alike, i.e. their times (t) and limbs rotations amplitudes (A) are assumed to be equal, therefore for parameterization only two values are used: $\mathrm{d}A = A_a = A_o$ and $\mathrm{d}t = t_a = t_o$;
2. times and amplitudes of these phases are limited and cannot extend beyond subjectively accepted values: $\mathrm{d}t \in (0;10)$ [frames], $\mathrm{d}A \in (0;0,265\pi)$ [rad];
3. *moving holds* are calculated as random movements, with times and amplitudes taken from the limited ranges, however this is not in the scope of the paper.

4 Motion Parameterization and Processing

The processed motion parameters are presented in Fig. 2. Parameters related to poses, taken from the director/animator's requirements, are stored in vector A. Parameters connected to subjective features and related to additional motion phases are stored in vector B:

$$A = [V,A,t] \quad B = [dA,dt] \tag{1}$$

where: $A = a_3 - a_2$, $t = t_3 - t_2$, $V = A/t$.

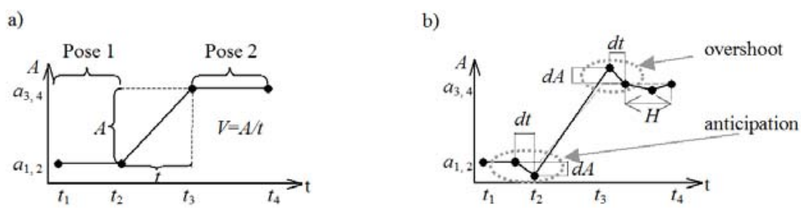


Fig. 2: Motion data: a) transition between two poses is described with parameters $A = [V,A,t]$, b) additional phases of anticipation, overshoot and moving hold. “H” depicts range of subtle random variation of the target pose.

4.1 Proportionality Coefficients

Traditional animation rules describe a way of the additional phase utilization for achieving particular subjective results. Namely, fast motion with a big amplitude should be preceded by a big anticipation and finished with a big overshoot; a long motion should be preceded with a long anticipation and long overshoot. Taking these into consideration one can assume that the proportionality occurs between these parameters, which can be described as:

$$dA = alpha \cdot V \cdot A \qquad dt = beta \cdot V \cdot t \tag{2}$$

where $alpha$ and $beta$ are yet unknown proportionality coefficients. In case when dA or dt exceeds the assumed maximal value[1], following $f(x)$ and $g(x)$ saturation functions are used:

$$\begin{gathered} dA = f(alpha \cdot V \cdot A), \text{ where } f(x) = 0{,}265\pi \cdot tgh(x/0{,}22\pi) \\ dt = g(beta \cdot V \cdot t), \text{ where } g(x) = 10 \cdot tgh(x/0{,}125) \end{gathered} \tag{3}$$

Considering Eqs. (2) and (3) variables dA and dt depend only on coefficients alpha and beta. If changing dA and dt influences subjective meaning of animation therefore a relation should exist also between subjective features and $alpha$ and $beta$. Relations between objective numerical values of $alpha$, $beta$ and subjectively evaluated animation features: *style* and *fluidity*, are specified during data mining results of subjective test. In tests a simple animations where used, containing two poses and a transition, with additional anticipation, overshoot and moving hold. Test participants' task was to name features of motion utilizing following discrete scales: $style = \{natural, middle, exaggerated\}$; $fluidity = \{fluid, middle, abrupt\}$; $quality = \{1, 2, 3, 4, 5\}$. Results of fluidity and style evaluation are stored in vector $Q = [style, fluidity]$, quality scores are processed individually as a additional criterion QS. The evaluation of visual stimuli was performed with respect to ITU-T and ITU-R recommendations for subjective determination of transmission quality and assessment of the quality of television picture [8][9]. A correlation between $alpha$ and $beta$ of each evaluated animation and its $style$ and $fluidity$ scores was calculated. The results are presented in Table 1. Strong correlation indicates that certain connection between selected subjective feature and proposed coefficient exists, and rules describing that relation can be created.

[1] It was verified during visual tests what maximal values of amplitudes dA and times dt are subjectively accepted by the viewers.

Table 1: Correlation coefficients between subjective and objective parameters of animations.

	beta-style	beta-fluidity	beta-quality	alpha-style	alpha-fluidity	alpha-quality	style-fluidity	style-quality	fluidity-quality
R	-0.14	0.86	0.81	0.82	0.16	0.09	-0.21	-0.27	0.94

4.2 Inverse of Evaluation Function

For creating rules that connect objective and subjective parameters of animation a relation between them should be known. Some information about that relation is obtained during subjective evaluation tests, when participants for an animation described with A and B select values for subjective features. A particular animation described by $A = [V, A, t]$ and $B = [dA, dt]$ is not evaluated identically by every test participant. Projection from B to Q with given A is denoted as $f_A : B \rightarrow Q$. That ambiguous function is called **evaluation function** and reflects viewers' answers. It is assumed that A and Q are given as a director's requirements. For the output animation only B parameters are unknown. Therefore the inverse function is being sought, $f_A^{-1} : Q \rightarrow B$, which for a given A and required values of features Q picks correct B. That function is also ambiguous. First, for each A and Q a result is generated as a set of objects - animations that have a motion as the one defined in A and that where subjectively evaluated as having values matching given $Q = [style, fluidity]$, but differentiated by B. From this set one object is selected based on an additional criterion, i.e. maximization of *quality score* **QS**[2]. Therefore for any given A, it is possible to generate unambiguous rules connecting values of given subjective requirements Q with parameters of additional phases B. These rules describe what values of additional phases to use if the working animation is described by values stored in vector A and animation subjective features values should match those given in Q. Fig. 3 is the graphical representation of the unambiguous result searching problem described above. Discrete hyper cubes of parameters visualize finite and discrete domains of multidimensional parameters vectors Q and B. A given Q value for the inverse function $f_A^{-1} : Q \rightarrow B$ points to various B values and the result is ambiguous. On the other hand for any B there exists an additional mediation value **QS**, quality score obtained in the subjective test. Maximal **QS** is sought for all Bs obtained in first step (area marked by dashed line), and only one result B is selected.

Equations (2) and (3) describe proportionality between V, A and t values stored in vector A, and parameters of additional motion phases dA and dt stored in vector B. Coefficients of these proportionalities, i.e. $alpha$ and $beta$, are used for simplification of searching for $f_A^{-1} : Q \rightarrow B$. The problem is first reduced to defining relations $Q \rightarrow [alpha, beta]$ as seen in Table 1, then the inverse function $f_A^{-1} : Q \rightarrow B$ is

[2] Instead of subjective quality other criterions can be used: new subjective feature (its values should be evaluated during tests), or objective measures, e.g. total energy used for the movement, leading to less or more energetic movement.

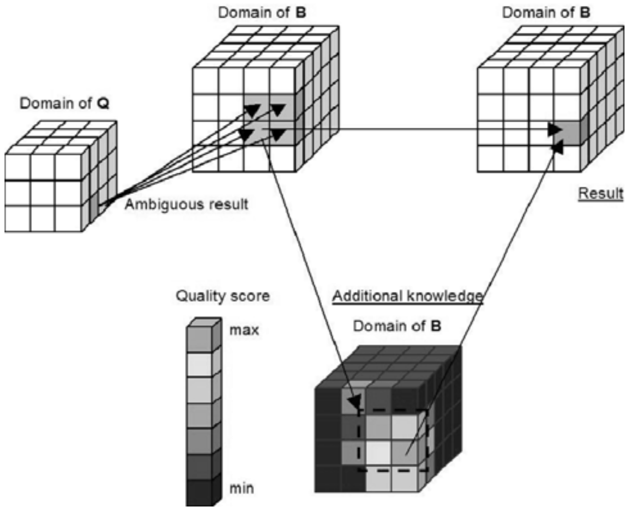


Fig. 3: Illustration of unambiguous result searching for the given A. Other As lead to different mapping between Q and B.

calculated for any given A based on relation (3). Obtained relations between subjective variables and proportionality coefficients are presented in Table 2.

Table 2: Calculation of $alpha$ and $beta$ based on given fluidity and style.

alpha		fluidity		
		abrupt	middle	fluid
style	natural	0.7	0.5	0.3
	middle	0.9	0.7	0.5
	exaggerated	1.3	1.1	0.9

beta		fluidity		
		abrupt	middle	Fluid
style	natural	3	5	7
	middle	1	5	5
	exaggerated	3	5	7

4.3 Generation of rules

The set of rules was obtained based on knowledge gathered during performed animation subjective evaluation tests. These rules allow calculation of parameters of additional motion phases that are enhancing movement of the animated character, in a way required by the user. These rules are implemented in the fuzzy inference system, where input and output linguistic variables describing motion features are processed by fuzzy logic. These rules are as follows:

$$\text{IF } V = \dots \wedge t = \dots \wedge \text{style} = \dots \wedge \text{fluidity} = \dots \text{ THEN } dt = \dots \tag{4}$$
$$\text{IF } V = \dots \wedge A = \dots \wedge \text{style} = \dots \wedge \text{fluidity} = \dots \text{ THEN } dA = \dots$$

For all parameters fuzzy membership functions are required, therefore input parameters V, A, t are first discretized: $V = \{0.0, 0.05, 0.1, 0.15, \dots, 0.4\}$, $A = \{0.1, 0.2, \dots, 1.0\}$, $t = \{5, 10, 15, dots, 50\}$. Then calibration animations are prepared for evaluation, presenting an animated character arm motion with speed, amplitude or time, chosen as one of the above discrete values.
These animations are evaluated utilizing linguistic descriptors:
$speed = \{low, medium, high, veryhigh\}$, $amplitude = \{low, medium, high, veryhigh\}$, $time = \{short, medium, long, verylong\}$. Based on evaluation results membership functions (mf) are created. For example for mf $speed = low$ as a kernel[3] a range of discrete values of feature speed for which participants selected linguistic value *low* more often than in 80% of cases is selected. Membership functions for variable amplitude are presented in Fig. 4a.
Crisp values obtained by calculation of Eq. (3) for all combinations of discrete values of input parameters V, A, t are also fuzzified. In the first step all crisp values were clustered using k-means algorithm, then triangle mf were created, each having maximum in a center value of respective k-th cluster, and spanning in a way that fulfills sum-to-one condition, as presented in Fig. 4b.

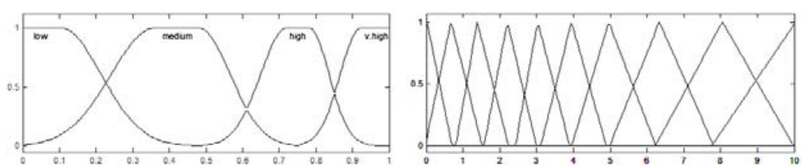


Fig. 4: Membership functions obtained for: a) linguistic variable $amplitude$, b) variable dt.

Fuzzy rules (4) are formulated for all three linguistic values of variables style and fluidity and for all four values of the discretized V, A, t. First for the given style and fluidity based on Table 2 coefficients $alpha$ and $beta$ are calculated, then from Eq. (3) the outcome values are calculated, which are finally fuzzified. Repeating that process for all combinations of input values, all needed rules are being formulated. Examples of rules obtained for calculation of fuzzy dA value are presented in Table 3.

[3] values with membership equal 1.

Table 3: Rules for calculating fuzzy value of dA for a given $naturalstyle$ and $fluidanimation$.

dA		V			
		low	medium	high	v. high
A	low	mf1	mf1	mf2	mf3
	medium	mf2	mf2	mf3	mf5
	high	mf2	mf3	mf4	mf6
	v. high	mf2	mf3	mf4	mf7

mf$n$ – n-th membership function for variable dA

5 Results

The effectiveness of the animation processing system was verified in visual subjective tests. Five animated actions served as the test material. These were enhanced using the system and all combinations of input descriptors for *style* and *fluidity*. All animations were rated using 5-point scale. For verification also non-processed versions of animations were displayed. These animations are the input animations presented in Fig. 1 on the left side, as schematic animations described only by A. Mean Opinion Score values for all animations are presented in Table 4. Processed animations obtained statistically valid higher scores than non-processed (Fig. 5). Moreover fluid motion was always rated higher than the abrupt one (Fig. 6a). Finally, variation of animation style does not have influence on quality scores (Fig. 6b), therefore the method developed for the animation enhancement can be applied to generate many versions of the single base motion, matching user's requirements for style and fluidity.

Table 4: Mean opinion score for animations depending on values of subjective descriptors used for enhancement.

Mean opinion score		style		
		natural	medium	exaggerated
fluidity	abrupt	2.08	2.06	2.08
	medium	3.28	3.22	3.1
	fluid	4.24	3.92	4.02
non-processed			1.5	

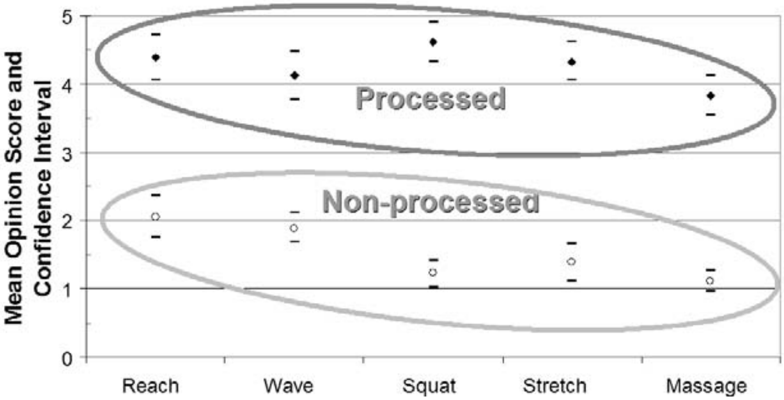


Fig. 5: Mean Opinion Score values with confidence intervals for processed and non-processed animations of five actions utilizing values $style : natural, fluidity : fluid$.

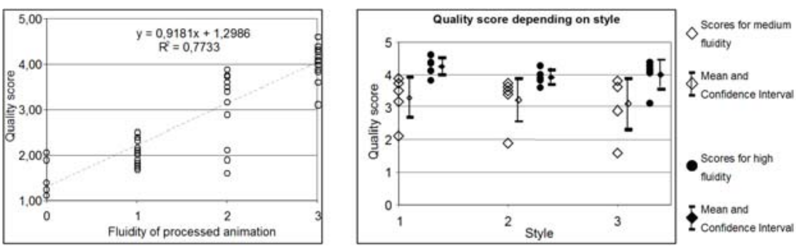


Fig. 6: a) Correlation between fluidity of animation and obtained scores for all five test animations. Fluidity “0” is for *non-processed*, “1” for *abrupt*, “2” for *medium*, “3” for *fluid*. b) Quality scores depending on style for two levels of fluidity: medium and high. Corresponding scores and mean scores are shown. Style “1” is for *natural*, “2” for *medium*, “3” for *exaggerated*.

6 Conclusions

Utilizing methodology and rules of traditional animation combined with fuzzy processing the system for animation enhancement was developed. In the system fuzzy rules are used for calculating parameters of the additional motion phases that should be inserted to animation for enhancement of its subjective quality and to introduce new subjective features such as stylization and fluidity of motion. New proportionality coefficients $alpha$ and $beta$ were defined that are strongly correlated with subjective features of animation.

References

1. Blair, P.: Cartoon Animation. Walter Foster Publishing, Laguna Hills (1995)
2. Dontcheva, M., Yngve, G., Popović, Z.: Layered Acting For Character Animation. In: Proc. SIGGRAPH, Los Angeles, pp. 409–416 (2003)
3. Lasseter, J.: Principles of Traditional Animation Applied to 3D Computer Animation. In: Proc. SIGGRAPH, Los Angeles, pp. 35–44 (1987)
4. Li, Y., Gleicher, M., Xu, Y., Shum, H.: Stylizing Motion with Drawings. In: Proc. SIGGRAPH, Los Angeles, pp. 309–319 (2003)
5. Milenkovic, V.J., Schmidl, H.: Optimization-Based Animation. In: Proc. SIGGRAPH, Los Angeles, pp. 37–46 (2001)
6. Neff, M., Fiume, E.: Artistically based computer generation of expressive motion. In: Proc. Symposium on Language, Speech and Gesture for Expressive Characters, pp. 29–39 (2004)
7. Popović, J., Seitz, S.M., Erdmann, M.: Motion Sketching for Control of Rigid-Body Simulations. Association for Computing Machinery Transactions on Graphics 22(4), 1034–1054 (2003)
8. Recommendation ITU-R BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Picture (2002)
9. Recommendation ITU-T P.800, Methods for Subjective Determination of Transmis-sion Quality. Geneva, Switzerland (1996)
10. Remondino, F., Schrotter, G., Roditakis, A., D'Apuzzo, N.: Markerless Motion Capture from Single or Multi-Camera Video Sequence. In: Int. Workshop on Modelling and Motion Capture Techniques for Virtual Environments (CAPTECH), Zermatt, Switzerland (2004)
11. Szczuko, P., Kostek, B.: Analysis and Generation of Emotionally Featured Animated Motion. In: The 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing (RSFDGrC), Regina, pp. 333–341 (2005)
12. Szczuko, P.: Application of fuzzy rules in computer animation. Ph.D Thesis, Gda?sk University of Technology, Poland (2008), `http://sound.eti.pg.gda.pl/animacje`
13. Thomas, F., Johnston, O.: Disney Animation - The Illusion of Life. Abbeville Press, New York (1981)
14. Vercher, J.L.: Perception and Synthesis of Biologically Plausible Motion: From Human Physiology to Virtual Reality. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS (LNAI), vol. 3881, pp. 1–12. Springer, Heidelberg (2006)

Comparative performance evaluation of classifiers for Facial Expression Recognition

Aristomenis S. Lampropoulos, Ioanna-Ourania Stathopoulou, and George A. Tsihrintzis

Abstract Towards building new, friendlier human-computer interaction and multimedia interactive services systems, we developed a image processing system which consists of the face detection module, which first determines automatically whether or not there are any faces in given images and, if so, returns the location and extent of each face and a facial expression classification module, which allow the classification of several facial expressions. In order to increase the accuracy of the facial expression classification module, we developed four different classifiers, namely:(1) Multilayer perceptrons, (2) Radial basis networks, (3) K-nearest neighbor classifiers and, (4) Support vector machines. In this paper we make an evaluation of performance of these classifiers versus the human's expression recognition performance for five expression: 'neutral', 'happy', 'surprised', 'angry' and 'disgusted'.

1 Introduction

During human-to human interaction and interpersonal relations, understanding the other's psychological state is quite common. While in human relations this is done almost automatically and by combining many indications of someone's state, e.g. facial expression, tone of voice, body posture, etc., identifying someone's expression only by his face image is quite challenging [1]. When mimicking human-to-human communication, human-computer interaction systems must determine the psychological state of a person, so that the computer can react accordingly. Indeed, in the design of advanced human-computer interfaces, images that contain faces are instrumental in the development of more effective and friendly methods in

University of Piraeus, Department of Informatics,
80 Karaoli and Dimitriou St, Piraeus 18534, Greece.
e-mail: `arislamp,iostath,geoatsi@unipi.gr`

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 253–263.
springerlink.com

multimedia interactive services and human computer interaction systems. Vision-based human-computer interactive systems assume that information about a user's identity, state and intent can be extracted from images, and that computers can then react accordingly. Similar information can also be used in security control systems or in criminology to uncover possible criminals.

Previous attempts to address problems of facial expression classification in images fall within one of two main directions in the literature:(1) methods using image sequences (video) ([2] ,[3], [4], [4], [5]) and (2) methods using static images. Approaches using image sequence often apply optical flow analysis to the image sequence and use pattern recognition tools to recognize optical flow patterns associated with particular facial expression. This approach requires acquisition of multiple frames of images to recognize expressions and thus has limitations in real-time performance and robustness. Facial expression recognition using still images can be divided in two main categories: (1) methods based on facial features ([6], [7],[8] ,[9], [10], [11]), and (2) methods that utilize image-based representations of the face ([12] ,[13], [14], [15], [16]). Methods that use facial features for facial expression recognition have fairly fast performance, but the challenge in this approach is to develop a feature extraction method that works well regardless of variations in human subjects and environmental conditions. Methods that utilize image-based representation have as an input the whole facial image which is preprocessed using various methods (e.g. Gabor Filters) or is given to classifier in order to recognize the facial expression. The aforementioned methods usually work well in generalizing for other face images, not in the database, but it's fairly difficult to train such a classifier. Finally, with the advances in technology, there are some new methods based on thermal imagery (e.g. [17]), but, in this cases, there is a need for a more sophisticated hardware, which makes it difficult to use in everyday human computer interaction.

Another fundamental issue about the facial expression classification is to define a set of categories we want to deal with. A related issue is to devise mechanisms of categorization. Facial expressions can be classified in various ways: (1) Methods that are trying to classify the image face in discrete facial emotions (e.g. [18], [19], [20]) and, (2) Methods that are trying to classify the image in terms of facial actions that cause an expression (e.g. [21], [22], [16], [23]), the majority of these methods use the Facial Action Coding System (FACS) [24]. An extended survey about all the aforementioned methods can be found in [25].

In this paper we identified five emotions that occur during a typical human-computer interaction, namely expressions corresponding to the "neutral", "happiness", "surprise", "anger" and "disgusted" psychological states. In order to achieve better results we evaluate the performance of four classifiers for the classification task: (1) Multilayer perceptron, (2) Radial Basis function network, (3) K-nearest neighbor classifier and (4) Support vector machines. Specifically, the paper is organized as follows: in Section 2, we present our system, where we describe briefly how the facial expression databases were acquired and facial which features we use for the classification task. We also describe the architecture of each classifier. In Section 3, we evaluate the performance of our system for each classifier. We draw conclusions and point to future work in Sections 4 and 5, respectively.

2 System Overview

2.1 Facial Expression Database

Since our search in the literature and World Wide Web didn't result to a complete facial expression database we built our own facial expression database. The process of acquiring image data and building this database is described extensively in [1]. The final dataset consisted of 250 different persons, each forming the seven expressions: ”neutral”, ”happy”, ”sad”, ”surprised”, ”angry”', ”disgusted” and ”bored-sleepy”.

2.2 Feature Description

From the collected dataset, we identified differences between the ”neutral” expression of a model and its deformation into other expressions. This led us to the identification of the some important facial features [1], that can represent these changes in mathematical terms, so as to form the feature vector. These facial points are widely used in facial processing systems and they can help us in the in the computation of the facial features which will be used as an input to the classifiers. The aim of feature extraction process is to convert pixel data into a higher-level representation of shape, motion, color, texture and spatial configuration of the face and its components. Specifically, we locate and extract the corner points of specific regions of the face, such as the eyes, the mouth and the eyebrows, and compute variations in size or orientation from the ”neutral” expression to another one. Also, we extract specific regions of the face, such us the forehead or the region between the eyebrows, so as to compute variations in texture. Namely, the extracted features are:

- Mouth Ratio
- Left Eye Ratio
- Right Eye Ratio
- Head ratio
- Texture of the forehead: Measurement of the changes of the texture of the forehead compared to 'neutral' expression
- Texture of the chin: Measurement of the changes of the texture of the chin compared to 'neutral' expression
- Texture of the region between the eyebrows: Measurement of the changes of the texture f the region between the eyebrows compared to 'neutral' expression

The above features form the resulting feature vector which is fed to the classifiers for training and testing as we describe in the next Section. The feature extraction process and systems results are analyzed and presented for various stages of the development of our system in [26], [27], [28], [29], [30], [31], [32].

2.3 Classifier description

Humans are able to classify facial expressions in a few seconds. How do we make sense of the expression of a face we look? Although this is a question that is studied by psychology researchers and cognitive scientists alike, there is no clear answer. What is known is that the ability is something we learn over time, there is a good deal of information present in a few seconds when we look in a face, we can decode and judge this information quickly, and our methods for judgment are not always introspectively accessible. This process is thought to be both highly parallel and to use multiple stages of filtering to extract higher level features from lower level features. Using the above ideas we compared the classification performance of following classifiers in facial expression classification: 1)**R**adial **B**asis **F**unctions neural networks, 2)**K**-th **N**earest **N**eighbour classifiers, 3)**S**upport **V**ector **M**achines and 4)**M**ultilayer **P**erceptron neural networks.

2.3.1 SVM classifier

The support vector machine (SVM) is a supervised classification system that finds an optimal hyperplane which separates data points that will generalize best to future data. Such a hyperplane is the so called maximum margin hyperplane, which maximizes the distance to the closest points from each class. Let $S = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ where $\mathbf{s}_j \in \mathbf{R^d}$ be a set of d-dimensional feature vectors corresponding to the image files of a facial expression database. Any hyperplane separating the two data classes has the form Eq. 1

$$f(\mathbf{s}) = \mathbf{w} \cdot \Phi(\mathbf{s}) + b \tag{1}$$

where $f : \mathbf{R^d} \rightarrow [-1, +1]$. The SVM classifier is obtained by solving a quadratic programming problem of the form:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{2n} \xi_i, \tag{2}$$

subject to the constraints

$$y_i (\mathbf{w} \cdot \Phi(\mathbf{s}_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 \forall i \in \{1, \ldots, 2n\}. \tag{3}$$

The optimal solution gives rise to a decision function of the following form:

$$f(\mathbf{s}) = \textstyle\sum_{i=1}^{2n} y_i w_i \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{s}_j) + b \tag{4}$$

A significant characteristic of SVMs is that only a small fraction of the w_i coefficients are non-zero. The corresponding pairs of $\mathbf{s}_i$ entries (known as margin support vectors) and y_i output labels fully define the decision function. Given that the training patterns appear only in dot product terms $\Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{s}_j)$, we can employ a positive definite kernel function $K(\mathbf{s}_i, \mathbf{s}_j) = \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{s}_j)$ to implicitly map into a higher dimensional space and compute the dot product. Specifically, in our approach we utilize the Gaussian kernel function which is of the form $K(\mathbf{s}_i, \mathbf{s}_j) = \exp\{-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{2\sigma^2}\}$.

2.3.2 RBF neural network classifier

Our system uses radial basis function (RBF) networks for facial expression classification. RBF networks have the advantages over other classifiers that they use initially unsupervised learning methods to find clusters of facial expressions without presupposed class labels. Then the RBF network distinguishes facial expression classes using the weights that are learned during training when the class labels for the samples are included. Also, RBF network can quickly classify new facial images of expressions once it has been trained. However, training can require a large amount of time because it traditionally involves finding good parameters for each basis function using gradient descent.

The input layer is determined by the dimensionality d of feature vector of each data point.Thus, the input to our RBF network is a vector $\mathbf{s}_j \in \mathbf{R^d}$

We will choose M basis functions for our network where each function computes the distance from $\mathbf{s}_i$ to a prototype vector $\mathbf{s}_j$. We use Gaussians for our basis functions: $K(\mathbf{s}_i, \mathbf{s}_j) = \exp\{-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{2\sigma_j^2}\}$. The parameters $\mathbf{s}_j$ and σ_j for each function are determined using unsupervised or supervised methods. So, the RBF network consisted of fifty basis functions (50 neurons) in the hidden layer.

The number of neurons in the output layer is determined by the number of classes we want to classify in each experiment. The equation for a single output

$$y_k(\mathbf{s}) = \sum_{i=1}^{M} w_{ki} \Phi_i(\mathbf{s}) + b \tag{5}$$

where $b = w_{k0}$ is the weight of the bias

The network was trained with the Expectation Maximization algorithm for two hundred (200) cycles and its output estimates the degree of membership of the input feature vector in each class. Thus, the value at each output necessarily remains between 0 and 1.

2.3.3 MLP neural network classifier

The Multi-layer Perceptron neural network which constructed in our system is a two feed-forward layered network. In this network the dimensionality of input layer is d, M hidden units and c output units. The output of the jth hidden unit is given by a weighted linear combination of the d input values:

$$a_j = \sum_{i=1}^{d} w_{ji}^{(1)} \mathbf{s}_i + b^{(1)} \tag{6}$$

where $w_{ji}^{(1)}$ denotes a weight in the first layer going from input i to hidden unit j and $b^{(1)}$ is the bias for the first layer. Similarly, the outputs for the second layer is given in the following form:

$$a_k = \sum_{i=1}^{M} w_{kj}^{(2)} \mathbf{z}_i + b^{(2)} \tag{7}$$

The activation of the kth output unit is obtained by transforming the linear combination using a non-linear activation function, to give:

$$y_k(\mathbf{s}) = \tilde{g}(a_k) \tag{8}$$

where g is the activation function.

In other words, an explicit expression for the complete function represented by our network is given in the form:

$$y_k(\mathbf{s}) = \tilde{g}\left(\sum_{i=1}^{M} w_{kj}^{(2)} g\left(\sum_{i=1}^{d} w_{ji}^{(1)} \mathbf{s}_i + b^{(1)}\right) + b^{(2)}\right) \tag{9}$$

The number of hidden units is five (5). The two-layer network with linear outputs is trained by minimizing a sum-of-squares error function using the scaled conjugate gradient optimizer.

2.3.4 KNN classifier

The KNN classifier was based on the class label prediction of the 10 nearest neighbours.

The NetLab toolbox was utilized in order to construct the RBF network, MLP, network and KNN classifiers, while the SVM classifier was implemented with the OSU-SVM toolbox. More details for the classifiers can be found in [33, 34, 35, 36].

3 Experimental performance evaluation

Classification results were calculated using 10-fold cross-validation evaluation, where the dataset to be evaluated was iteratively partitioned so that 90% be used for training and 10% be used for testing for each class. This process was iterated with different disjoint partitions and the results were averaged. This ensured that the calculated accuracy was not biased because of the particular partitioning of training and testing.

The results have shown that the SVM classifiers achieved higher results than the other three classifiers. The results presented in Table 1 illustrate the SVM classifier as the more adequate for this task. Also, by conducting empirical studies [1], we were able to measure the performance of human observers for the facial expression classification task. As we can observe, all the classifiers perform better than the human classifiers, results are also shown in Table 1

Table 1: Human versus computer classifiers

Classifiers	Accuracy for the five classes	Humans
MLP	81.52%	65.20%
RBF	92.72%	
KNN	94.8%	
SVM	96.00%	

In Table 2 we can observe the classification accuracy for each of the five classes depicting the ‘neutral’, ‘happiness’, ‘surprise’, ‘angry’ and ‘disgusted’ expression, respectively. The results of the four classifiers are in agreement with the results from the human responses. As we can observe in Table 2, the expressions corresponding to ‘angry’ and ‘disgusted’ achieved the lower success rates not only from the classifiers but also from the humans.

Table 2: Classification rates for each expression

Expressions	MLP	RBF	KNN	SVM	Human responses
Neutral	100%	100%	100%	100%	**80%**
Happy	94.80%	100%	100%	100%	**90%**
Surprised	83.60%	99.60%	99.60%	100%	**95%**
Angry	69.20%	75.20%	98.40%	94.40%	**55%**
Disgusted	60.00%	88.80%	76.00%	85.60%	**65%**

Based on these results we consider the SVM classifier as the more adequate classifier for this problem. Also, as we can observe from the results in Table 2, RBF classifier achieved the best classification rate in classifying the ‘disgusted’ expression. In the following Tables 3 and 4 we can see the confusion matrix for the 250 images for the SVM and RBF classifier, respectively.

Table 3: Confusion maxrix for the SVM Classifier

	Neutral	Happy	Surprised	Angry	Disgusted
Neutral	250	0	0	0	0
Happy	0	250	0	0	0
Surprised	0	0	250	0	0
Angry	0	0	0	236	14
Disgusted	0	0	0	36	214

The results in Tables 3 and 4, show that the misclassification for the ‘angry’ and ‘disgusted’ expressions are confided to these two expressions. The best results for the ‘disgusted’ expression are given from the RBF Classifier, where, as we can see

Table 4: Confusion maxrix for the RBF Classifier

	Neutral	**Happy**	**Surprised**	**Angry**	**Disgusted**
Neutral	250	0	0	0	0
Happy	0	250	0	0	0
Surprised	0	0	249	1	0
Angry	0	0	2	188	60
Disgusted	0	0	2	26	222

in Table 4, it manage to correctly classify 222 images of the 'disgusted' expression, while 26 images were misclassified as 'angry' and 2 images as 'surprised'. Despite this, we consider SVM Classifier as more adequate for this problem because it achieved the best results for all the expressions.

4 Conclusions

Automatic face detection and expression classification in images is a prerequisite in the development of novel human-computer interaction modalities. However, the development of integrated, fully operational such detection/classification systems is known to be non-trivial, a fact that was corroborated by our own statistical results regarding expression classification by humans. Towards building such systems, we developed four different classifiers, namely:(1) Multilayer perceptrons, (2) Radial basis networks, (3) K-nearest neighbor classifiers and, (4) Support vector machines. In this paper we make an evaluation of performance of these classifiers versus the human's expression recognition performance for five expression: 'neutral', 'happy', 'surprised', 'angry' and 'disgusted'. The results pointed out SVM Classifiers as the best choice for this kind of problem.

5 Future Work

In the future, we will extend this work in the following three directions: (1) we will improve our system by using wider training sets so as to cover a wider range of poses and cases of low quality of images. (2) We will investigate the need for classifying into more than the currently available facial expressions, so as to obtain more accurate estimates of a computer user's psychological state. In turn, this may require the extraction and tracing of additional facial points and corresponding features. (3) We plan to apply our system for the expansion of human-computer interaction techniques, such as those that arise in mobile telephony, in which the quality of the input images is too low for existing systems to operate reliably.

Another extension of the present work of longer term interest will address several problems of ambiguity concerning the emotional meaning of facial expressions by processing contextual information that a multi-modal human-computer interface may provide. For example, complementary research projects are being developed [39],[38], [37] that address the problem of emotion perception of users through their actions (mouse, keyboard, commands, system feedback) and through voice words. This and other related work will be presented on future occasions

6 Acknowledgement

This work has been sponsored by the General Secretary of Research and Technology of the Greek Ministry of Development as part of the PENED basic research program.

References

1. Stathopoulou, I.O., Tsihrintzis, G.: Facial Expression Classification: Specifying Requirements for an Automated System. In: Proceedings of the 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems, Bournemouth, United Kingdom (October 2006)
2. Cohn, J., Kanade, T.: Use of automated facial image analysis for measurement of emotion expression. In: Allen, J.B., Coan, J.A. (ed.) The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science (2006) (in press)
3. Goneid, A., e.K.R.: Facial feature analysis of spontaneous facial expression. In: Proceedings of the 10th International AI Applications Conference (2002)
4. Lien, J.J.J., Kanade, T., Cohn, J.F., Li, C.C.: Detection, tracking and classification of action units in facial expression. Journal of Robotics and Autonomous Systems
5. Cohen, I., Sebe, N., Chen, L., Garg, A., Huang, T.S.: Facial expression recognition from video sequences: Temporal and static modeling. Computer Vision and Image Understanding, 160–187 (2003)
6. Terzopoulos, D., Waters, K.: Analysis and synthesis of facial image sequences using physical and anatomical models. IEEE Trans. Pattern Anal. Mach. Intell. 15(6), 569–579 (1993)
7. Essa, I.A., Pentland, A.P.: Coding, analysis, interpretation, and recognition of facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. 19(7), 757–763 (1997)
8. Zhang, Z., Lyons, M., Schuster, M., Akamatsu, S.: Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In: FG 1998: Proceedings of the 3rd International Conference on Face & Gesture Recognition, Washington, DC, USA, p. 454. IEEE Computer Society, Los Alamitos (1998)
9. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surv. 35(4), 399–458 (2003)
10. Bartlett, M.S., Littlewort, G., Fasel, I., Movellan, J.R.: Real time face detection and facial expression recognition: Development and application to human computer interaction. In: CVPR Workshop on CVPR for HCI, pp. 139–157 (2003)

11. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: FG 1998: Proceedings of the 3rd International Conference on Face & Gesture Recognition, Washington, DC, USA, p. 200. IEEE Computer Society, Los Alamitos (1998)
12. Black, M.J., Yacoob, Y.: Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In: ICCV, pp. 374–381 (1995)
13. Lisetti, C.L., Schiano, D.J.: Automatic facial expression interpretation: Where human-computer interaction, ai and cognitive science intersect. Pragmatics and Cognition 8(1), 185–235 (2000)
14. Dailey, M.N., Cottrell, G.W., Adolphs, R.: A six-unit network is all you need to discover happiness. In: TwentySecond Annual Conference of the Cognitive Science Society, pp. 101–106. Erlbaum, Mahwah (2000)
15. Rosenblum, Y.Y., Davis, L.: Human expression recognition from motion using a radial basis function network architecture. IEEE Transactions on Neural Networks 7(5), 1121–1138 (1996)
16. Kotsia, I., Buciu, I., Pitas, I.: An analysis of facial expression recognition under partial facial image occlusion. Image Vision Comput. 26(7), 1052–1067 (2008)
17. Hernández, B., Olague, G., Hammoud, R., Trujillo, L., Romero, E.: Visual learning of texture descriptors for facial expression recognition in thermal imagery. Comput. Vis. Image Underst. 106(2-3), 258–269 (2007)
18. Hammal, Z., Couvreur, L., Caplier, A., Rombaut, M.: Facial expression classification: An approach based on the fusion of facial deformations using the transferable belief model. Int. J. Approx. Reasoning 46(3), 542–567 (2007)
19. Kim, D.H., Jung, S.U., Chung, M.J.: Extension of cascaded simple feature based face detection to facial expression recognition. Pattern Recogn. Lett. 29(11), 1621–1631 (2008)
20. Liang, D., Yang, J., Zheng, Z., Chang, Y.: A facial expression recognition system based on supervised locally linear embedding. Pattern Recogn. Lett. 26(15), 2374–2389 (2005)
21. Xiang, T., Leung, M.K.H., Cho, S.Y.: Expression recognition using fuzzy spatio-temporal modeling. Pattern Recogn. 41(1), 204–216 (2008)
22. Pantic, M., Rothkrantz, L.: Expert system for automatic analysis of Facial Expression. Image and Vision Computing Journal 18(11), 881–905 (2000)
23. Chung-Lin Huang, Y.M.H.: Facial expression recognition using model-based feature extraction and action parameters classification. Journal of Visual Communication and Image Representation 8(3), 278–290 (1997)
24. Ekman, P., Friesen, W.V.: Manual for the Facial Action Coding System. Consulting Psychologists Press, Palo Alto (1977)
25. Pantic, M., Rothkrantz, L.: Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12), 1424–1445 (2000)
26. Stathopoulou, I.O., Tsihrintzis, G.: A neural network-based facial analysis system. In: Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services, Lisboa, Portugal (April 2004)
27. Stathopoulou, I.O., Tsihrintzis, G.: An Improved Neural Network-Based Face Detection and Facial Expression Classification System. In: IEEE International Conference on Systems, Man, and Cybernetics, The Hague, Netherlands (October 2004)
28. Stathopoulou, I.O., Tsihrintzis, G.: Pre-processing and expression classification in low quality face images. In: Proceedings of 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services (July 2005)

29. Stathopoulou, I.O., Tsihrintzis, G.: Evaluation of the Discrimination Power of Features Extracted from 2-D and 3-D Facial Images for Facial Expression Analysis. In: Proceedings of the 13th European Signal Processing Conference, Antalya, Turkey (September 2005)
30. Stathopoulou, I.O., Tsihrintzis, G.: Detection and Expression Classification Systems for Face Images (FADECS). In: Proceedings of the IEEE Workshop on Signal Processing Systems (SiPS 2005), Athens, Greece (November 2005)
31. Stathopoulou, I.O., Tsihrintzis, G.: An Accurate Method for eye detection and feature extraction in face color images. In: Proceedings of the 13th International Conference on Signals, Systems, and Image Processing, Budapest, Hungary (September 2006)
32. Stathopoulou, I.O., Tsihrintzis, G.: NEU-FACES: A Neural Network-based Face Image Analysis System. In: Proceedings of the 8th International Conference on Adaptive and Natural Computing Systems, Warsaw, Poland (April 2007)
33. Bishop, C.M.: Neural Networks for Pattern Recognition, 1st edn. Oxford University Press, Oxford (1995)
34. Haykin, S.: Neural Networks, 2nd edn. Prentice Hall, Englewood Cliffs (1999)
35. Duda, R.O., Hart, P.E., Strock, D.G.: Pattern Classification, 2nd edn. John Wiley, Chichester (2000)
36. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 1st edn. Academic Press, London (1998)
37. Alepis, E., Virvou, M., Kabassi, K.: Knowledge Engineering Aspects of Affective Bi-Modal Educational Applications. In: E-business and Telecommunications, Book Series CCIS, vol. 23, pp. 328–339 (2008)
38. Alepis, E., Virvou, M., Kabassi, K.: Affective reasoning based on bi-modal interaction and user stereotypes. Studies in Computational Intelligence 142, 523–532 (2008)
39. Alepis, E., Virvou, M., Kabassi, K.: Development process of an affective bi-modal Intelligent Tutoring System. Intelligent Decision Technologies 142, 117–126 (2007)

Human-Computer Interaction Approach Applied to the Multimedia System of Polysensory Integration

Lech Michal and Kostek Bozena

Abstract In the paper an approach of utilizing an interaction between the human and computer in a therapy of dyslexia and other sensory disorders is presented. Bakker's neuropsychological concept of dyslexia along with therapy methods are reviewed in the context of the Multimedia System of Polysensory Integration, proposed at the Multimedia Systems Department of Gdansk Univ. of Technology. The system is presented along with the training methods proposed. Three therapeutic exercises, stimulating respectively left hemisphere only, right hemisphere only, and both hemispheres simultaneously are described. Their utility in a therapy of disorders such as dyslexia, ADD/ADHD (Attention Deficit Disorder/Attention Deficit Hyperactivity Disorder) and lateralization problems is discussed in the context of human brain functioning. Also, their implementation process in the Multimedia System of Polysensory Integration is provided and a sample of tests carried out in a primary school is described.

1 Introduction

HCI (*Human-Computer Interaction*) is defined by Hewett, Baecker, Card, Carey, Gasen, Mantei, Perlman, Strong and Verplank as "a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" [12]. In this context one may think about creating. Certainly, this definition is mainly associated with designing usable, useful, and intuitive interfaces of communication between the human and computer, however it seems that other concepts may meet these requirements as well. The paper presents a different approach to HCI. The interaction between human and computer has been treated here as an innovative tool for a therapy of sensory disorders, especially dyslexia. High emphasis has been placed on utilizing

Multimedia Systems Dept., Gdansk University of Technology, Gabriela Narutowicza 11/12, 80-952 Gdansk, Poland, e-mail: mlech,bozenka@sound.eti.pg.gda.pl

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 265–274.
springerlink.com

various multimedia components to attract person's attention and on designing an adaptable graphical user interface taking into account functioning of the brain of a person with a particular disorder.

Although dyslexia problems have been known for more than one hundred years still majority of schools of any degree do not use educational methods meeting needs of dyslexic persons, or their teachers even neglect existence of this disorder. Because of the fact that long-lasting learning using typical methods does not lead to any improvement, dyslexic pupils are often treated as less intelligent. In fact, their brain is able to work 400 to 2000 times faster than the brain of person without dyslexia and can handle 32 mental images per 1 second while verbally thinking person can have only 2 to 5 thoughts constituting single words (in dyslexia of type P) [7]. This creates a great potential when information is prepared in such a way that it can be processed with understanding. Unfortunately, using teaching methods aimed at dyslexic pupils' needs is still not well corresponding to actual demand in schools. In such a state of affairs improving methods of dyslexia therapy and methods preventing from this disorder at the earliest stage of education seems to be the best way to help dyslexic persons. In this context, in the age of highly developed multimedia technologies, HCI can be seen as the basis for developing potentially highly profitable therapeutic system. In the paper the system developed for a therapy of dyslexia of types L and P, and of the mixed type is presented. The system can be also used for a therapy of lateralization problems, ADD/ADHD (*Attention Deficit Disorder/Attention Deficit Hyperactivity Disorder*) and for general development of a child.

2 Types of dyslexia and therapy

The basis for the development of the Multimedia System of Polysensory Integration was the Dirk Bakker's neuropsychological conception of dyslexia [1]. The conception connects particular dyslexia disorders with anatomical structure and physiology of brain and its functioning in the context of cognitive processes [2] [3][4]. Bakker distinguishes between two types of dyslexia, type L (linguistic) and P (perceptual), depending on which hemisphere should dominate and actually dominates when performing the particular task [2][4]. In dyslexia of type P the right hemisphere dominates over the left one. Person with such a disorder bases on perceptual characteristics of written texts. An exception is a preliminary phase of learning to read, i.e. elementary reading, when letters are new for a person and do not have meaning. Elementary reading should be based on right hemisphere strategies. As a person learns to read and letters gradually obtain meaning, right hemisphere strategies should be displaced by the left ones. The moment of passing from using right hemisphere to left hemisphere strategies in reading processes occurs at the age of 8. If after this period a child still uses right hemisphere strategies this constitutes the bases for diagnosing dyslexia of type P or risk of having dyslexia of type P in later years [2][4][5].

In dyslexia of type P reading is slow, with pauses, syllabication and without understanding. A therapy for this disorder, regarding Bakker's neurological conception, is based on exercises stimulating left hemisphere. Examples of exercises distinguished by Bakker as HSS (*Hemisphere Specific Stimulation*) and HAS (*Hemisphere Alluding Stimulation*) methods have been described in next two sections. In dyslexia of type L the right hemisphere is taken over by the left one. This kind of dyslexia can be observed at early stages of learning to read. Persons suffering from such a disorder start using left-hemisphere strategies of reading before they actually learn to recognize letter symbols properly. Reading is fast but with a lot of mistakes and also without understanding [1]. A therapy of dyslexia of type L bases on exercises stimulating right hemisphere. Examples of exercises for L-dyslexics have been described in next two sections. Apart from Bakker's neuropsychological conception of dyslexia also dyslexia of the mixed type can be discerned. In such a disorder both hemispheres are underdeveloped and need stimulation [6].

2.1 HSS

HSS (*Hemisphere Specific Stimulation*) stimulation is based on addressing stimulus directly to particular hemisphere using three sensory modalities, i.e. sight, hearing, and sensorial or motor activity, individually or simultaneously. As an example, a situation in which a sound stimulus reaches left ear and in consequence stimulates right hemisphere mainly can be mentioned [1].
In a therapy of L-dyslexia a training based on recognizing letters, made of various materials like wood, plastic or sponge, by touching them with left hand and having eyes closed can be given as an example. Letters in such a therapy are of complex shapes and are often to be used by a child to create untypical and abstract concepts [1]. For P-dyslexics one can mention an exercise based on reading or spelling words loudly, written with a simple font, presented in the right visual field. Simultaneously voices of a therapist and a child are audible in the right headphone [1].

2.2 HAS

The aim of the HAS method (*Hemisphere Alluding Stimulation*) is to activate particular hemisphere not by its direct stimulation but by preparing stimulating material in such a way that its processing forces one of the hemispheres activation. As an example a therapy during which a child is asked to read perceptually complex texts, i.e. containing various font styles, colors and sizes, can be mentioned. Although reading activates both hemispheres, such preparation of the text materials employs necessity of symbol recognition and in consequence involves right hemisphere in a greater degree [1].

The above-mentioned exercise is used in a therapy of dyslexia of type L. In Fig. 1 an example of an appropriately prepared material for such a task is presented.

Fig. 1: Text appropriately prepared for HAS stimulation in a therapy of dyslexia of type L.

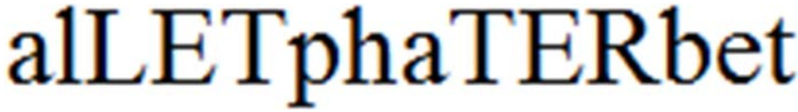


Fig. 2: Exercise stimulating the right hemisphere in HAS therapy of dyslexia of type L.

In a therapy of dyslexia of type P stimulating materials are very simple, i.e. font is easily readable, style, color and size are consistent with the whole content. Typical exercises associated with this kind of therapy involve ability of solving various linguistic tasks, such as creating a sentence from words written in random order, creating rhymes for given words or choosing the word which does not fit to the set of words.

3 Multimedia system of polysensory integration

The Multimedia System of Polysensory Integration bases both on HSS and HAS therapy methods. The system integrates various sensory modalities, i.e. visual, hearing, and motor activity and as such provides polysensory training.

3.1 Equipment

The platform for the training consists of the following equipment: a part of the floor designated for tests, called therapeutic pad in further parts of this paper (3x3 squares), four speakers, five computer monitors, two USB cameras, a PC computer with two-output graphic card, and video splitter (1 input, 4 outputs). The view of the therapeutic pad positions and speaker placement has been presented in Fig. 3.

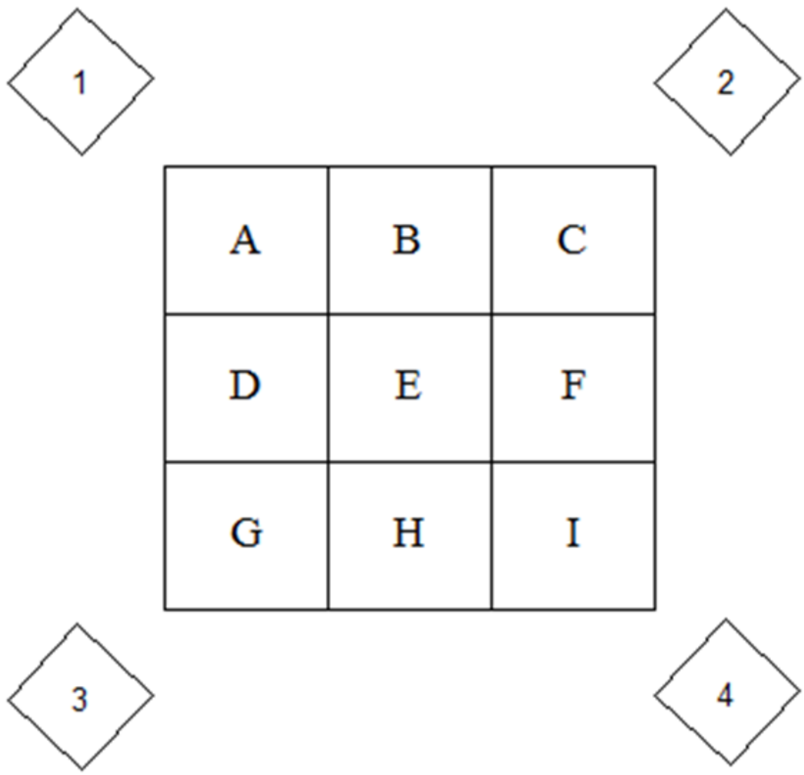


Fig. 3: Therapeutic pad positions and speaker placement.

3.2 Training methods

The Multimedia System of Polysensory Integration contains three therapeutic polysensory training schemes, stimulating the particular hemisphere or both hemispheres, intended for therapy of various disorders.

3.2.1 Metronome training for a therapy of dyslexia of type P

A child takes up a middle position of the therapeutic pad. Sequences (bars) of various numbers of metronome ticks are generated. The first sequence (e.g. four ticks) of the metronome ticks is being played and simultaneously corresponding sequence of shapes (e.g. rectangles) is displayed for a child's on the screen monitor (Fig. 3). One of the shapes is different from the others. The task of a child is to count the metronome ticks and analyze the displayed sequence, and bounce when the tick associated with the highlighted shape is played. The bounce should be made in synchronization with the metronome tick. After bouncing, the child gets graphical information about its precision (Fig. 3) and also gets the audio feedback. If the child bounces too early, the audio feedback mechanism returns the signal from the left side. Else, the audio feedback is given from the right side.

Fig. 4: A graphical user interface for a child, used with training for a therapy of type P dyslexia or ADD/ADHD.

The training is mainly intended for therapy of ADD/ADHD and dyslexia of type P. In such a type of dyslexia the right hemisphere takes over the left one [2][4]. To support the desired lateralization and proper cooperation of both hemispheres the left one needs to be activated and this constitutes the purpose of the training. Listening to the metronome ticks, tracking the rectangles on the screen, and synchronizing a jump according to them activates this particular hemisphere as it is responsible for time sequencing and motor control [9][10][11]. Tick sounds are given from the right side of the child, thus as they reach the right child's ear they activate the left hemisphere in the first place [1]. The audio feedback is given from the corresponding side depending on the fact whether a child performs a jump earlier than the particular metronome tick or later. If a jump is performed too early, the audio feedback is given from the left side and activates the right hemisphere. In such a case, the early jump may be performed because the left hemisphere, which is quick and inaccurate, had taken over the right hemisphere. That is the reason why the right hemisphere needs to be stimulated. In the case of a late jump the situation is reversed. Shapes displayed on the screen are very simple, thus they do not activate right hemisphere as perceptually complex shapes do [1].

3.2.2 Metronome training for a therapy of dyslexia of the mixed type

The training in case of dyslexia of the mixed type is the same regarding necessity of analyzing metronome and graphical sequences by the child. The difference is that the task of the child is not to bounce but jump on the appropriate position of the therapeutic pad (changing a position). The position on which the jump should be performed is determined by recognizing a direction (speaker) from which the metronome sequence is being played. A jump is performed on a neighboring position in this direction. After performing a jump the child gets, just like in the training for a therapy of dyslexia P, graphical information about a jump precision (Fig. 5) and also gets the audio feedback. Additionally, information about jump's direction is given (Fig. 5). The training is mainly intended for a therapy of dyslexia of type M (mix) and therapy of lateralization problems. The purpose of the training is to stimulate simultaneously both hemispheres. The stimulation of the left hemisphere is provided by the same means as in the therapeutic training for dyslexia P. Metronome sounds are played alternately from the left and right sides of the child so both hemispheres are stimulated alternately. Right hemisphere is activated by the necessity of recognizing a direction from which the metronome sounds come and jumping in the corresponding direction. Such a task is associated with a spatial orientation for which the right hemisphere is mainly responsible [1][5].

3.2.3 Letter-mix training (for the therapy of dyslexia of type L)

Before the training begins, the child takes up a position in the middle of the therapeutic pad, a part of the floor designated for the training. A letter mix created using various font styles and colors, e.g. "GiTIraGERfE", is displayed on the monitor screen in front of the child. The letter mix contains syllables composing words naming animals. The syllables are mixed in a manner presented in the example above. There is one word consisting of letters of the same font style and color contained in the letter mix. The first task of the child is to find this word in the letter mix. In the given example the proper word is: "TIGER". There are other sensible words contained in the letter mix, e.g. word "GiraffE" in the sample letter mix, to make the task more difficult. Other words than the one to be found is written using inconsistent font style or color thus the child knows that this is not the word of interest. With each position of the therapeutic pad there is one chosen animal associated. One of the positions must be associated with the animal that is assigned to the word to be

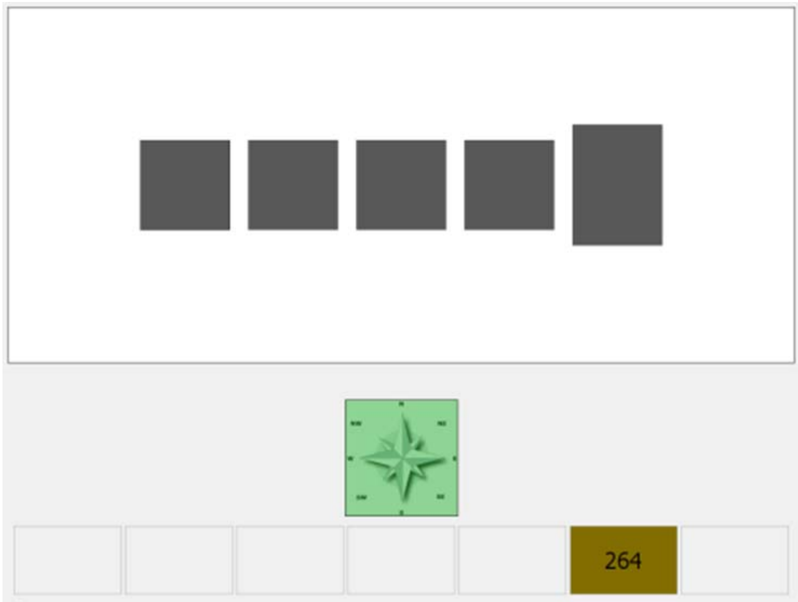


Fig. 5: The graphical user interface for a child, used with training for a therapy of dyslexia of the mixed type.

found in the letter mix. When a child changes the position walking on the floor, an animal assigned to the occupied position is displayed on the monitor screen. The task of the child is to find the position to which the animal associated with the appropriate word contained in the letter mix is assigned, by walking on the floor and memorizing associations between positions and displayed animals, like in “Memory” game. A child signalizes that he/she has found the right position (animal) by a bounce. Audio feedback posing animal sound is given from the speaker on the left side of the child. If a bounce is performed on the proper position a green label with the name of the animal is displayed and a new letter mix is created, new set of animals is assigned to the positions of the therapeutic pad, and the training continues. If a bounce is performed on the position to which the animal not corresponding to the word found in the letter mix is assigned, red label with the name of the animal is displayed and the set of position-animal associations stays the same until the child finds the correct position. The training is mainly intended for a therapy of dyslexia of type L and can be also used for general development of a child up to age of 8. It employs two senses: seeing and hearing and together with involving motor activity it poses highly profitable in treatment of sensory disorders polysensory training [1]. Complex colorful font of the letter mix, displayed in the left side of the screen to address the left field of vision, activates mainly the right hemisphere while searching for the word [1][8]. This hemisphere is also activated by looking at images of animals and necessity of remembering where on the therapeutic pad is a place with the

assigned particular animal (spatial orientation) [1][5][9][11]. Playing animal sounds from speakers on the left side of the child additionally stimulates right hemisphere [1].
The prototype version of the system, having functionality completely covering training methods described, was tested in one of primary schools in Gdansk, Poland (Fig. 6). Both, the letter-mix training and the metronome training reliability, measured as the number of feedback information signals adequate to the child's movements per all feedback information signals sent during one training session, were up to level of 80%. It was discovered during the tests that stability of results depends highly on the system calibration and proper camera placement in the therapeutic environment. Calibrating the system appropriately for a particular child instead of setting it to the fixed value (like during the tests when threshold was set basing on the author's own jumps) can be a first step to increase system reliability.

Fig. 6: Tests of the Multimedia System of Polysensory Integration in the primary school.

4 Conclusions

The Multimedia System of Polysensory Integration was proposed and discussed in the context of human-computer interaction. Three therapeutic methods have been implemented and some preliminary tests were carried out in a primary school. Children's reactions regarding the Multimedia System of Polysensory Integration were very spontaneous. They executed therapeutic commands fast, with pleasure and a complete understanding. They also expressed a desire for continuing the training after the tests had been completed.

The children's approval of the system along with the results of research on polysensory training potentially makes the Multimedia System of Polysensory Integration a very powerful tool for helping dyslexics, especially taking into account the number of sensory modalities that can be simultaneously used during the training. The system may also serve as a tool for general development of a child.

Acknowledgments

The research work is partially supported by the Polish Ministry of Science and Higher Education within the project entitled: "Elaboration of a series of computer multimodal interfaces and their implementation to education, medicine, defence and to industry" (POIG 1.3).

References

1. Bakker, D.J.: Neuropsychological Treatment of Dyslexia. Oxford University Press, Oxford (1990)
2. Bakker, D.J., Moerland, R., Goekoop-Hoefkens, M.: Effects of hemisphere - specific stimulation on the reading performance of dyslexics boys: A pilot study. Journal of Clinical Psychology 3, 155–159 (1981/1990)
3. Bakker, D.J., Satz, P.: Specific reading disability. In: Advances in theory and method. Rotterdam University Press, Rotterdam (1970)
4. Bakker, D.J., Vinke, J.: Effects of hemisphere - specific stimulation on brain activity and reading in dyslexics. Journal of Clinical and Experimental Neuropsychology 7, 505–525 (1985)
5. Bates, E., Roe, K.: Language Development in Children with Unilateral Brain Injury. In: Nelson, C., Luciana, M. (eds.) Handbook of Development Cognitive Neuroscience. The MIT Press, Cambridge (2001)
6. Brunsdon, R., Hannah, T., Nickels, L., Coltheart, M.: Successful treatment of sublexical reading deficits in a child with dyslexia of the mixed type, journal article. Neuropsychological rehabilitation 12(3), 199–229 (2002)
7. Davis, R., Braun, E.: The Gift of Dyslexia. In: Why Some of the Smartest People Can't Read and How They Can Learn. Perigee Books (1997)
8. De Haan, M.: Development of face recognition. In: Nelson, C., Luciana, M. (eds.) Handbook of Development Cognitive Neuroscience. The MIT Press, Cambridge (2001)
9. Gazzaniga, M., Hutsler, J.: Hemispheric Specialization. In: Wilson, R., Keil, F. (eds.) The MIT Encyclopedia of the Cognitive Sciences, pp. 369–370. The MIT Press, Cambridge (1999)
10. Kujala, T.: Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. PNAS 98(18) (2001)
11. Leckman, J., Peterson, B., Schultz, R., Cohen, D.: Neuropsychological findings implicate brain regions involved in visual-motor integration. In: Nelson, C., Luciana, M. (eds.) Handbook of Development Cognitive Neuroscience. The MIT Press, Cambridge (2001)
12. ACM SIGCHI - Association for Computing Machinery Special Interest Group on Computer-Human Interaction, `http://sigchi.org/cdg/cdg2.html#2_1` (last updated April 11, 2008)

Fuzzy Simple Additive Weighting for Evaluating a Personalised Geographical Information System

Katerina Kabassi

Abstract This paper presents how a fuzzy multi-criteria decision making theory has been used for evaluating personalised software. More specifically, the fuzzy simple additive weighting theory has been used for the evaluation of a Geographical Information System. The Geographical Information System has the ability to process information about its users in order to adapt its interaction to each user dynamically. The proposed system has been evaluated in comparison to a standard GIS using linguistic terms. These terms were further translated to fuzzy numbers in order to calculate a crisp value for each GIS aggregating all the criteria taking into account. The comparison of the crisp value for the two GISs revealed that the personalised GIS is friendlier and more useful than the standard one.

1 Introduction

Geographical Information Systems have been used lately extensively in many different domains, e.g. economical and regional development, environmental management, tourism sector etc. However, a main problem of such systems is that not many users have the specialised knowledge that is required to use them. Furthermore, the increasing complexity of the world and information overload makes it nearly impossible for a Geographical Information System (GIS) to be able to address the needs of large and diverse user populations. As a result, lately, a lot of research energy has been put on developing systems that are user friendly and can provide personalised interaction.

Therefore, several AI methods and approaches such as Bayesian Networks, multi-criteria decision making and fuzzy logic have been used for personalising interaction in a GIS [1, 2, 3, 4]. In view of the above, a Geographical Information System called ADAPTIGIS (Adaptive GIS) was developed [5]. The particular system contains data

Department of Ecology and the Environment, Technological Educational Institution of Ionian Islands, 2 Kalvou Sq., 29100 Zakynthos, Greece e-mail: kkabassi@teiion.gr

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 275–284.
springerlink.com

about the physical and anthropogenic environment of Zakynthos, an island of Greece. More specifically, ADAPTIGIS has the ability to process information about the users so that the system can adapt its interaction to each user dynamically. The information is processed using a multi-criteria decision making theory [6].

However, the incorporation of some kind of theory does not guarantee the successful reasoning of the system. Mc Tear [7] points out that the relationship between theory and practice is particularly important in Intelligent Interface Technology as the ultimate proof of concept here is that the interface actually works and that it is acceptable to users; for this reason practical issues such as performance, reliability and usability would seem to be more important than theoretical issues such as choice of system design methodology or specification notations.

Indeed, Chin [8] points out that empirical evaluations are needed to determine which users are helped or hindered by user-adapted interaction in user modelling systems. He adds that the key to good empirical evaluation is the proper design and execution of the experiments so that the particular factors to be tested can be easily separated from other confounding factors. However, he notes that empirical evaluations are not so common in the user modelling literature. Many researchers have also stated that it is important when evaluating adaptive systems to assess whether the system works better with the user modelling component as opposed to a system deprived of this component [8, 9, 10].

In view of the above, ADAPTIGIS was evaluated in comparison to a standard GIS that does not incorporate any intelligence using Fuzzy Simple Additive Weighting (FSAW) [11]. Indeed, the evaluation of software involves several criteria and, therefore, can be considered as a multi-criteria problem. Especially, the FSAW theory was selected because it is rather simple and has the advantage of allowing multi-criteria problems to accommodate linguistic terms represented as fuzzy numbers. This facilitates the creation of a decision procedure that is more realistic than other existing theories [11].

2 Fuzzy Simple Additive Weighting

Zadeh [12] pioneered the use of Fuzzy Set Theory (FST) to address problems involving fuzzy phenomena. In a universe of discourse X, a fuzzy subset $\widetilde{A}$ of X is defined with a membership function $\mu_{\widetilde{\alpha}}(x)$ that maps each element x in X to a real number in the interval $[0,1]$. The function value of $\mu_{\widetilde{\alpha}}(x)$ signifies the grade of membership of x in $\widetilde{A}$. When $\mu_{\widetilde{\alpha}}(x)$ is large, its grade of membership of x in $\widetilde{A}$ is strong [13]. A fuzzy set $\widetilde{A} = (a,b,c,d)$ on R, $a < b < c < d$, is called a trapezoidal fuzzy number if its membership function is

$$\mu_{\widetilde{\alpha}}(x) = \begin{cases} \frac{(x-a)}{(b-a)}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{(x-d)}{(c-d)}, & c \leq x \leq d \\ 0, & \text{otherwise} \end{cases}$$

where a, b, c, d are real numbers [13, 14]. Trapezoidal fuzzy numbers are the most widely used forms of fuzzy numbers because they can be handled arithmetically and interpreted intuitively. The FSAWS procedure based on above conceptual model is as follows:

Step 1: Form a committee of decision-makers. Choose the attributes and identify the prospective alternatives. A committee of decision-makers is formed to determine the most appropriate alternative.

Step 2: Determine the degree of importance (or reliability) of the decision-makers. If the degrees of importance (or reliability) of decision-makers are equal, then the group of decision-makers is deemed a homogeneous group $I_1 = I_2 = \cdots = I_k = \frac{1}{k}$; otherwise, the group of decision-makers is called a heterogeneous (non-homogeneous) group.

Step 3: Introduce linguistic weighting variables (Table 1) for decision-makers to assess attributes importance, and compute aggregated fuzzy weights of individual attributes. Let $\widetilde{W}_{jt} = (a_{jt}, b_{jt}, c_{jt}, d_{jt}),\ j = 1, 2, \ldots, n;\ t = 1, 2, \ldots, k$ be the linguistic weight given to subjective attributes $C_1, C_2, \ldots, C_h$ and objective attributes $C_{h+1}, C_{h+1}, \ldots, C_n$ by decision-maker D_t. The aggregated fuzzy attribute weight, $\widetilde{W}_j = (a_j, b_j, c_j, d_j),\ j = 1, 2, \ldots, n$ of attribute C_j assessed by the committee of k decision makers is defined as $\widetilde{W}_j = (I_1 \otimes \widetilde{W}_{j1}) \oplus (I_2 \otimes \widetilde{W}_{j2}) \oplus \cdots \oplus I_{1k} \otimes \widetilde{W}_{jk})$, where $a_j = \sum_{t=1}^{k} I_t a_{jt}$, $b_j = \sum_{t=1}^{k} I_t b_{jt}$, $c_j = \sum_{t=1}^{k} I_t c_{jt}$, $d_j = \sum_{t=1}^{k} I_t d_{jt}$.

Step 4: Defuzzify the fuzzy weights of individual attributes; compute the normalized weights and construct the weight vector.

To defuzzify the weights of the fuzzy attributes, the signed distance is adopted. The defuzzification of $\widetilde{W}_j$, denoted as $d(\widetilde{W}_j)$ is therefore given by

$$d(\widetilde{W}_j) = \frac{1}{4}(a_j + b_j + c_j + d_j),\ j = 1, 2, \ldots, n$$

The crisp value of the normalized weight for attribute C_j denoted as W_j, is given by:

$$W_j = \frac{d(\widetilde{W}_j)}{\sum_{j=1}^{n} d(\widetilde{W}_j)},\ j = 1, 2, \ldots, n$$

where $\sum_{j=1}^{n} W_j = 1$.The weight vector $W = [W_1, W_2, \ldots, W_n]$ is therefore formed.

Step 5: Use linguistic rating variables (Table 2) for decision-makers to assess fuzzy ratings of alternatives with respect to individual subjective attributes, and then pool them to obtain the aggregated fuzzy ratings. Let $\widetilde{x}_{ijt} = (o_{ijt}, p_{ijt}, q_{ijt}, s_{ijt})$, $i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, h,\ t = 1, 2, \ldots, k$ be the linguistic suitability rating assigned to alternative location A_i for subjective attribute C_j by decision-maker D_t. Let us further define $\widetilde{x}_{ijt}$ as the aggregated fuzzy rating of alternative A_i for subjective attribute C_j, such that

$$\widetilde{x}_{ij} = (I_1 \otimes \widetilde{x}_{ij1}) \oplus (I_2 \otimes \widetilde{x}_{ij2}) \oplus \cdots \oplus I_{1k} \otimes \widetilde{x}_{ijk})$$

which can subsequently be represented and computed as

$$\widetilde{x}_{ij} = (o_{ij}, p_{ij}, q_{ij}, s_{ij}),\ i = 1,2,\ldots,m,\ j = 1,2,\ldots,h$$

where $o_{ij} = \sum_{t=1}^{k} I_t o_{ijt}$, $p_{ij} = \sum_{t=1}^{k} I_t p_{ijt}$, $q_{ij} = \sum_{t=1}^{k} I_t q_{ijt}$, $s_{ij} = \sum_{t=1}^{k} I_t s_{ijt}$.
Step 6: Construct a fuzzy rating matrix based on fuzzy ratings. The fuzzy rating matrix $\widetilde{M}$ can be concisely expressed in matrix format

$$\widetilde{M} = \begin{bmatrix} \widetilde{x}_{11} & \widetilde{x}_{12} & \cdots & \widetilde{x}_{1n} \\ \widetilde{x}_{21} & \widetilde{x}_{22} & \cdots & \widetilde{x}_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & & \cdot \\ \widetilde{x}_{m1} & \widetilde{x}_{m2} & \cdots & \widetilde{x}_{mn} \end{bmatrix}$$

where $\widetilde{x}_{ij}, \forall i, j$ is the aggregated fuzzy rating of alternative A_i, $i = 1,2,\ldots,m$ with respect to attribute C_j.
Step 8: Derive total fuzzy scores for individual alternatives by multiplying the fuzzy rating matrix by their respective weight vectors.

Obtained total fuzzy score vector by multiplying the fuzzy rating matrix $\widetilde{M}$ by the corresponding weight vector W, i.e.,

$$\widetilde{F} = \widetilde{M} \otimes W^T = \begin{bmatrix} \widetilde{x}_{11} & \widetilde{x}_{12} & \cdots & \widetilde{x}_{1n} \\ \widetilde{x}_{21} & \widetilde{x}_{22} & \cdots & \widetilde{x}_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & & \cdot \\ \widetilde{x}_{m1} & \widetilde{x}_{m2} & \cdots & \widetilde{x}_{mn} \end{bmatrix} \otimes \begin{bmatrix} W_1 \\ W_2 \\ \cdot \\ \cdot \\ \cdot \\ W_n \end{bmatrix} = \begin{bmatrix} \widetilde{x}_{11} \otimes W_1 \oplus \widetilde{x}_{12} \otimes W_2 \oplus \ldots \widetilde{x}_{1n} \otimes W_n \\ \widetilde{x}_{12} \otimes W_1 \oplus \widetilde{x}_{22} \otimes W_2 \oplus \ldots \widetilde{x}_{2n} \otimes W_n \\ \cdot \\ \cdot \\ \cdot \\ \widetilde{x}_{m1} \otimes W_1 \oplus \widetilde{x}_{m2} \otimes W_2 \oplus \ldots \widetilde{x}_{mn} \otimes W_n \end{bmatrix} = \begin{bmatrix} \widetilde{f}_1 \\ \widetilde{f}_2 \\ \cdot \\ \cdot \\ \cdot \\ \widetilde{f}_n \end{bmatrix} = \left[\widetilde{f}_i\right]_{m*1}$$

where $\widetilde{f}_i = (r_i, s_i, t_i, u_i)$, $i = 1,2,\ldots,m$.
Step 9: Compute a crisp value for each total score using a defuzzification method and select the alternative(s) with the maximum total score.
Rank total fuzzy scores $\widetilde{f}_1, \widetilde{f}_2, \ldots, \widetilde{f}_m$ by the signed distance to determine the best location. Determine crisp total scores of individual locations by the following defuzzification equation:

$$d(\widetilde{f}_i) = \frac{1}{4}(r_i + s_i + t_i + u_i),\ i = 1,2,\ldots,m$$

where $d(\widetilde{f}_i)$ gives the defuzzified value (crisp value) of the total fuzzy score of location A_i by using the signed distance. The ranking of the locations can then be preceded with the above crisp value of the total scores for individual alternatives.

3 Geographical Information System

ADAPTIGIS is a Geographical Information System that contains data about the physical and anthropogenic environment of Zakynthos, an island of Greece [5, 6]. The particular island has great touristic as well as ecological interest due to Lagana Bay, where the turtles Caretta-Caretta live and breed. For this purpose, the information that is maintained in such a GIS would be of interest to a great variety of users. However, different kinds of users may have different interests, needs and background knowledge. For example, tourists and/or residents of the islands would prefer to find information about roads, summer resorts and cultural information, e.g. monuments and churches. Environmentalists, ecologists and researchers, on the other hand, may seek low/high resolution satellite data, which are used for the estimation, charting, characterization and classification of various environmental parameters such as land cover/usage, geomorphological features, etc. In view of the above, the main characteristic of ADAPTIGIS is that it can adapt its interaction with each individual user. In order to adapt the information provided to the interests and background knowledge of each user interacting with ADAPTIGIS, the system incorporates a user modelling component. This component maintains information about the interests, needs and background knowledge of all categories of potential users. The information that is collected for every category of users has been based on the analysis of the results of the empirical study [6].

More specifically, the potential users of the GIS were divided into five stereotypes in accordance to the five main categories of users that were identified: Residents of the islands, Tourists, Local authorities of Zakynthos, Environmentalists / Researchers, Students of the department of ecology and the environment in the Technological Educational Institution of the Ionian islands, which is located in Zakynthos. Users were also categorised into one of three stereotypes taking into account his/her believed level of expertise in computers and the Internet: novice, intermediate and expert and correspond to low, medium and high of expertise in ICT, respectively. For example, if a user has no computer skills then s/he is categorised as novice.

The body of the stereotype consists of some default assumptions about the users belonging to that specific stereotype. More specifically, in the proposed approach in ADAPTIGIS, default assumptions are parameterized and are given in the form of values for the criteria defined during analysis. In this way, stereotypes can easily be combined with the decision making model for providing personalized interaction.

The main feature of ADAPTIGIS is that it can adapt its interaction with each user. In order to evaluate different information, the system uses a simple decision making model. The suitability of each map for the particular user interacting with the system is estimated taking into account some criteria.

- Degree of Interest (i): The values of this criterion show how interesting each information about Zakynthos is for the particular user.
- Need for information (n): This criterion shows how important each information about Zakynthos is for the particular user.

- Comprehensibility of the information(c): This criterion also shows how comprehensible each information about Zakynthos is to the particular user.
- Level of computer skills (l): This criterion shows how comprehensible the way of presentation of each information about Zakynthos is to the particular user.

The values of these criteria are estimated taking into account the information that is stored in the user modeling component of the system. This component stores information about each individual user interacting with the GIS.
For the evaluation of the geographical information, the reasoning mechanism of the system uses the SAW method [15, 16]. According to the SAW method the multi-criteria function is calculated as a linear combination of the values of the four criteria that had identified in the previous experiments:
$U(X_j) = \sum_{i=1}^{4} w_i c_{ij}$, where w_i are the weights of criteria and c_{ij} are the values of the criteria for the X_j geographical information (map). The criteria used for the evaluation of the geographical information are considered equally important and, therefore, the formula for the calculation of the multi-criteria function is formed:

$$U(X_j) = 0.37i + 0.30n + 0.20c + 0.13l \tag{1}$$

In view of the values of the multi-criteria function for the different geographical information, the maps are ranked and the one with the highest value is considered to be the most suitable for the user that interacts with the system.

4 Evaluation Experiment

For the evaluation of the Adaptive GIS, the evaluation experiment was designed taking into account the steps of the FSAW method. For this purpose a committee of the decision-makers was formed to evaluate ADAPTIGIS. The committee of decision makers was consisted of 12 users that were randomly selected; 2 environmentalists, 4 students, 4 tourists and 2 software engineers. The group of decision makers was homogeneous as the reliability (importance) of the decision-makers was equal. The criteria that were taken into account for the evaluation of the Adaptive GIS was:

- Interest satisfaction (Is): this criterion reveals how successful the system was in addressing the users' interests.
- Information completeness (Ic): this criterion shows how successful the system was in providing all the appropriate information to the user interacting to the system.
- Needs fulfillment (Nf): this criterion shows how successful the system was in addressing the users' needs and presenting the appropriate information to the user.

- User friendly (Uf): the specific criterion shows what the users think about the interaction with the adaptive system, whether it is intrusive or not, whether it provides natural interaction etc.

The decision makers were first asked to evaluate the above mentioned attributes with respect to their importance in adaptive software evaluation. This procedure resulted in aggregated fuzzy weights of individual attributes. For the evaluation of the above mentioned criteria the decision makers used the linguistic variables presented in table 1.

Linguistic variables	**Fuzzy numbers**
Very low (VL)	(0, 0, 0, 3)
Low (L)	(0, 3, 3, 5)
Medium (M)	(2, 5, 5, 8)
High (H)	(5, 7, 7, 10)
Very high (VH)	(7, 10, 10, 10)

Table 1: Linguistic variables and fuzzy numbers for the importance weight.

As soon as all the linguistic variables of the criteria weights were collected, the fuzzy weights of the criteria are calculated. More specifically, each linguistic variable is translated to a fuzzy number as this is presented in table 1. These values are used for calculating the aggregated fuzzy weight for each one of the four attributes: $\widetilde{W}_{I_s} = (15.5, 23, 23, 29)$, $\widetilde{W}_{I_c} = (9.25, 17, 17, 25.75)$, $\widetilde{W}_{N_f} = (23, 20.5, 20.5, 28)$, $\widetilde{W}_{U_f} = (11.5, 19.5, 19.5, 27)$. These weights are defuzzified according to the step 4 of the theory: $d(\widetilde{W}_{I_s}) = 22.63$, $d(\widetilde{W}_{I_c}) = 17.25$, $d(\widetilde{W}_{N_f}) = 20.50$, $d(\widetilde{W}_{U_f}) = 19.38$. The crisp values of the normalized weights are calculated:

$$W_{I_s} = \frac{d(\widetilde{W}_{I_s})}{\sum_{j=1}^{4} \widetilde{W}_{I_s}} = \frac{22.63}{79.76} = 0.28 \ , \ W_{I_c} = \frac{d(\widetilde{W}_{I_c})}{\sum_{j=1}^{4} \widetilde{W}_{I_c}} = \frac{17.25}{79.76} = 0.22 \ ,$$

$$W_{N_f} = \frac{d(\widetilde{W}_{N_f})}{\sum_{j=1}^{4} \widetilde{W}_{N_f}} = \frac{20.50}{79.76} = 0.26 \ , \ W_{U_f} = \frac{d(\widetilde{W}_{U_f})}{\sum_{j=1}^{4} \widetilde{W}_{U_f}} = \frac{19.38}{79.76} = 0.24 \ ,$$

The weight vector is, therefore, formed to $W = (0.28, 022, 026, 024)$.

5 Results

The 12 users were asked then to evaluate ADAPTIGIS using the linguistic rating variables presented in Table 2 with respect to the four attributes.

Linguistic variables	Fuzzy numbers
Very poor (VP)	(0, 0, 0, 20)
Between very poor and poor (B. VP & P)	(0, 0, 20, 40)
Poor (P)	(0, 20, 20, 40)
Between poor and fair (B. P & F)	(0, 20, 50, 70)
Fair (F)	(30, 50, 50, 70)
Between fair and good (B. F & G)	(30, 50, 80, 100)
Good (G)	(60, 80, 80, 100)
Between good and very good (B. G & VG)	(60, 80, 100, 100)
Very good (VG)	(80, 100, 100, 100)

Table 2: Linguistic variables and fuzzy numbers for the ratings.

The linguistic rating variables were processed using the formulae of step 5 of the algorithm in order to obtain the aggregated fuzzy ratings. The detailed presentation of the calculations is beyond the scope of this paper and, therefore, only the results of this processing are presented in the fuzzy rating matrix (Table 3).

	Is	Ic	Nf	Uf
ADAPTIGIS	(185,245,255,285)	(117.5,177.5,200,255)	(97.5,157.5,195,255)	(127.5,187.5,220,270)
Standard GIS	(45,85,85,145)	(67.5,117.5,137.5,197.5)	(22.5,72.5,87.5,147.5)	(142.5,202.5,235,285)

Table 3: Fuzzy rating matrix.

ADAPTIGIS seems to be winning the standard GIS in most attributes. However, the standard GIS was considered by 2 users as user friendlier and non intrusive. In order to derive the total fuzzy scores for individual alternative, the fuzzy rating matrix is multiplied by the corresponding weight vector W.

$$\widetilde{F} = \begin{bmatrix} 185,245,255,285 & 117.5,177.5,200,255 & 97.5,157.5,195,255 & 127.5,187.5,220,270 \\ 45,85,85,145 & 67.5,117.5,137.5,197.5 & 22.5,72.5,87.5,147.5 & 142.5,202.5,235,285 \end{bmatrix} \otimes \begin{bmatrix} 0.28 \\ 0.22 \\ 0.26 \\ 0.24 \end{bmatrix} =$$

$$\widetilde{F} = \begin{bmatrix} (134.2,194.2,219.4,267.3) \\ (67.5,117.1,133.2,190.8) \end{bmatrix}$$

The crisp value for each total score is computed using the defuzzification method.

$$d(\widetilde{f_1}) = \frac{1}{4}(134.2+194.2+219.4+267.3) = 203.8$$

$$d(\widetilde{f_2}) = \frac{1}{4}(67.5,117.1,133.2,190.8) = 127.2$$

The crisp value of the first alternative that corresponds to ADAPTIGIS is much higher than the crisp value of the standard GIS. Therefore, the users seem to prefer in general ADAPTIGIS than a standard GIS.

6 Conclusions

Many researchers have focused on the shortage of software evaluation methodologies and experiments (eg. [17], [18]). This is especially the case of software that incorporates intelligence and is considered more complex. In this paper, we described how a fuzzy multi-criteria decision making theory can be used for evaluating a personalised GIS. The theory that was selected was the FSAW theory. The particular theory is very simple and uses linguistic terms. Indeed, previous studies (e.g. [19]) revealed that users had a difficulty in quantifying criteria while evaluating a software system. Therefore, in this case, FSAW was selected that uses linguistic terms that can be further translated into fuzzy numbers. This results in making the procedure more realistic, suitable, user friendly and effective.

The evaluation experiment revealed that the users could express themselves better using linguistic terms and, therefore, preferred this evaluation setting than others that use number rating. Furthermore, the evaluation results proved that the personalised GIS described in this paper surpass a standard GIS in selecting the information that is most useful for a particular user.

References

1. Malaka, R.Y., Zipf, A.: Deep map-challenging IT research in the framework of a tourist information system. In: Proceedings of International Congress on Tourism and Communication Technologies in Tourism, pp. 15–27 (2000)
2. Laukkannen, M., Helin, H., Laamanen, H.: Tourists on the move. In: Klusch, M., Ossowski, S., Shehory, O. (eds.) CIA 2002. LNCS, vol. 2446, pp. 36–50. Springer, Heidelberg (2002)
3. Schmidt-Belz, b., Poslad, S., Nick, A., Zipf, A.: Personalized and location based mobile tourism services. In: Workshop on Mobile Tourism Support Systems, in conjunction with the Fourth International Symposium on Human Computer Interaction with Mobile Devices, pp. 18–20 (2002)
4. Gervais, E., Hongsheng, L., Nussbaum, D., Roh, Y.-S., Sack, J.-r., Yi, J.: Intelligent map agents - An ubiquitous personalized GIS. Phtogrammetry & Remote Sensing 62, 347–365 (2007)
5. Kabassi, K., Charou, E., Martinis, A.: Implementation Issues of a Knowledge-based Geographical Information System. In: Knowledge-Based Software Engineering. Frontiers in Artificial Intelligence and Applications (Proceedings of the 8th Joint Conference on Knowledge-Based Software Engineering JCKBSE 2008) (2008a)
6. Kabassi, K., Virvou, M., Charou, E., Martinis, A.: Software Life-cycle for an Adaptive Geographical Information System. In: International Conference on Signal Processing and Multimedia Applications (SIGMAP 2008), pp. 393–396 (2008b)
7. McTear, M.F.: Intelligent interface technology: from theory to reality? Interacting with Computers 12, 323–336 (2000)
8. Chin, D.N.: Empirical Evaluation of User Models and User-Adapted Systems. User Modeling and User Adapted Interaction 11(1/2), 181–194 (2001)
9. Micarelli, A., Sciarrone, F.: Anatomy and Empirical Evaluation of an Adaptive Web-based Information Filtering System. User Modeling and User-Adapted Interaction 14, 159–200 (2004)

10. Cheng, R., Vassileva, J.: Design and evaluation of an adaptive incentive mechanism for sustained educational online communities. User Modeling and User Adapted Interaction 16, 321–348 (2006)
11. Chou, S.-Y., Chang, Y.-H., Shen, C.-Y.: A fuzzy simple addtive weighting system under group decision-making for facility location selection with objective/subjective attributes. European Journal of Operational Research 189, 132–145 (2008)
12. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
13. Keufmann, A., Gupta, M.M.: Introduction to Fuzzy Arithmetic: Theory and Application. Van Nostrand Reinhold, New York (1991)
14. Dubois, D., Prade, H.: Operations on fuzzy numbers. International Journal of Systems Science 9, 613–626 (1978)
15. Fishburn, P.C.: Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments. Operations Research (1967)
16. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications. Lecture Notes in Economics and Mathematical Systems, vol. 186. Springer, Heidelberg (1981)
17. Virvou, M., Alepis, E.: Mobile educational features in authoring tools for personalised tutoring. Computers and Education 44(1), 53–68 (2005)
18. Virvou, M., Katsionis, G., Manos, K.: Combining software games with education: Evaluation of its educational effectiveness, Educational Technology & Society. Journal of International Forum of Educational Technology & Society and IEEE Learning Technology Task Force (2005)
19. Virvou, M., Kabassi, K.: Experimental Studies within the Software Engineering Process for Intelligent Assistance in a GUI. Journal of Universal Computer Science 8(1), 51–85 (2003)

Animated Agents in the User Interface of Recommender Systems

Anastasios Savvopoulos and Maria Virvou

Abstract Adaptivity helps these systems to be adjustable to users' needs and inferences. There many ways to achieve adaptivity in a system. The most common ones are adaptive suggestions, which may include product or action suggestions to users. Lately, there is another technology used to help users in an adaptive way. This technology is the incorporation of animated lifelike agents into adaptive systems. In this paper we present two cases of systems that sell products, how we incorporated animated agents into them and how we used them in order to help users buy these products in a more effective way

1 Introduction

Nowadays there are lots of systems trying to help users more efficiently. In order to do so, these systems incorporate techniques of adaptivity. Adaptivity helps these systems to be adjustable to users' needs and inferences. There many ways to achieve adaptivity in a system. The most common ones are adaptive suggestions, which may include product or action suggestions. Concerning products these suggestions are often called recommendations and the corresponding applications recommendation systems. Concerning user actions, these systems are called adaptive help systems. Both systems try to find out in an explicit or implicit way what the user wants. In the first case, what the user wants to bay and in the second what the user wants to do. In both cases there are many technologies offered in order to achieve recommendations and adaptive help. An example is the adaptive hypermedia technology that changes the user's interface appearance in order to suit to his/her needs. Another example is dynamic user interfaces technology that changes the whole user interface in order prevent the user from recurrent mistakes. Lately, there is another technology used to

University of Piraeus, Department of Informatics,
Pireaus, Greece, 18534
e-mail: asavvop,mvirvou@unipi.gr

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 285–295.
springerlink.com

help users in an adaptive way. This technology is the incorporation of animated life-like agents into adaptive systems. These agents have been widely used in e-learning systems in order to help students solve problems, and they were called pedagogical agents. Their usage has proved very effective and helped students learn more easily. More specifically, Lester and Stone [1], point out that because animated agents have a strong life-like presence, they can capture students imaginations and play a critical role in keeping the deeply engaged in a learning environments' activities. So in recent years, there is a growing research on transferring this technology of animated agents into recommending and adaptive help systems, so these systems can also benefit from this life-like presence.
In this paper we present two cases of systems that sell products, how we incorporated animated agents into them and how we used them in order to help users buy these products in a more effective way.

2 Related Work

In the past years the animated agent technology was mainly used by e-learning applications. The animated agent technology proved successful in helping students use the e-learning systems and enhance their learning experience (Baylor [2]). The success of the animated agent technology lead to the creation of a new research domain and researchers called the animated agents pedagogical agents. An interesting research in the field of e-learning has been conducted by Amy Baylor [2]. Her research involved the creation of three pedagogical agents with three different roles. The first had the role of the motivator, the second of the expert and the third the role of the mentor. The experimental results showed that all agents fulfilled their roles and supported students during learning. On the other hand, we used the animated agent technology on an entirely different field. In our field, sales support, we used the animated agent in order help users buy products they want more effectively. In both cases presented here the animated agent communicates with the user model of every user provided by the main mechanism of the application. Then the animated agent helps the user in a personalized way based on this user model. Another approach on the field of pedagogical agent has been made by Veletsianos et al.[3]. In their research, they created two pedagogical agents, one male and one female, and employed them in two teaching courses. The findings of their research suggest that learners found male agents to be more outgoing and agreeable than female agents. In our applications, we used male animated agents in order to help users. Our animated agents were not only created to help and motivate users buy products, but to provide product recommendations based on users' models in a novel way. An important work on the same field of pedagogical agents has also been made by Johnson et al. [4]. In their work they created Adele, a pedagogical agent that assists students as they assess and diagnose medical and dental patients in clinical settings. The experimental results showed that Adele can be applied to future educational applications of interface agent technology. On the other, we used the animated agent technology

to support customer buy products and use commerce applications. Our agents interacted with users and provided them with useful suggestions and information based on their user models. Despite the fact, that the main research on animated agents has been done on the field of e-learning, there are other fields that incorporated animated agents. An example is the work done by Andre et al. [5]. In their work they created multipurpose animated presentation agent that was used in different applications in order to point gesture at points of interest in the user interface. On the other hand, our animated agent was not only used for pointing out suggested products but helped users in an adaptive way. Our agents interacted with every user in a personalized way according to their needs and interest extracted from their user model. Another example of interesting work with animated agents is the work of Kirschning and Rueda [6]. They incorporated microsoft agents into html documents and used them to motivate children to interact with digital encyclopedias. Their results showed that children were helped by the use of the animated agent. In our work shown here, we incorporated an animated agent not only in an internet based application but also in an interactive TV application. We also used the animated agent not only for helping users interact with application but also for suggesting products in a personalized way.
Lastly, on the field of e-commerce and product recommendation much work has been done by Kieling et al. [7]. They created the COSIMA project, which a smart e-sales assistant. The COSIMA is java based server that controls the animated agent. The agent incorporates chatting with customers and smart sales advices. Despite the fact, that our technology was incorporated on the same field and had the same purpose of selling products more efficiently, we followed a different approach. Instead, of relying on sql based search engines of products in order to extract customer preferences, used in the COSIMA project, we connected our animated agent architectures with the user models created from the main reasoning mechanism of the application we incorporated the animated agent. In this way, our technology is not affected by the product being sold. We have also incorporated the animated agent in two different mediums, the web and the interactive TV, using very similar architectures. This shows that our technology is not affected by the medium that incorporates it.

3 Incorporating animated agent into e-commerce: Vision.Com

The first case we applied the animated agent was Vision.Com. Vision.Com is an e-commerce video store that learns from customers' preferences. A screenshot of Vision.Com is illustrated in Figure 1. Its aim is to provide help to customers choosing the best movie for them. Vision.Com operates on top of the .net framework technology and uses client server techniques. The web based system runs at a local network with IIS (internet information services) playing the role of the web server. For every user the system creates a different record at the database. There are two types of information saved for every user, the explicit and implicit. In Vision.Com every

customer can visit a large number of movies by navigating through four movie categories. These four movie categories are: social, action, thriller and comedy movies. All customers have their own personal shopping cart. If a customer intends to buy a movie she/he must simply move the movie into her/his cart by pressing the specific button. Users also have the ability to remove one or more movies from their cart by choosing to delete them. After deciding which movies to buy a customer can easily purchase them by pressing the button “buy”. All navigational moves of a customer are recorded by the system in the statistics database. In this way Vision.Com saves statistics considering the visits in the different categories of movies and movies individually. The same type of statistics was saved for every customer and every movie that was moved to the buyers’ cart. The same task is conducted for the movies that are eventually bought by every customer. All of these statistical results are moderated from one to zero and saved in the statistics database. In particular, Vision.Com interprets users’ actions in a way that results in the calculation of users’ interests in individual movies and movie categories. Each user’s action contributes to the individual user model by implying degrees of interest into one or another movie category or individual movie. The features measured by vision.com are movie categories, actors, directors and price range.
The architecture that combines the animated agent with the system is shown in figure 1. The architecture includes the animation space, the balloon space and the balloon dialogues. The animation space is a database that contains all the animations of the Microsoft Agent Character. This character includes 94 animations that help him to create a lifelike behaviour in a similar way as in other adaptive systems (e.g. Lester et al. 1999). It also contains the library file of creating the character. The balloon space contains one type of balloon that gives the agent the opportunity to communicate one-way with the user. This is why this balloon space contains only the phrases that the agent can speak. To help the user communicate in both-ways we implemented a second balloon space of balloon dialogues that consists of a library of files of different types of balloons. These balloons include different forms that allow the user to enter data, questions and choices in order to communicate with the agent more freely.
The animated agent uses the user’s model to perform two types of recommendations, general recommendations and specific recommendations. General Recommendations are recommendations based only on one characteristic measured in this specific user model. For example, if a user’s model shows a high interest degree in a specific movie category then the animated agent collects this information and when the user enters the movie categories page the animated agent takes action. This involves three steps. First, the agent moves and points the category title. Second, the agent with the help of the balloon dialog informs the user about this category and third, with the help of the text to speech technology informs suggests this category to the user by sound. An example illustration of general recommendations can be seen in figure 2.

The second types of recommendations are specific recommendations made in the specific category movie page. The main difference between general and specific recommendations is that specific recommendations take into concern all the

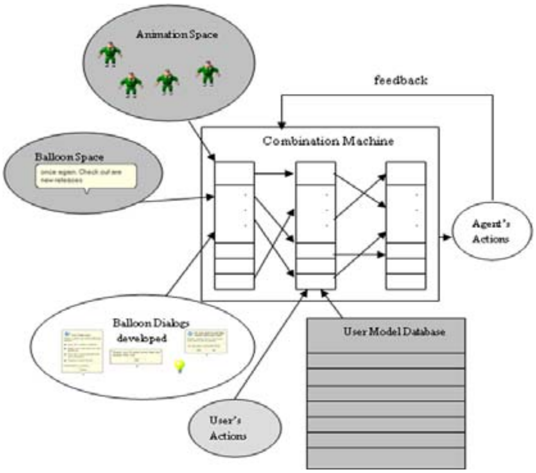


Fig. 1: The Animated Agent Architecture.

Fig. 2: Animated agent's general recommendation example.

characteristics in a user's model. Then the animated agent chooses the highest degrees of interest in all these characteristics and based on these degrees searches the movies databases to find movies matching these criteria. If no movie is found based on these criteria then the lowest of these highest degrees of interest is excluded and the search begins again. This process continues until a certain amount of movies matching these criteria are found. After gathering this list of favorite movies the agent is ready to make specific recommendations. The process of making a specific movie recommendation involves four steps. The first three are the same as general recommendation meaning that the agent moves next to the specific movie title, points out and suggests the user this movie. The final step involves suggesting the user to move this movie to his cart without the user moving to this movie's specific page. If the user chooses so, then the movie is automatically placed to his shopping cart. An example illustration can be seen in figure 3.

Lastly, the animated agent except recommendation actions performs general actions as well. These actions involve social actions and help actions. Social actions involve welcoming the user to application and humor actions. These actions help

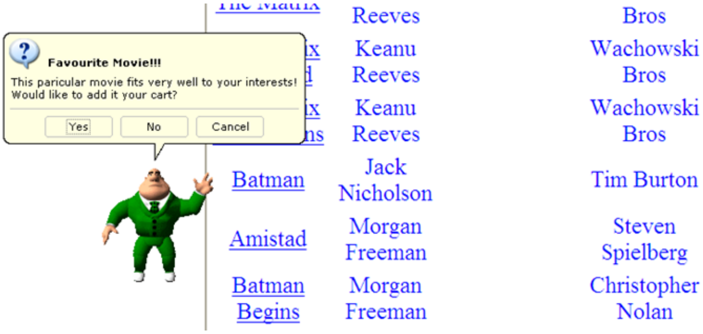


Fig. 3: Move to cart suggestion example.

the users feel friendlier while interacting with the application (figure 4). Next, help actions can help the users find out how to use the vision.com application. For example, if a user double clicks the animated agent then a balloon dialog is shown that contains possible frequently questions that concern the usage of the system. By selecting on these questions the balloon dialog presents the corresponding answer. In this way the user does not have to browse through the entire help catalogue to find help about his/her problem. An example of help action can be seen in figure 4 (right).

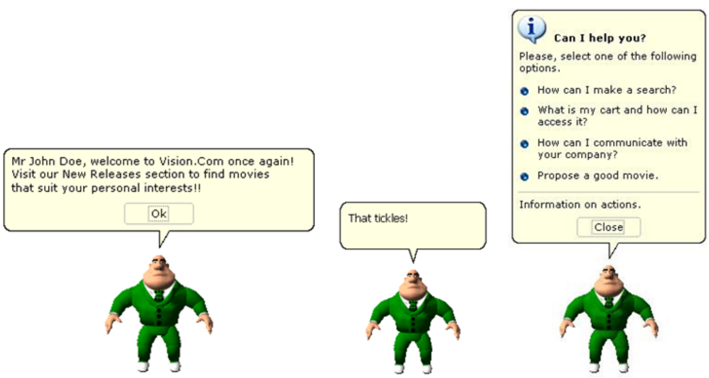


Fig. 4: Social and help actions examples.

4 Incorporating animated agent into TV-commerce: iTVMobi

The second case we incorporated our animated agent technology was iTVMobi. Here we extended our animated agent functionality. In iTVMobi the animated agent not only suggested products but with the help of the adaptive help system of the application involved into helping users in an adaptive to their specific problems way. iTVMObi is a mobile shop created for the interactive television that learns from customer preferences. iTVMobi was built on Microsoft TV (MSTV) technology. The core components of MSTV are available in the Windows XP operating system and can be run on personal computers (PCs). MSTV technology can be utilized within a familiar and mature Integrated Development Environment (IDE). Microsoft Visual Studio offers a multitude of tools for designing, developing, testing and deploying an application. iTVMObi can be used by a telemarketing channel to sell mobile phones in a personalized way. Its aim is to provide help to customers of all ages by suggesting the best mobile phone for them. In order to help customers make the best buy iTVMObi performs two functions. The first is a recommender system that makes suggestions concerning mobile phones and the second is the adaptive help system that provides adaptive help generated for the users. The system can learn about the users' preferences and mistakes and provide more helpful responses.
For every user, iTVMObi creates a different record at a database. There are two types of information saved for every user, the explicit and implicit. The explicit information is saved on the Explicit Information In iTVMObi every customer can visit a large amount of mobile phones. For the purposes of our research we have implemented the system for five popular mobile brands. Every customer has her/his own personal shopping cart. If customers intend to buy a phone they must simply move the phone into their cart by pressing the specific button or they can press the buy button at their remote control at the time that the specific product is shown on their TV screen. They also have the ability to remove one or more phones from their cart by choosing to delete them. After deciding which phones to buy a customer can easily purchase them by pressing the button "buy" at their shopping cart. All navigational moves of a customer are made through the TV remote control and are recorded by the system in the statistics database. In this way iTVMObi saves statistics considering the visits in the different brands and specific phones individually. The same type of statistics is saved for every customer and every phone that is moved to the buyer's cart. The same task is conducted for the mobile phones that are eventually bought by every customer. All of these statistical results are scaled to the unit interval [0, 1]. In particular, ITVMObi interprets users' actions in a way that results into two different functions. The first is the calculation of users' interests in individual phones and production companies and the second is the interpretation of users' actions concerning possible navigation mistakes. Each user's action contributes to the individual user model by showing degrees of interest into one or another company or individual phone or by showing likelihood on a specific mistake. The features measured for interest degrees are the five popular brads, phone price range, phone technical features, phone size, phone connectivity, phone display features, phone memory capabilities, phone battery autonomy and price range. As for mistakes

degrees we consider the following: difficulty of the user to see brands' names, difficulty to see phones' names and pictures and the confusion degree. The architecture for combing the animated agent with the rest of the system's modules is similar to the one used in Vision.Com. The animated agent in iTVMobi uses the user's model and checks both interest and mistake degrees. The agent then performs two types of actions, recommendations and help actions. As for the recommendations the process is similar to the previous case of Vision.Com. Again, here we have two types of recommendations, general and specific. General Recommendations suggest users' phones or companies based on a single feature measured (figure 5). For example if a user has shown great interest on wireless communications on mobile phones than the agent will search the phone with the greatest wireless communications features from the company that the user clicked into. The steps of suggestion are similar as Vision.Com. Again, here the agent moves next to the phone, points it and suggests it concerning the feature measured.

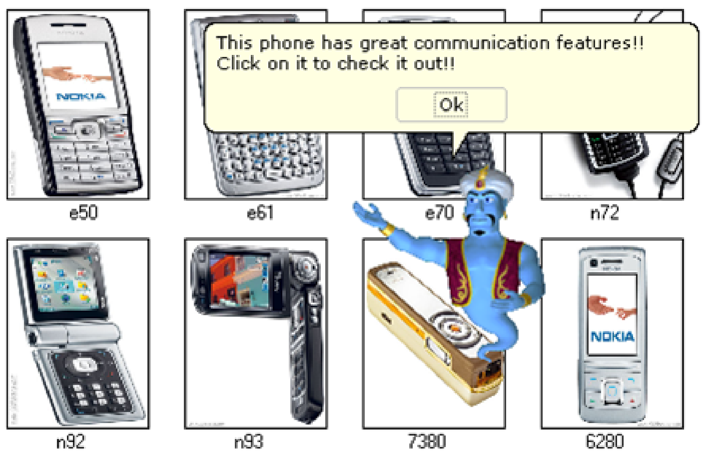


Fig. 5: Animated agent general suggestion example.

Specific recommendations are also made in a similar way to Vision.Com. Again, here all highest features are taken into consideration. Then the agent searches the phones database to find phones that match those criteria. If no phones are found the lowest interest degree is excluded from the criteria and the search begins again. The search continues until a certain number of phones are found from each of the phones companies. The suggestion is performed similar to Vision.Com (figure 6).

In iTVMobi we extended the animated agent actions by introducing adaptive help actions. By saving to every user's model degrees concerning mistakes we managed to predict users' mistakes and change the system behavior accordingly. This meant changing the user interface dynamically as well the agent's actions dynamically. Every time a customer uses the system the adaptive help system finds his/her

Fig. 6: Move to cart suggestion example.

representative and responds with adaptive help actions. An example can be seen in Figures 7, 8. In this particular example the system observes the user's navigation moves between two neighboring mobile phones and counts his/her mistakes. If a user has made a lot mistakes in this section, like browsing two neighboring phones repeatedly without putting any of phones in his cart at the meantime, then the system identifies that the user cannot view the phone pictures clearly and chooses to enlarge them. If the mistakes between the two neighboring phones continue then the system identifies that the user has confused only these two phones. The action taken by the system is to change the location of these two phones and move the one away from the other, while bringing a different phone close in order not to destroy the whole arrangement in the screen of the phones. If the user continues to make the same kind of mistakes then the system uses the animated agent in order point the phones by moving next to them, showing them with its "hand" and then telling with its "voice" the model of the phone. The system also increases the sound volume in order to help people with hearing problems understand more clearly the point out function of the animated agent.

Again, in iTVMobi, the animated agent except recommendation and adaptive help actions performs general actions as well. These actions involve social actions and help actions. These actions help the users feel friendlier while interacting with the application.

5 Conclusions

In this paper we presented two different cases of applications, that we incorporated the technology of animated agents. We created two similar architectures for both

Fig. 7: **Left** First stage. Small phone pictures. **Right** Second stage. The user has made mistakes. Bigger phone pictures and a next button showing that phones are split in two pages.

Fig. 8: **Left** Third stage. The user has confused the first two phones on the bottom. The system has changed their locations. **Right** Fourth stage. The user continues to confuse the phones. The system enables the animated agent in order to point the phones and increases the sound volume the agent..

applications in order to incorporate and control the animated agent. The animated agent on both cases helped users buy products in a personalized way. Moreover, in the second case the animated agent helped users overcome their navigation problems. Last but not least, in both cases every user could interact with the animated agent in order acquire useful information about the usage of the application. The personalization was achieved with the help of every user's model created by the reasoning mechanism of the main application that the animated agent was incorporated into. The animated agent technology helped in creating friendlier user interfaces and increased the user's ability to understand the usage of every application.

References

1. Lester, J.C., Stone, B.A.: Increasing believability in animated pedagogical agents. In: Proceedings of the First international Conference on Autonomous Agents. AGENTS 1997, Marina del Rey, California, United States, February 05 - 08, pp. 16–21. ACM, New York (1997)

2. Baylor, A.L., PALS (Pedagogical Agent Learning Systems) Research Group 2003: The impact of three pedagogical agent roles. In: Proceedings of the Second international Joint Conference on Autonomous Agents and Multiagent Systems. AAMAS 2003, Melbourne, Australia, July 14 - 18, pp. 928–929. ACM, New York (2003)
3. Veletsianos, G., Yerasimou, T., Doering, A.: The Role of Intelligent Agents on Learner Performance. In: Kommers, P., Richards, G. (eds.) Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005, pp. 3638–3643. AACE, Chesapeake (2005)
4. Johnson, W.L., Shaw, E., Marshall, A., LaBore, C.: Evolution of user interaction: the case of agent adele. In: Proceedings of the 8th international Conference on intelligent User interfaces. IUI 2003, Miami, Florida, USA, January 12 - 15, pp. 93–100. ACM, New York (2003)
5. André, E., Müller, J., Rist, T.: The PPP persona: a multipurpose animated presentation agent. In: Catarci, T., Costabile, M.F., Levialdi, S., Santucci, G. (eds.) Proceedings of the Workshop on Advanced Visual interfaces. AVI 1996, May 27 - 29, pp. 245–247. ACM, New York (1996)
6. Kirschning, I., Rueda, J.O.: Animated agents and TTS for HTML documents. In: Proceedings of the 2005 Latin American Conference on Human-Computer interaction. CLIHC 2005, Cuernavaca, Mexico, October 23 - 26, pp. 148–154. ACM, New York (2005)
7. Kieβling, W., Fischer, S., Holland, S., Ehm, T.: Design and Implementation of COSIMA -A Smart and Speaking E-Sales Assistant. In: Proceedings of the Third international Workshop on Advanced Issues of E-Commerce and Web-Based information Systems (WECWIS 2001). WECWIS, June 21 - 22. IEEE Computer Society, Washington (2001)

A Web-Based Interactive System for Learning NL to FOL Conversion

Ioannis HATZILYGEROUDIS and Isidoros PERIKOS

Abstract In this paper, we present NLtoFOL SIP system, a web-based interactive system aimed at helping students to learn how to convert/translate natural language (NL) sentences into first-order logic (FOL) formulas. It tries to achieve it by providing (a) a structured and interactive process (SIP) for the conversion and (b) guidance and help during that process. The system provides guidance and help of various levels in an intelligent way based on the user's responses. Also, the user interface is dynamically configured during the user interaction to reflect the steps of SIP. Different NL sentences may require the implementation of different number of SIP steps. According to our knowledge, there is no other system that tackles the problem of NL to FOL conversion in such a systematic way. Small scale evaluation has given quite satisfactory results.

1 Introduction

Knowledge Representation & Reasoning (KR&R) is a fundamental topic of Artificial Intelligence (AI). A basic KR language is First-Order Logic (FOL), the main representative of logic-based representation languages, which is part of almost any introductory AI course and textbook (Russell and Norvig 2003, Luger, 2004). In some proposals, emphasis is given to the propositional logic (PL) (Neller et al 2006) as an easier way of understanding logic. However, PL is weaker than FOL, in that we cannot use variables. This makes PL less interesting and less expressive. So, FOL is more exciting and expressive in representing knowledge and reason with it, although more complicated. To make automated inferences, Clause Form (CF), a special form of FOPC, is used. Teaching FOL as a knowledge representation and reasoning language includes many aspects. One of them is translating natural language (NL) sentences into FOL formulas, often called logic formalization of NL

Department of Computer Engineering & Informatics, School of Engineering University of Patras, Greece

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 297–307.
springerlink.com

sentences. It is an ad-hoc process; there is no specific algorithm that can be automated within a computer. This is mainly due to the fact that NL has no clear semantics as FOL does. Also, most of existing textbooks do not pay the required attention to that. They simply provide the syntax of FOL and definitions of the logical symbols and terms (Russell and Norvig 2003, Luger 2004). Even more specialized textbooks do the same (Brachman and Levesque 2004). At best, they provide a kind of more extended explanations and examples (Genesereth and Nilsson 1987). They do not provide any systematic guidance towards it. Given the above, students usually find difficulties in learning the task of formalizing NL sentences in FOL, which confronts to tutors' common experience. We consider that the NL to FOL conversion is of great importance in teaching logic as a knowledge representation (KR) language. Understanding of when and for what to use a predicate, a function or a logical connective gives a more complete insight and understanding of FOL as a KR language. Furthermore, given that a knowledge engineer will often face the fact of converting NL sentences into FOL, while acquiring knowledge, exact knowledge of the factors of the conversion is very crucial. There are several systems that are characterized as logic educational software (for an account see `http://www.cs.otago.ac.nz/staffpriv/hans/logiccourseware.html`). However, most of them deal with how to construct formal proofs mainly using natural deduction rules, restricting themselves to PL. None of them is treating logic as a KR&R language. More specifically, none of them is dealing with the process of translating/converting NL expressions to FOL formulas. The main reason is that there is no a systematic way to do that. In (Hatzilygeroudis 2007), we introduced a structured process for guiding students in translating a NL sentence into a FOL one. This is a fully interactive process that can be implemented on a computer. In this paper, we introduce an interactive web-based system implementing the above mentioned process. It is used to assist students in learning how to translate/convert a NL sentence into a FOL one.

2 Related Work

There various systems that are used or have been created for teaching or helping in teaching logic and logic-based reasoning. Plato (Plato) and Logic Toolbox (Logic Toolbox) are two examples of logic educational software. However, they deal with how to construct formal proofs mainly using natural deduction rules, restricting themselves to PL. The same happens with Jape (Bornat and Sufrin 1997), where the user can interact with a proof process at various stages. PROOFWEB (Kaliszyk et al 2007) has been motivated by Jape and can be considered is an "improved Jape-clone". It is again for teaching natural deduction proofs. It actually provides an interactive interface to Coq (`http://coq.inria.fr/doc/main.html`), a higher order proof assistant. Finally, OLIVER (Wildenberg and Scharff 2002) is another similar web-based system, where instant feedback is provided to the students about the correctness or not of their proposal at each step. Logic Tutor (Abraham et al 2001), on the other hand, is an intelligent tutoring system (ITS) for learning formal proofs in PL based on natural deduction. As an intelligent system, it adapts to the needs of the students via keeping user models. So, it provides context-sensitive feedback and exercises tailored to the users. Logic-ITA (Abraham and Yacef 2002, Lesta and Yacef 2002) is actually an extension to Logic Tutor, where a tutor part has been added, used as a teaching assistant system. P-Logic Tutor (Lukins et al 2002) is also a kind of intelligent tutoring system aiming at teaching students fundamental aspects of PL and theorem proving. To this end, it provides an interactive web-based interface. In the same vein, Hoare Logic Tutor (Goshi et al 2001) is an ITS for teaching and learning Hoare logic, a special kind of logic. All the above systems, although deal with learning and/or teaching logic, they are not concerned with how to use predicate logic as a KR&R language. More specifically, they do not deal at all with how to formalize a NL sentence into FOL. As far as we are aware of, there is only one system that claims doing the latter. It is called KRRT (Knowledge Representation and Reasoning Tutor) (Alonso et al 2007). It is a web-based system that aims at helping students to learn FOL as a KR&R language. It is based on FITS (Alonso et al 2006), its predecessor system, and deals with both knowledge representation in and reasoning with FOL. The translation form NL to FOL takes place in its KR part. However, the only help provided to the students is at syntactic and logical equivalence levels. The student gives his/her FOL proposal sentence and the system checks its syntax and whether it is the correct one (here equivalent sentences are acceptable). However, it does not provide any guidance about how to make that translation or even what is the kind of error made by the student.

3 A Structured and Interactive Process for NL to FOL Conversion

One problem in converting natural language into first order logic has to do with the unclear semantics that natural language has. Natural language has no clear semantics as FOL has. However, the main difficulty comes from the lack of a systematic way of making the conversion. In a previous work (Hatzilygeroudis 2007), we introduced SIP (Structured and Interactive Process), a process that guides a student in translating/converting a NL sentence into a FOL one. SIP, which has been somewhat modified from that in [xx], has as follows:

1. Spot the verbs, the nouns and the adjectives in the sentence and specify the corresponding predicates or function symbols.
2. Specify the number, the types and the symbols of the arguments of the function symbols (first) and the predicates (next).
3. Specify the quantifiers of the variables.
4. Construct the atomic expressions (or atoms) corresponding to predicates.
5. Divide produced atoms in groups of the same level atoms.
6. Specify the connectives between atoms of each group and create corresponding logical formulas.

7. Divide produced formulas in groups of the same level formulas.
8. If only one group of formulas is produced, specify the connectives between formulas of the group, create the next level formula and go to step 10.
9. Specify the connectives between formulas of each group, create the next level formulas and go to step 7.
10. Place quantifiers in the right points in the produced formula to create the final FOL formula

To demonstrate the steps of the above process, we present the conversion of the NL sentence “Every city has a dog-catcher of it that has been bitten by every dog living in the city” into a FOL formula.

Step 1. specify predicates/functions
city $\rightarrow$ predicate: *city*
dog $\rightarrow$ predicate: *dog*
has a dog-catcher $\rightarrow$ predicate: *dog-catcher*
has been bitten $\rightarrow$ predicate: *is-bitten-by*
living in $\rightarrow$ predicate: *lives-in*

Step 2. Number, types and symbols of arguments

Predicate	Arity	Types	Symbols
city	1	variable	x
dog	1	variable	y
dog-catcher	2	variable, variable	z, x
is-bitten-by	2	variable, variable	z, y
lives-in	2	variable, variable	y, x

Step 3. Quantifiers
x $\rightarrow \forall$
y $\rightarrow \exists$
z $\rightarrow \forall$

Step 4. Atoms
Atom 1: *city(x)*
Atom 2: *dog(y)*
Atom 3: *dog-catcher(z,x)*
Atom 4: *is-bitten-by(z, y)*
Atom 5: *lives-in(y, x)*

Step 5. Groups of atoms of the same level
GroupAtom1: {*city(x)*}
GroupAtom2: {*dog-catcher(z, x)*}
GroupAtom3: {*is-bitten-by(z, y)*}
GroupAtom4: {*dog(y), lives-in(y, x)*}

<u>Step 6. Connectives and formulas of groups</u>
GroupAtom1 → Form1: *city(x)*
GroupAtom2 → Form2: *dog-catcher(z, x)*
GroupAtom3 → Form3: *dog(y)* ∧ *lives-in(y, x)*
GroupAtom4 → Form4: *is-bitten-by(z, y)*

<u>Step 7. Groups of the same level formulas</u>
GroupForm1-1: {*city(x)*}
GroupForm2-1: {*dog-catcher(z, x)*}
GroupForm3-1: {*dog(y)* ∧ *lives-in(y, x), is-bitten-by(z, y)*}

<u>Step 9. Connectives and formulas of groups</u>
GroupForm1-1 → Form1-1: *city(x)*
GroupForm2-1 → Form2-1: *dog-catcher(z, x)*
GroupForm3-1 → Form3-1: *(dog(y)* ∧ *lives-in(y, x))*⇒ *is-bitten-by(z, y)*

<u>Step 7. Groups of the same level formulas</u>
GroupForm1-2: {*city(x)*}
GroupForm2-2: {*dog-catcher(z, x), (dog(y)* ∧ *lives-in(y, x))* ⇒ *is-bitten-by(z, y)*}

<u>Step 9. Connectives and formulas of groups</u>
GroupForm1-2 → Form1-3: *city(x)*
GroupForm2-2 → Form2-3: *dog-catcher(z, x)*∧ *((dog(y)* ∧ *lives-in(y, x))* ⇒ *is-bitten-by(z, y))*

<u>Step 7. Groups of the same level formulas</u>
GroupForm1-3: {*city(x), dog-catcher(z, x)* ∧ *((dog(y)* ∧ *lives-in(y, x))* ⇒ *is-bitten-by(z, y))*}

<u>Step 8. Connectives and formula of last group</u>
GroupForm1-3 → Form1-4: *city(x)* ⇒ *(dog-catcher(z, x)* ∧ *((dog(y)* ∧ *lives-in(y, x))* ⇒ *is-bitten-by(z, y)))*

<u>Step 10. Final formula</u>
(∀x) city(x) ⇒ *((∃z) dog-catcher(z, x)* ∧ *(y) (dog(y)* ∧ *lives-in(y, x))* ⇒ *is-bitten-by(z, y))*

4 System Description

In this section, we present the system that implements the NLtoFOL SIP, discussed in the previous section. First, we present the architecture of the system and then we concentrate in the interaction process and the interface of the system.

4.1 System Architecture

The NLtoFOL SIP system is a web-based interactive system intended to help students in learning how to translate a NL sentence into a FOL one. Its architecture is illustrated in Figure 1. It consists of four units: the *system database*, the *interface configuration* unit, the *intelligent data analysis* unit and the *user interface*.
The system database is mainly used to store sentences and related information. A number of NL sentences with corresponding steps of SIP and related information are pre-stored in the database and used during the system operation. The interface configuration unit is responsible for configuring the user interface according to the current step in SIP and the sentence the user is dealing with. So, the user interface of the system is dynamically re-configured to adapt to the needs of the specific session. This is achieved via the guidance given by the intelligent analysis unit, which is a rule-based system that based on the input data from user interface decides on which reconfigurations should be made to the user interface or which kind of interaction will be allowed or given to the user. It also is responsible for tracing user's mistakes and handling them in terms of appropriate feedback to the user.

4.2 User Interaction and Feedback

As soon as the user registers the system, it asks him/her to choose the difficulty level (1-5) of the sentence to be dealt with. Then the system presents a list of sentences (from its database) that are of the selected difficulty level. Then the user/student selects one of them to proceed and the system displays the selected sentence and a structure with empty boxes for step 1 to be filled in by the user (see Figure 2). Boxes are always more than required.

At step 1, as well as at each step, the user is given a brief explanation of what he should do, i.e. how to fill in the boxes. Also, a number of guidelines related to the process are given to him/her. For example, for step 1, guidelines such as the following are given:

- Usually predicates are expressed in third person.
- Auxiliary verbs like various forms of 'be' and 'have' usually do not correspond to separate predicates, but are merged with other verbs or nouns.

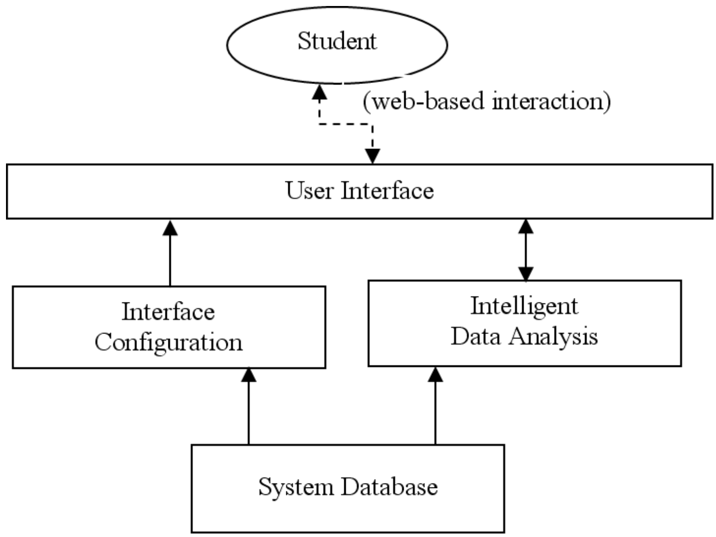


Fig. 1: NLtoFOL SIP System Architecture.

After having filled in the corresponding boxes, the user submits his/her answer to the system. Then the system responds in various ways depending on the answer of the user. If the answer is right, the system informs the user about it and calls him/her to proceed to the next step. If not, the system returns an error message that informs the user about the possible error. Possible errors are stored in the database. The intelligent data analysis component of the system finds which is the most probable error made by the user and returns corresponding message. For example it finds whether the error is related to the number of predicates or to a wrong predicate etc. Also, more than one level of help or hints is available to the user at each step that guides him/her how to correct the error(s). The help depends on the specific sentence and error. So, the content of the 'help' button is dynamically filled in too. The user may or may not use the provided help. In any case, to proceed to the next step, the current step should be correctly answered in all its parts. This is necessary, because each next step depends on its previous one and any wrong part will be propagated in the next steps. Recall that the aim of the system is not student evaluation but learning. After a step is correctly completed, the system proceeds to the next and so on and so forth as far as the last step.

4.3 System Interface

The interface of the NLtoFOL SIP system, as stated earlier, is dynamically re-configured, during a user session. It is designed to reflect SIP, presented in Section 3.

In Figures 2 and 3 two snapshots of a session with NLtoFOL SIP system are presented, in the current form of interface. In Figure 2 the first step of SIP for sentence “All humans eat some food” is depicted, whereas in Figure 3 the situation after the completion of the third step and display of the fourth step structure is depicted. The ‘Word’ boxes, in step 1, are filled in by the user, whereas in the ‘Type’ boxes the user selects one of the possible predefined types of a ‘Word’, i.e. one of ‘predicate’ or ‘function’ provided via a pop-down menu.
The structure of step2 depends on the answers to step1. So, given that there are three predicates specified in step1, there are three rows in step 2, each corresponding to one of the predicates. Also, in step 3, as many rows as the specified variables in the previous step are created and displayed.

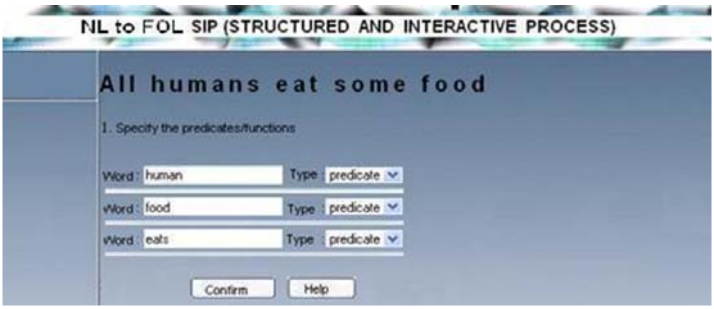


Fig. 2: A snapshot of the step 1 of SIP.

5 Conclusions

In this paper, we present NLtoFOL SIP system, a web-based interactive system aimed at helping students to learn how to convert NL sentences into FOL formulas. It tries to achieve it by providing (a) a structured and interactive process (SIP) for the conversion and (b) guidance and help during that process. The system provides guidance and help of various levels in an intelligent way based on the user’s responses. Also, the user interface is dynamically configured during the user interaction to reflect the steps of SIP. Different NL sentences may require the implementation of different number of SIP steps.
According to our knowledge, there is no other system that tackles the problem of NL to FOL conversion in such a systematic way. Existing tools and textbooks do not pay as much attention as required to teaching formalization of NL sentences into FOL.
We performed a small scale evaluation of a first version of the system using a group of five students. Their reaction was really encouraging. They found the system very

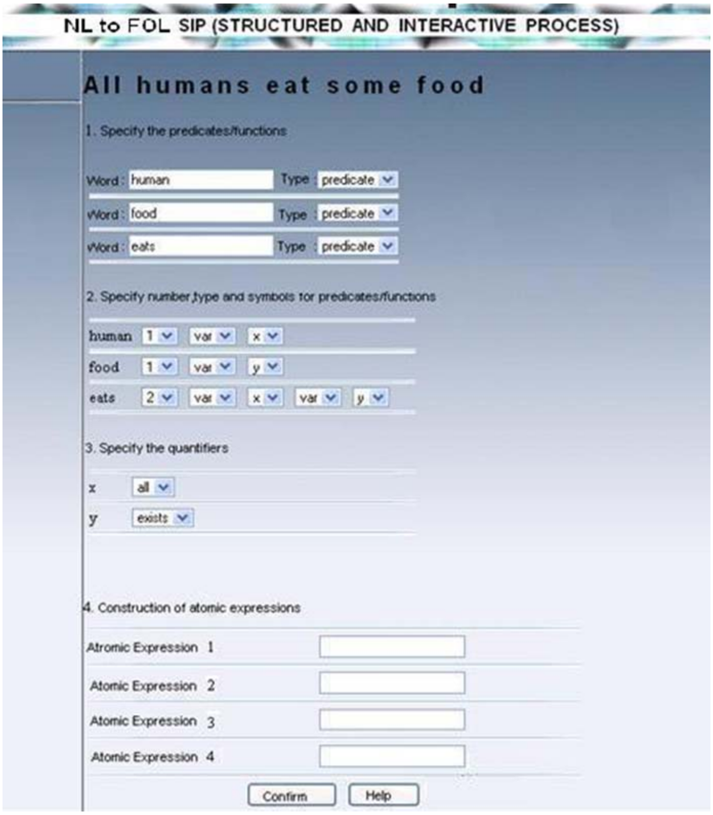


Fig. 3: A snapshot of SIP interface after completion of the third step.

useful and user friendly. They however noticed some kind of problem when dealing with somewhat complicated sentences, resulting in formulas with more than one level of implication. In such cases, steps 7-9 are difficult to follow. Also, they suggested that the first two steps could be united in one.

Apart from the above points, there are a number of ways that the system could be improved. One way is to make it more intelligent, i.e. more adaptable to the user, by introducing a user model which will keep information about the user, like the sentences he/she has dealt with, the mistakes made etc. Then that information could be used for better selection of the proposed sentences.

Also, what is missing is a tutor component that would help the tutor to manage the sentences and the related information (e.g. help, hints, error messages etc) in the database. Although it is relatively easy to implement an appropriate interface towards this end, the big problem remains: it is quite time consuming and tedious to store in the database for each sentence so much information, i.e. its analysis in all steps, with possible alternatives in some or all of them, corresponding help, hint and error items for each step etc. So, an interesting further research effort would be to semi-automate or even automate the above process. Of course, this is not possible for any NL sentence. However, it could be easier if it was done for a number of sentence templates of predefined structure.

References

Abraham et al 2001. Abraham, D., Crawford, L., Lesta, L., Merceron, A., Yacef, K.: The Logic Tutor: A multimedia presentation. Electronic Journal of Computer-Enhanced Learning (2001)

Abraham and Yacef 2002. Abraham, D., Yacef, K.: Adaptation in the Web-Based Logic-ITA. In: De Bra, P., Brusilovsky, P., Conejo, R. (eds.) AH 2002. LNCS, vol. 2347, pp. 456–461. Springer, Heidelberg (2002)

Alonso et al 2006. Alonso, J.A., Aranda, G.A., Martín.-Matceos, F.J.: FITS: Formalization with an Intelligent Tutor System. In: Proceedings of the IV International Conference on Multimedia and Information and Communication Technologies in Education (2006)

Alonso et al 2007. Alonso, J.A., Aranda, G.A., Martín-Matceos, F.J.: KRRT: Knowledge Representation and Reasoning Tutor. In: Moreno Díaz, R., Pichler, F., Quesada Arencibia, A. (eds.) EUROCAST 2007. LNCS, vol. 4739, pp. 400–407. Springer, Heidelberg (2007)

Bornat and Sufrin 1997. Bornat, R., Sufrin, B.: Jape: A calculator for animating proof-on-paper. In: Mc-Cune, W. (ed.) CADE 1997. LNCS, vol. 1249, pp. 412–415. Springer, Heidelberg (1997)

Brachman and Levesque 2004. Brachman, R.J., Levesque, H.J.: Knowledge Representation and Reasoning. Elsevier, Amsterdam (2004)

Genesereth and Nilsson 1987. Genesereth, M.R., Nilsson, N.J.: Logical Foundations of AI. Morgan Kaufmann, Palo Alto (1987)

Hatzilygeroudis 2007. Hatzilygeroudis, I.: Teaching NL to FOL and FOL to CL Conversions. In: Proceedings of the 20th International FLAIRS Conference, Key West, FL, May 2007, pp. 309–314. AAAI Press, Menlo Park (2007)

Kaliszyk et al 2007. Kaliszyk, C., Wiedijk, F., Hendriks, M., van Raamsdonk, F.: Teaching logic using a state-of-the-art proof assistant. In: Proceedings of the International Workshop on Proof Assistants and Types in Education (PATE 2007), June 25, pp. 37–50 (2007)

Goshi et al 2001. Goshi, K., Wray, P., Sun, Y.: An Intelligent Tutoring System for Teaching and Learning Hoare Logic. In: Proceedings of the Fourth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2001), pp. 293–297 (2001)

Lesta and Yacef 2002. Lesta, L., Yacef, K.: An intelligent teaching assistant system for Logic. In: Cerri, S.A., Gouardéres, G., Paraguaçu, F. (eds.) ITS 2002. LNCS, vol. 2363, pp. 421–431. Springer, Heidelberg (2002)

Logic Toolbox. Logic Toolbox, `http://philosophy.lander.edu/~jsaetti/Welcome.html` (accessed, November 2006)

Lukins et al 2002. Lukins, S., Levicki, A., Burg, J.: A tutorial program for propositional logic with human/computer interactive learning. In: SIGCSE 2002, pp. 381–385. ACM, New York (2002)

Luger, 2004. Luger, G.F.: Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 5th edn. Addison-Wesley, Reading (2004)

Neller et al 2006. Neller, T., Markov, Z., Russell, I.: Clue Deduction: Professor Plum Teaches Logic. In: Proceedings of the 19th International FLAIRS Conference (FLAIRS 2006), Melbourne Beach, Florida, May 11-13, pp. 214–219 (2006)

Plato. Plato, `http://www.utexas.edu/courses/plato/info.html` (accessed, February 2009)

Russell and Norvig 2003. Russell, S., Norvig, P.: Artificial Intelligence: a modern approach, 2nd edn. Prentice Hall, Upper Saddle River (2003)

Wildenberg and Scharff 2002. Wildenberg, A., Scharff, C.: OLIVER: an OnLine Inference and VERification system. In: The Proceedings of the 32 ASEE/IEEE Frontiers in Education Conference (2002)

Document Design with Interactive Evolution

Juan C. Quiroz, Amit Banerjee, Sushil J. Louis, and Sergiu M. Dascalu

Abstract We present human guided evolution of brochure documents. The user interacts with a genetic algorithm, which evolves placeholders, each placeholder represented with one of three shapes: (1) ellipse, (2) rectangle, and (3) rounded rectangle. The user guides the evolutionary process by evaluating a small subset taken from a large population of documents. Along with the subjective user input, individuals in the population of the genetic algorithm are evaluated on a set of objective heuristics for document design. We present pretest results, including an evaluation of the tool and documents created.

1 Introduction

Design is a fundamental, purposeful, pervasive and ubiquitous activity and can be defined as the process of creating new structures characterized by new parameters, aimed at satisfying predefined technical requirements. It consists of several phases, which differ in details such as the depth of design, kind of input data, design strategy, procedures, methodology and results [12]. Usually the first stage of any design process is the preliminary or the conceptual design phase, followed by detailed design, evaluation and iterative redesign [3]. We are interested in supporting the creative conceptual design phase by allowing for the exploration and the ability to assess alternative design solutions. In the work presented in this paper we look at supporting a simple design task of creating a brochure document, where we specify a small set

Juan C. Quiroz · Sushil J. Louis · Sergiu M. Dascalu
University of Nevada, Reno,
Reno, NV 89557, USA
e-mail: quiroz,sushil,dascalus@cse.unr.edu

Amit Banerjee
The Pennsylvania State University, Harrisburg
Middletown, PA 17057, USA e-mail: aub25@psu.edu

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 309–319.
springerlink.com

of requirements, but leave the rest of the problem to be defined by the user through exploration of designs.

We use genetic algorithms (GAs) to explore document designs. GAs are search algorithms based on the principles of genetics and natural selection [6]. GAs consist of a population of individuals, where each individual is a potential solution to the problem being solved. Assigning a fitness to individuals in the population plays a central role in the search success of a GA. However, there are times, such as when we are dealing with subjective criteria, when it is difficult if not impossible to determine the fitness of individuals in a GA population [13]. Interactive genetic algorithms (IGAs) differ from GAs in that the objective fitness evaluation is replaced with user evaluation, thus allowing for the user to guide the evolutionary process through subjective input [13]. By doing so, IGAs incorporate human knowledge, emotion, intuition, and preference into GAs. Figure 1 shows the process involved in an IGA.

Effective IGAs have to overcome several issues. GAs usually rely on large population sizes running for many generations, but asking a user to make hundreds or thousands of choices is likely an unrealistic task. A user would rapidly fatigue and/or lose interest. Furthermore, because of the subjective nature of human input, it can lead to users changing their goals through the IGA run, leading to noisy fitness landscapes – which coupled with user fatigue can result in suboptimal solutions [8].

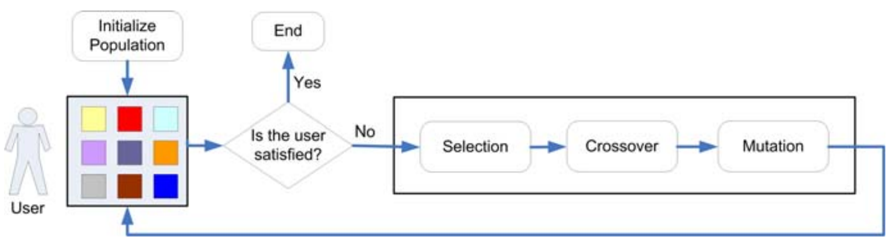


Fig. 1: Interactive genetic algorithm

The work presented here is specifically interested in the development of a tool based on human guided evolution, which would allow users to create documents with an IGA. Document templates can be found on various websites and some come as default installs in popular programs such as word processors. However, the support for users to modify an existing template to suit their particular needs is lacking. Many times the options given to users is to use an existing template or to start from scratch. GAs have been used for document design, but not with direct user feedback [5]. Our work is different in that not only is the evolution of documents driven by subjective criteria, but also by objective criteria taken from document style guidelines. We further address how the affordance of IGAs affects the user experience and the user's ability to create satisfactory document designs.

The remainder of this paper is organized as follows. Section 2 discusses background information on the use of GAs to create document layouts. Section 3 illustrates how the document design problem was mapped to IGAs. In Section 4 we present a discussion and evaluation of the system and created documents. Section 5 takes a look at affordance as a key challenge for IGAs. Finally, Section 6 presents our conclusions and directions for future work.

2 Background

Geigel and Loui present the use of a GA to evolve album page layouts [5]. A collection of pictures are loaded into the system, which then separates the pictures into various pages. Each image in a page is encoded using a 4-tuple: (1) x coordinate, (2) y coordinate, (3) scaling factor, and (4) rotation angle. These floating point values are arranged in an array, resulting in a chromosome of floating point values instead of the traditional binary string representation used in GAs. User interaction is limited to the specification of user preferences at the start of the evolutionary process, including attributes such as balance, spacing, emphasis, and unity [5]. The user specifies a value between 0 and 1 for each of the preference attributes, with 0 representing a criteria not considered important for the resulting album page layouts and 1 representing an important attribute. During fitness evaluation each of these attributes is computed by analyzing the images in a page, with the resulting values scaled by the preference values specified by the user. In our system the user does not specify such preference values for attributes. Instead, the user directly guides the evolutionary process by participating during each generation by selecting the best document in the population.

3 Interactive Evolution of Documents

In our previous work we applied interactive evolution to floorplanning [2, 10]. For floorplanning we used a simple graphic representation to represent rooms in an apartment. The boundary of the floorplan was depicted using a rectangular shape. We then recursively created a tree, by subdividing rooms either vertically or horizontally (starting with the entire floorplan as a single room and as the root node). In the end, the resulting floorplan (originally a single rectangle) ended up subdivided into smaller areas, each area representing a different room in the floorplan. Because of the way we originally coded the floorplan representation, only rectangular or square rooms could be created.

We wanted to build on the framework we had used for floorplanning. We figured that by taking the rectangular rooms created, and allowing for various transformations to be applied to each room while drawing each room respectively, we would be able to get some degree of overlap and interesting shape combinations. If we further allowed for various shapes to be drawn, such as drawing a circle instead of a rectangle for where a room should be, we would be able to achieve an even greater degree of variation and possibly creative and interesting documents.

We allowed for three types of shapes to be drawn: (1) rectangles, (2) ellipses, and (3) rounded rectangles. Each of these shapes could be scaled respectively along the x or y axis, by up to 10%, either scaling up or scaling down. These changes allow for the original floorplan representation to be transformed into a collection of shapes, where either each shape can represent a placeholder for content (such as text or an image) or where the collection of shapes could represent a background design.

We initialize the documents similar to the floorplans. Except that once we know the allocation of the rooms, we assign the shapes to one of four quadrants, based on the shapes' locations, using the shape's center as the point of reference. This results in a quad-tree of depth one.

3.1 Fitness Evaluation

The fitness of every individual in the population was assigned by combining objective and subjective heuristics. Subjective heuristics consisted of comparing attributes of individuals in the population with attributes of the document selected as the best by the user. Objective heuristics consisted of coded measures, which would help guide the evolutionary process towards documents with proper style and design guidelines, but which also reflected the user's preferences.

Harrington et al. present a set of aesthetic measures for automated document layout, which we use as criteria for the objective fitness evaluation [7]. We evaluate the use of white space, the degree of overlap in the shapes, and the spatial balance. To evaluate spatial balance, first we bisect the page vertically, computing how much area is covered in the left half versus the right half. This gives us a measure of vertical balance. Next we bisect the page horizontally, computing how much area is covered in the top half versus the bottom half, giving us a measure of horizontal balance. We combine horizontal and vertical balance to give us an objective measure of spatial balance of a document.

We evaluate white space by computing the area of every shape in a document, adding these areas together, and dividing this total by the area of the entire document. For example, if there are a lot of large shapes in a document, then the total sum of all shape areas will be close to the area of the entire document. The amount of overlap is computed by determining whether any two shapes overlap, and if so, figuring out the area of the overlap region. This is computed for every shape in the document. This objective measure enforces a low degree of overlap between shapes. Finally, the three objective criteria are averaged together.

What if a user likes documents which conflict with any of the objective criteria? While the objective criteria are meant to allow a user to create documents which adhere to proper style guidelines for documents, a user might want to create a document which does not follow all of the objective criteria. A user might want or need to create a document in which a high degree of overlap is desired. This is where the subjective criteria comes into play. The subjective and objective criteria, which might be contradictory at times, are optimized using Pareto optimality.

3.2 Generational Algorithm

The interaction between the user and the GA is key in IGAs. Instead of asking the user to evaluate all individuals in the IGA's large population (population size of 100), we select a subset of size nine from the large population, and display this to the user to be evaluated [11]. In case the user does not find any satisfactory individuals, the user has the option to scroll down on the user interface and view the rest of the population. We introduce a visible gap and a label of "Rest of Population" to make it clear to the user that viewing the rest of the population is optional, since having to view all individuals in the population over many generations can lead to user fatigue. We allow the user to guide the evolutionary process by selecting the best individual in the population, with the user having the choice of picking from the subset, or by selecting any other individual in the population.

The selection of the best individual is used to interpolate the fitness of every other individual in the population (for further details see [11, 2]). However, the manner in which offspring are generated from the population can have drastic effects in the behavior of the IGA, which can range from many diverse individuals to rapid convergence to individuals similar to the user's selected best. We explore three generational algorithms. The first one used was the Non-dominated Sorted multi-objective Genetic Algorithm, abbreviated as NSGA-II [4]. The NSGA-II creates fronts of non-dominated individuals, where within a front none of the individuals are any worse than any other individual across all optimization criteria. All individuals within a front are said to have the same rank. We select parents by using the crowded distance tournament operator. We pick two individuals to participate in the tournament, and we select the individual with the higher rank to be part of the mating pool. In case the two individuals have the same rank, and consequently belong to the same front, then the crowded distance of both individuals is computed, and we select the individual with the highest crowded distance to be part of the mating pool. This translates to the individual being in a less crowded region of the front and hence, the crowded distance selection favors the most diverse individuals within a front.

The second generational algorithm was the same as NSGA-II, except that every offspring was generated by selecting an individual from the population and mating it with the user selected best. The third generational algorithm generates offspring by applying mutations to the user selected best individual.

3.3 Customizing

In order to allow users to be able to create a document from the tool, we added support for customization of the evolved document designs. Any document can be edited by the user, however in the current version we do not allow changes made by the user to a document to be inserted back into the IGA population. Customizing opens a new window, where the document is displayed on a larger scale, allowing the user to appreciate detail and to facilitate editing. We support for the user to move any of the shapes around, to add text to one of the existing shapes, to overlay an image on a shape, to resize a shape, to add free floating textboxes and images (not attached to one of the shapes), and to delete any shape, text, or image. The user can also change either the color of any individual shape, or one of the predefined color schemes can be selected. If a color scheme is selected, the colors in the scheme are assigned sequentially to the shapes in the document. While this functionality is limited compared to the full breadth of options that come with tools such as Photoshop or a word processor, it is meant to allow users to be able to visualize how content would look on one of the evolved documents. After customization, the user also has the option to save the current document as an image.

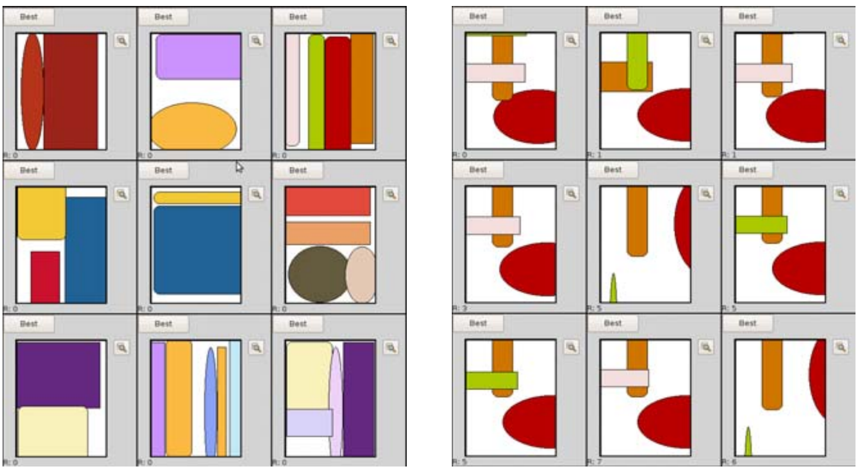
4 Results

Figure 2 shows a sample snapshot of the user interface at the beginning of the process (generation 0) and after 10 generations of user input. The IGA process begins with documents created randomly, as shown on the left side of the figure. Through feedback, the user can steer the evolutionary process in various ways, ranging from documents where all shapes (which stand for place-holders) are circular and overlapping, to documents with a small degree of overlap and lots of white space. As the IGA population converges (with the convergence rate depending on the generational algorithm) the focus of the search changes to fine-tunning, with documents varying by a small degree as seen on the right side of the figure. Some other examples of brochures created with our IGA tool are shown in Figure 3.

As an initial pretest, we had three participants test the tool. The participants were given the task of creating a brochure which advertised a new minor in interactive digital games to be offered by the computer science department. The requirements given to the participants were that the brochure was to include the following elements: (1) a header; (2) at least one paragraph, allowing for text to be distributed over the brochure as desired; and (3) at least two images. We interviewed the participants afterwards to get their thoughts on the usability of the system.

Due to the small sample, we cannot generalize the feedback given, but it allowed us to find areas that needed further work and had to be addressed before a full set of user studies. With regards to the generational algorithm, the participants preferred the two generational algorithms based on the NSGA-II. The participants found the generational algorithm which used mutations on the user selected best individual to be too sporadic, with too many random changes introduced from generation to generation. We attribute this to having used too high a mutation rate. However, the dilemma is finding the right mutation rate which would make this generational algorithm intuitive to users, since a too low of a mutation rate would not allow the user to effectively explore document designs.

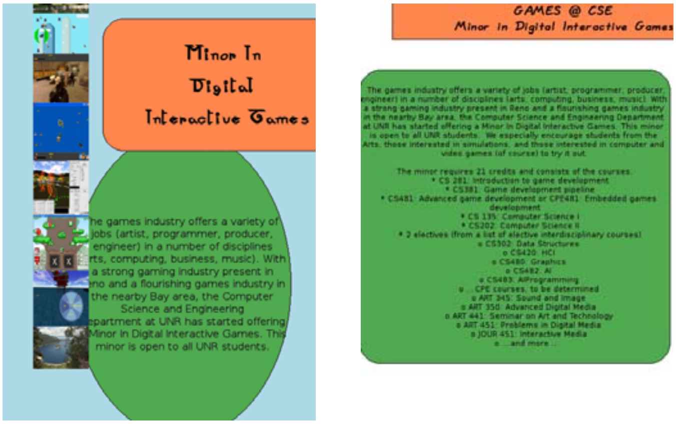
All three users were able to create brochures which met the given requirements as shown in Figure 4. The users liked the ability to explore alternative designs, but found the interaction with the IGA to be too limiting at times. For example, the users could not modify a document and subsequently inject it back into the population. There were many times when a document design was close to the one desired by the participants, but through the picking the participants were not able to fine tune it to what they desired.



(a) Generation 0 (b) Generation 10

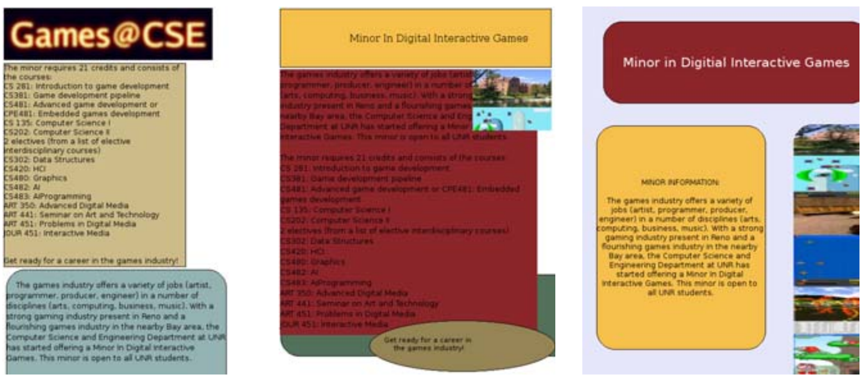
Fig. 2: Brochure templates as displayed to the user for evaluation.

All three participants were asked whether they believed it would have been faster for them to create the brochure from scratch instead of using the tool, and all three agreed. Such a statement is understandable given the current limitations in the systems' functionality. However, the system is meant as more than a simple document creator, and all three participants recognized that and expressed positive feedback in the ability to view and assess many document designs in a matter of minutes, which helped them build a conceptual model of the brochure they wanted. A user with a clear set of goals, and who has a clear conceptual model of what he/she wants, might benefit from just using a standard design tool such as a word processor or a drawing system to create a document. However, when the requirements are open-ended, and if the user has to create a conceptual model for the given requirements, then the concept of this tool would be useful. The IGA would allow the user to explore many



(a) Brochure 1 (b) Brochure 2

Fig. 3: Sample brochures.



(a) User 1 (b) User 2 (c) User 3

Fig. 4: Brochures created by users.

options, exposing the user to many possibilities which the user would not have come up with on his/her own, especially in a short period of time.

While evolutionary computation is a powerful technique, the interaction with a user presents many challenges and limitations. In the next section we propose that the main challenges faced by researchers in creating IGA tools for end-users stem from affordance issues.

5 Affordance of IGAs

One of the challenges faced when designing an IGA is affordance. Affordance, as discussed by Don Norman in his book The Design of Everyday Things, means that the designer must make “appropriate actions perceptible and inappropriate ones invisible” [9]. It is key for the designers to provide a conceptual model of the system, otherwise users end up making up their own conceptual model, which is bound to be erroneous. For example, a properly designed door knob lends its affordance to turn, just like a button presents the affordance of pushing.

We face such a dilemma when designing an IGA, especially an IGA not to be used by experts, but by end-users to create documents or brochures. If a user was asked to create a document, specifically a brochure for a new major or minor being offered by a department at a university, it would be unrealistic to expect the user to create such a brochure in a matter of minutes, even half an hour would be an unreasonable amount of time. However, spending 30 minutes in front of an IGA doing evaluation picks gets boring and frustrating quickly. Most IGA research places an implicit blame on the user, suffering from fatigue, where instead a tireless computer would be able to find satisfactory solutions by participating in evolution over many generations and evaluating hundreds of individuals. The problem is that most IGA applications suffer from conceptual models targeted to expert users, and in some cases the conceptual model makes sense only to the researchers who made the particular IGA application.

The participants in our pretest faced some of these issues while using our IGA. They found the interaction with the IGA (picking only the best document) to be limiting. However, this was done in order to mitigate user fatigue, since previous work has shown that too much feedback results in user fatigue [13]. Another challenge is conducting the IGA session in a manner which allows the user to build a correct conceptual model of the IGA. To an end-user, the concept and understanding of the workings of a GA should not be integral to understanding the system. However, the user must be able to develop an intuition of what is occurring from generation to generation. This introduces a final road-block, which is that if an IGA is failing to work properly, it is hard for the end-user to realize this. An error on a crossover or a mutation operator could introduce enough error in the overall process, resulting in either premature convergence or convergence to local optima. An end-user could be led to believe that he/she is not picking the right solutions, or that the problem is “too hard” and cannot be solved using evolution. IGA applications suffer from some or all of these challenges, and it is critical for the deployment of IGA based tools to end-users to address these issues, in order to fully embrace the computational power and benefits of IGAs.

6 Conclusions and Future Work

We have presented an approach to document design based on human guided evolution. Users guide the creation of brochure documents by picking the brochure they like the best from a subset displayed from a large population size. The IGA combines the user feedback with objective criteria taken from document design guidelines. We had three users successfully create brochure documents which met a set of given requirements, and with each brochure differing based on each of the users' preferences. We believe that a limiting factor of IGAs in general is affordance issues.

Interactive evolutionary design of documents is a promising direction of future work. First, we would like to conduct user studies in order to assess the utility of the tool. For example, we want to test whether participants are indeed capable of creating documents from scratch faster than by using our IGA tool. We also plan to determine whether the quality of the documents created with our IGA tool is greater than those created from scratch by the participants, using evaluation criteria from the Creative Product Semantic Scale [1].

Finally, we are also interested in supporting the exploration of alternative designs of an already created document by a user. Instead of starting the evolutionary process with documents created from scratch, the user would create a simple document, and then load this into the IGA. Thus, we can use this initial document as the seed from which to start the evolutionary process, allowing the user to search alternatives in fewer generations, and enabling him/her to visualize the desired content on the brochures, instead of evolving placeholders as in the current version. Another alternative would be to allow the user to specify the elements which must be included in the document, such as a set of images and text blocks. The IGA could then work on layouts and transformations on the elements specified by the user.

Acknowledgements. We thank the study participants for their time. This work was supported in part by contract number N00014-0301-0104 from the Office of Naval Research and the National Science Foundation under Grant no. 0447416.

References

1. Andrews, J., Smith, D.C.: In search of the marketing imagination: Factors affecting the creativity of marketing programs for mature products. Journal of Marketing Research 33, 174–187 (1996)
2. Banerjee, A., Quiroz, J.C., Louis, S.J.: A Model of Creative Design Using Collaborative Interactive Genetic Algorithms, pp. 397–416. Springer, Heidelberg (2008)
3. Bentley, P.J., Wakefield, J.P.: Conceptual evolutionary design by genetic algorithms. Engineering Design and Automation Journal 3, 119–131 (1997)
4. Deb, K.: Multi-Objective Optimization Using Evolutionary Algorithms. John Wiley and Sons, Chichester (2001)
5. Geigel, J., Loui, A.: Using genetic algorithms for album page layouts. IEEE Multimedia 10, 16–27 (2003)

6. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Professional, Reading (1989)
7. Harrington, S.J., Naveda, J.F., Jones, R.P., Roetling, P., Thakkar, N.: Aesthetic measures for automated document layout. In: Proceedings of the 2004 ACM symposium on Document engineering, pp. 109–111. ACM, Milwaukee (2004)
8. Llorà, X., Sastry, K., Goldberg, D.E., Gupta, A., Lakshmi, L.: Combating user fatigue in iGAs: partial ordering, support vector machines, and synthetic fitness, pp. 1363–1370. ACM, Washington (2005)
9. Norman, D.A.: The Design of Everyday Things. Basic Books (2002)
10. Quiroz, J.C., Banerjee, A., Louis, S.J.: IGAP: interactive genetic algorithm peer to peer, pp. 1719–1720. ACM, Atlanta (2008)
11. Quiroz, J.C., Louis, S.J., Dascalu, S.M.: Interactive evolution of XUL user interfaces, pp. 2151–2158. ACM Press, London (2007)
12. Renner, G.: Genetic algorithms in CAD. Computer-Aided Design 35, 707–708 (2003)
13. Takagi, H.: Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation. Proceedings of the IEEE 89, 1275–1296 (2001)

Enhancing 2D environments with 3D data input

Constantinos Patsakis, Nikolaos Alexandris, and Elina Flerianou

Abstract In this paper we study how we can enhance the two dimensional environment that a user has, using a three dimensional input device. We mainly focus on the Wiimote device developed for Nintendo Wii console, but the results can be generalized for other 3D input devices.

1 Introduction

User interaction with computer has as primary input in most cases motion, it can be either the motion of the hands on the keyboard, or the movement of the mouse. Modern user interfaces succeed in offering better user interaction experience by getting more intuitive interaction with the user. Two very good examples are iPhone from Apple [3] and Opera browser .

iPhone uses a touch screen as many other devices, yet the intuitive use of two fingers creates a better user interaction, the user can use his fingers to rotate a photo as he would in real life, or even zoom in or zoom out depending on his movement while on the Internet or viewing photos or documents. Furthermore, iPhone uses the so called accelometers in order to track the tilting of the device, so the interface can be rotated automatically when the user rotates the device etc. On the other hand Opera browser uses gestures, motion patterns of mouse, in order to complete some tasks that the user makes often, like back, forward, create bookmarks etc. This feature is so good that other popular browsers like Firefox [5], Safari [3] and Chrome are adopting it.

If the environment can get more data from the free movement of the user, then we can achieve better user interaction. With this motive we tried to explore something more intuitive, 3D user movement. In order to make this possible, we used the

University of Piraeus

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 321–326.
springerlink.com

Wiimote device (Figure 1) from Nintendo Wii console to get 3D data from the user, while using an ordinary 2D user environment.

In this paper we give the basic principles about Wiimote and its use on PCs, we present the experiment that we conducted and it's results.

2 Wiimote essentials

The Wiimote is the main device for interacting with the user of Nintendo Wii, a very successful game console, developed in 2006. Wiimote is a wireless device that has bluetooth and IR sensors embeded, which provides audio and rumble feedback as well. Moreover, it has some standard buttons and triggers attached. Finally it can be used in parallel with other Wiimotes in order to provide more motion data or even be expanded with other devices like nunchuk or Wii motion plus.

Fig. 1: The Wiimote.

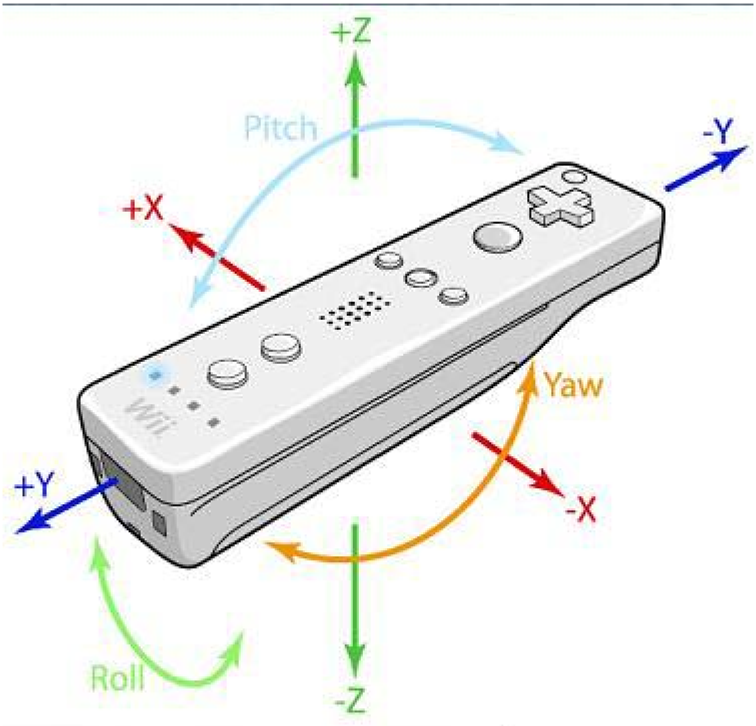


Fig. 2: Wiimote coordinates.Image courtesy of www.osculator.net

In order to get close to the coordinate system in Figure 2 we used a low level API, Brian Peek's [1] and used the following transformations in order to get a smoother cursor movement in the screen.

- $x = x_{wii} - c_x$
- $q_x = x|x|$
- $x' = x + q_x * SamplePeriod * SpeedGain$
- If $x' > 1$ then $x' = 1$
- If $x' < -1$ then $x' = -1$
- $x_{new} = \frac{Screen_x}{2}(x' + 1)$

The exact calculations are made for coordinate y, by replacing the coordinate x with y. The new position that the cursor has on screen is (x_{new}, y_{new}). For coordinate z we kept the above transformations only that we had:

- $z = z_{wii} - c_z$
- $q_z = z|z|$
- $z_{new} = z + q_z * SamplePeriod * SpeedGain$

since monitors do not have z coordinates and we wanted to track all the z changes.

3 Conducting the experiment

In our experiments we made a simple Visual Basic application with the user interface depicted in Figure 3. The application picks twenty random questions from a database of total 100 and provides the user with four possible answers to each one. The user just has to click with a Wiimote on what answer he thinks is most appropriate and then click on the next button. The questions that were used, were from a national skills test in Greece and had to do with common knowledge, things that anyone should know. The interface is quite minimalistic, it has one button, for next question, four radio buttons for the answers and one label for the question.

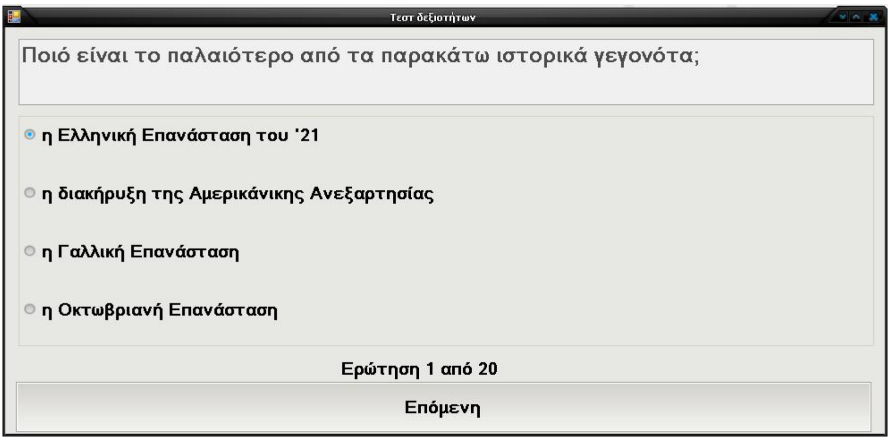


Fig. 3: The user interface of the application

Based on the fact that almost anyone should know the answers, the application made some comments when the users didn't select the correct answer, in order to irritate them with a nagging screen, possibly causing more stress. Throughout the skill test the application records in a file the x,y,z coordinates to create a three dimensional profile of the user's movement. The application also recorded whether the user answered correctly the question or not.

The sample used for this experiment was 50 people that hadn't been tested on the national test, so that they couldn't remember the test questions and all of them having small or big experience with computers, in order to understand the concept of using a Wiimote tracking device instead of mouse or other tracking device.

4 Results

Almost all users had no difficulty in adopting the new user interaction and were capable of using it with very good precision after almost one or two minutes. Of course all pointed out that it cannot yet substitute the mouse, as it's movement is more smooth and more precise. Furthermore, most users were surprised by the fact that the application criticized them when they were wrong and not giving any sign when they were right. The comment that was made to all of them was that it was knowledge that should be taken for granted.

In figure 4 we can see a typical x,y movement of the cursor from the user.

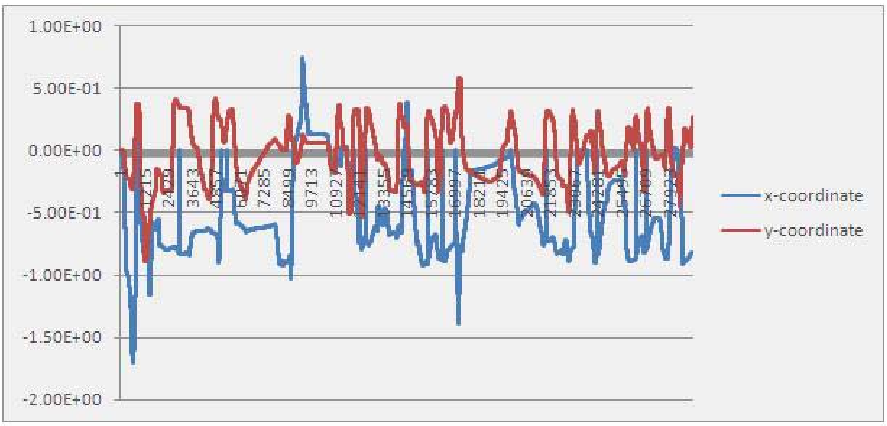


Fig. 4: xy Coordinates

In figure 5 we can see a typical z movement of the Wiimote from the user. In this figure we must note that in most peaks, the user had answered wrong.

It is obvious that the application caused some stress on the user, leading to changes in the way he moves the tracking device. More precisely the movement appeared to be more smooth in xy and less smooth in z axis.

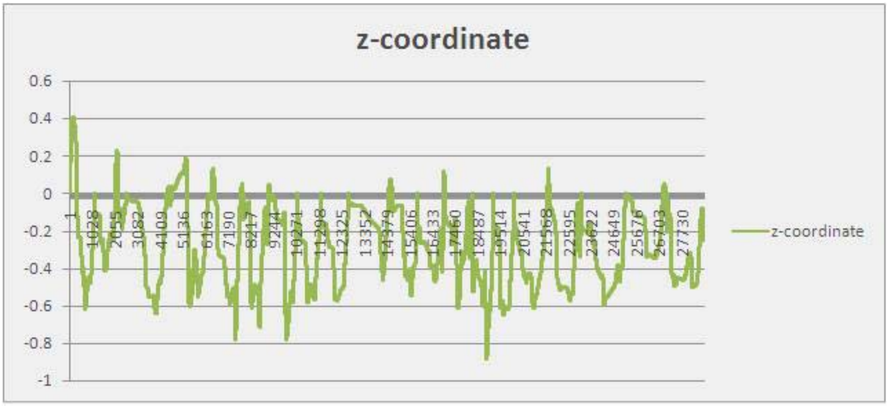


Fig. 5: z Coordinates

5 Conclusions

It is sure that 3D tracking devices can be used in 3D virtual environments, offering much more creative space for the user, yet in most 2D environments they can offer many more facilities, than usual 2D tracking devices, like mice.

The conducted experiment can lead to better user interfaces, that are more sensitive to users' reactions during their use. Of course the user does not need to be irritated by the application that he is using, never the less sudden changes in the z axis can mean that the user is not feeling well with the response of the application. Getting such data, an application can change it's behavior by using for example wizards instead of menus, in order to facilitate the user at that instant, leading him to his target instead of leaving him alone on stressful moments. The application of course keeps it's full features when used in other instances, while having a "stress mode".

References

1. Brian Peek's API, `www.coding4fun.com`
2. Nintendo, `www.nintendo.com`
3. Apple, `www.apple.com`
4. Opera browser, `http://www.opera.com/`
5. Mozilla Firefox, `www.mozilla.com/firefox/`
6. Chrome, `www.google.com/chrome`

Determining a Discrete Geometric Random Sum in Computational Intelligence for Supporting a Complex Human-Computer Interaction System under a Major Risk

Constantinos T. Artikis and Panagiotis T. Artikis

Abstract A type of geometric random sum of discrete, independent and identically distributed random variables taking values in the set of nonnegative integers is a useful analytical tool of computational intelligence for describing and analyzing the evolution of a complex human-computer interaction system under a major risk. The present paper concentrates on the determination of a particular case of such a geometric random sum and the application of the determined discrete geometric random sum in realizing the most probable and best scenario for the evolution of a complex human-computer interaction system under a major risk.

1 Introduction

Computational intelligence is generally recognized as a scientific discipline with very interesting practical and theoretical applications. The contribution of computational intelligence to designing of complex human-computer interaction systems is of particular importance. Generalization, discovery, association and abstraction are characteristics of computational intelligence [5, 9]. Moreover, cindynics is a scientific discipline providing a strong conceptual framework for investigating the behaviour of complex human-computer interaction systems under conditions of risk. Cindynics makes use of concepts and methods from various disciplines in order to

Constantinos T. Artikis
Department of Informatics, University of Piraeus,
80 Karaoli & Dimitriou Street, 18534, Piraeus, Greece
e-mail: ctartikis@gmail.com

Panagiotis T. Artikis
Section of Mathematics and Informatics, Pedagogical Department
University of Athens, 20 Hippocratous Str. 10680 Athens, Greece
e-mail: ptartikis@gmail.com

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 327–337.
springerlink.com

develop effective tools for treating situations of major risks and crises [8]. From a theoretical and practical point of view, stochastic models can be useful for creating powerful connections between the discipline of computational intelligence and the discipline of cindynics. Such connections can be productive for both disciplines. This means that research activities must be undertaken for developing stochastic models supporting connections between computational intelligence and cindynics. The present paper takes advantage of fundamental components of the cindynic framework and characteristics of computational intelligence for formulating a stochastic model suitable for design and operation of complex human-computer interaction systems. The formulated stochastic model is a particular case of the discrete geometric random sum considered in the last part of the present section. Let

$$\{X_n, n = 1, 2, \ldots\}$$

be a sequence of discrete, independent random variables with values in the set

$$\mathbf{N} = \{1, 2, \ldots\}$$

or in the set

$$\mathbf{N}_0 = \{0, 1, 2, \ldots\}$$

and following the same distribution with the random variable

$$X$$

having probability generating function

$$P_X(z) \tag{1}$$

We consider the discrete random variable

$$N$$

being independent of the sequence

$$\{X_n, n = 1, 2, \ldots\}$$

and following the geometric distribution with probability function

$$P(N = n) = pq^n,\ 0 < p < 1,\ q = 1 - p,\ n = 0, 1, 2, \ldots \tag{2}$$

The discrete geometric random sum

$$L = X_1 + X_2 + \ldots + X_N$$

is particularly important in theory and practice [7]. In general, an explicit analytical evaluation of the probability function of the above random sum is extremely difficult. In this case, it is suitable to evaluate the corresponding probability generating function. From (1) and (2) it is easily shown that the probability generating function of this random sum has the form

$$P_L(z) = \frac{p}{1 - qP_X(z)} \tag{3}$$

A stochastic derivation and an interpretation of the proposed discrete geometric random sum as a research tool of computational intelligence for describing and analyzing the evolution of a complex system under the occurrences of a major risk have been established by Artikis [4]. The present paper provides an extension of the results established by this author. More precisely, the paper concentrates on the implementation of two purposes. The first purpose is the determination of a particular case of the proposed discrete geometric random sum by formulating and solving a functional equation for probability generating functions. The second purpose is the interpretation of the determined discrete geometric random sum as a research tool of computational intelligence in supporting the realization of the most probable and best scenario for a complex human-computer interaction system under the occurrences of a major risk.

2 Definition and Stochastic Derivation of Discrete Renewal Distribution

The discrete renewal distribution is a strong tool of probability theory with applications in several practical disciplines. The present section considers the definition and the stochastic derivation of such a discrete distribution. We consider the discrete random variable

$$X$$

with values in the set

$$\mathbf{N}_0,$$

finite mean

$$\mu$$

and probability function

$$P(X = x) = p_x,\ x = 0, 1, \ldots$$

We also consider the discrete random variable

$$Y$$

taking values in the set

$$\mathbf{N}_0$$

and having probability function

$$P(Y=y)=q_y=\frac{1-\sum_{x=0}^{y} p_x}{\mu},\ y=0,1,2\ldots \tag{4}$$

Such a discrete random variable is said to follow the renewal distribution corresponding to the distribution of the random variable

$$X$$

or shortly the renewal distribution. Under certain conditions, the random variable

$$Y$$

denotes the residual lifetime to failure of an ageing component in a system. Upon failure, defective components are replaced by new components with lives being independent random variables distributed as the random variable

$$X.$$

In this case, the sequence of failure times is a renewal stochastic process in discrete time [6]. It is easily shown that

$$q_{y+1}<q_y\ y=0,1,\ldots$$

The above inequalities imply that the probability function of the renewal distribution has a unique mode at the point 0. If the probability generating function of the random variable

$$X$$

is

$$P_X(z),$$

then from (4) it follows that the probability generating function of the renewal distribution can be written in the form

$$P_Y(z)=\frac{1-P_X(z)}{(1-z)\mu},\ |z|\leq 1. \tag{5}$$

The above representation has been applied by Medgyessy [10] for establishing a characterization of discrete distributions defined on the set

$$\mathbf{N}_0$$

having a unique mode at the point 0. More precisely, he proved that the probability function of a discrete random variable

$$V,$$

taking values in the set

$$\mathbf{N}_0,$$

is unimodal with mode at the point 0 if , and only if, the corresponding probability generating function can be represented in the form

$$P_V(z) = \frac{1 - P_C(z)}{(1-z)\lambda},$$

where

$$P_C(z)$$

is the probability generating function of a uniquely determined discrete random variable

$$C$$

with values in the set

$$\mathbf{N}$$

and mean

$$\lambda = \frac{1}{P_V(0)}.$$

Theoretical properties and practical applications of the discrete renewal distribution have been established by several authors. A necessary and sufficient condition for the infinite divisibility of the discrete renewal distribution has been provided by Steutel [11]. Characterizations of the geometric distributions based on properties of failure rates of the random variables

$$X, Y$$

have been established by Gupta [6]. The canonical representation of the probability generating function of a discrete selfdecomposable random variable taking values in the set

$$\mathbf{N}_0$$

with the use of the probability generating function of the discrete renewal distribution has been established by Steutel and van Harn [12]. Some Poisson random sums incorporating discrete renewal random variables for investigating the discrete selfdecomposable random variables taking values in the set

$$\mathbf{N}_0$$

has been introduced by Artikis [1]. A characterization of the geometric distribution based on the equality in distribution of the random variables

$$X, Y$$

has been established by Wimmer and Kalas [14]. The formulation, investigation and applications of integral part models based on discrete renewal random variables have been considered by Artikis et al [2, 3]. The present paper makes quite clear that the existence of a unique mode at the point 0 for the probability function and the form of the probability generating function of the renewal distribution can contribute to the substantial extension of the practical applicability in computational intelligence of a discrete geometric random sum considered and stochastically derived by Artikis [4]. More precisely, it is shown that a particular case of such a discrete random sum can be of some practical importance for the description, analysis, selection and implementation of the most probable and best scenario for the evolution of complex systems under conditions of risk.

3 Determination of a Discrete Geometric Random Sum

The present section is devoted to the formulation and solution of a functional equation based on the probability generating function of the proposed discrete geometric random sum and the probability generating function of a discrete renewal distribution. We suppose that the discrete random variable

$$X$$

takes values on the set

$$\mathbf{N}$$

and has mean value

$$\mu = \frac{1}{p},\ 0 < p < 1$$

We also suppose that the renewal distribution corresponding to the distribution of the random variable

$$X$$

is infinitely divisible and the probability function of the discrete geometric random sum

$$L = X_1 + X_2 + \ldots + X_N$$

has a unique mode at the point 0. Moreover, we suppose that the above random sum and the random variable

$$Y,$$

following the renewal distribution corresponding to the distribution of the random variable

$$X,$$

are equal in distribution. From (3), (5) and the above assumptions we get the functional equation

$$\frac{1 - P_X(z)}{(1-z)1/p} = \frac{p}{1 - qP_X(z)} \tag{6}$$

It is easily shown that the functional equation in (6) can be written in the form

$$qP_X^2(z) - (1+q)P_X(z) + z = 0. \tag{7}$$

Since

$$P_X(z)$$

is a probability generating function, then from (7) we conclude that

$$P_X(z) = \frac{1 + q - \sqrt{(1+q)^2 - 4qz}}{2q} \tag{8}$$

is the unique probability generating function which satisfies the functional equation in (6). The probability generating function in (8) belongs to the random variable denoting the number of customers served during a busy period initiated by one customer for the M/M/1 queque having mean interarrival time

$$\frac{1}{q}$$

and mean service time 1 [13].

It is readily understood that the formulation and the solution of the above equation for probability generating functions determine a discrete geometric random sum with probability function having a unique mode at the point 0. Since the existence of this mode for the probability function of a discrete stochastic model, taking values in the set of nonnegative integers, substantially contributes to the applicability of such a model, then it seems to be of some importance to investigate the applications of the determined discrete geometric random sum in various practical disciplines. More precisely, the paper makes clear that the determined discrete random sum can be a strong analytical tool for computational intelligence and cindynics.

4 Extending the Applicability of a Discrete Geometric Random Sum

Formulation, stochastic derivation and interpretation in computational intelligence of a type of geometric random sum of discrete, independent and identically distributed random variables, taking values in the set of nonnegative integers have been established by Artikis [4]. A particular case of such a random sum is the random sum determined by the previous section. The present section of the paper extends the applicability of the determined discrete geometric random sum as a research tool of computational intelligence describing the most probable and best scenario for a complex system under the occurrences of a major risk. We consider the homogeneous Poisson process

$$\{N(t), t \geq 0\}$$

with intensity

$$\lambda$$

and the random variable

$$S$$

following the exponential distribution with parameter

$$\mu.$$

We suppose that the random variable

$$N(t)$$

denotes the frequency in the time interval

$$[0,t]$$

of a major risk threatening a complex system. We also suppose that the complex system implements a project in the time interval

$$[0,S]$$

The discrete random variable

$$N$$

denoting the frequency of the major risk in the time interval

$$[0,S]$$

follows the geometric distribution with probability function

$$P(N=n) = pq^n \ , \ n = 0,1,2,\ldots,$$

where

$$p = \frac{\mu}{\mu+\lambda} \tag{9}$$

Moreover, we suppose that the arbitrarily distributed discrete random variable

$$X_n, n = 1,2,\ldots$$

denotes the number of networks of the complex system destructed by the nth occurrence in the time interval

$$[0,S]$$

of the major risk. The discrete geometric random sum

$$L = X_1 + X_2 + \cdots + X_N$$

denotes the number of networks of the complex system destructed by

$$N$$

occurrences in the time interval

$$[0,S]$$

of the major risk [4]. If the random variables of the sequence

$$\{X_n, n = 1, 2, \ldots\}$$

are distributed as the random variable

$$X$$

with probability generating function

$$P_X(z) = \frac{1 + q - \sqrt{(1+q)^2 - 4qz}}{2q}$$

then the corresponding geometric random sum

$$L = X_1 + X_2 + \cdots + X_N$$

is of some particular practical importance for describing and investigating the evolution of the complex human-computer interaction system under the occurrences of the major risk. Such an importance is based on the existence of a unique mode at the point 0 for the probability function of the discrete geometric random sum determined by the previous section. This mode makes quite clear that ”zero destructed networks for the complex human-computer interaction system” is the most probable scenario of the occurrences of the major risk and hence the determined geometric random sum is a suitable stochastic model for studying the realization of such a scenario. Since the most probable scenario is also the best scenario for the evolution of the complex system under the occurrences of the major risk, then the determined geometric random sum seems to be useful for considering the behavior of such a system facing a crisis. More precisely, this random sum can be interpreted as an analytical tool of computational intelligence for facilitating intelligent behavior of a complex human-computer interaction system under the occurrences of a major risk [4]. It is readily understood that the probability of the most probable and best scenario is of extreme theoretical and practical importance for designing a complex system. From (3), (8) and (9) it follows that the probability of the most probable and best scenario or the probability of the event

$$L = 0$$

is given by

$$p = \frac{\mu}{\mu + \lambda}$$

This means that the intensity

$$\lambda$$

of the homogeneous Poisson process and the parameter

$$\mu$$

of the exponential distribution are constituent factors for investigating the probabilistic behaviour of a complex human-computer interaction system under the occurrences of the major risk. Since the homogeneous Poisson process is the most important counting process and the exponential distribution is the most frequently used distribution of time, then the probability of the most probable and best scenario of the occurrences of the major risk is structured on these two powerful tools of probability theory. It is quite obvious that the study of the probability

$$p = \frac{\mu}{\mu + \lambda}$$

as a function of the parameter

$$\mu$$

and the parameter

$$\lambda$$

can substantially contribute to investigating the evolution of a complex human-computer interaction system under the occurrences of the major risk. The results of the present section constitute a very good reason for determining other particular cases of the proposed geometric random sum of discrete, independent and identically distributed random variables taking values in the set of nonnegative integers as tools of computational intelligence and other areas of informatics, suitable for making predictions and decisions related to the intelligent behaviour of complex human-computer interaction systems operating under conditions of risk.

5 Conclusion

The theoretical contribution of the paper consists of determining a particular case of a geometric random sum of discrete, independent and identically distributed random variables taking values in the set of nonnegative integers. The determination of such a random sum is based on the formulation and solution of a functional equation for probability generating functions. Moreover, the practical contribution of the paper consists of interpreting the determined discrete geometric random sum as an analytical tool of computational intelligence for describing the most probable and best scenario for a complex human-computer interaction system operating under the occurrences of a major risk.

References

1. Artikis, T.: On certain discrete distributions. Mathematics Reports Toyama University 6, 1–6 (1983)
2. Artikis, T., Voudouri, A., Artikis, P.: Properties and applications in risk frequency reduction operations of an integral part model. Computers and Mathematics with Applications 42, 211–218 (2001)
3. Artikis, P.T., Artikis, C.T., Fountas, C.E., Hatzopoulos, P.: Discrete renewal and selfdecomposable distributions in modelling information risk management operations. Journal of Statistics & Management Systems 9(1), 73–85 (2006)
4. Artikis, C.T.: Formulating discrete geometric random sums for facilitating intelligence behavior of a complex system under a major risk. In: Tsihrintzis, G.A., Jain, L.C. (eds.) Multimedia Services in Intelligent Environments-Integrated Services. Studies in Computational Intelligence. Springer, Heidelberg (to appear, 2009)
5. Engelbrecht, A.: Computational Intelligence: An Introduction. Wiley, Chichester (2007)
6. Gupta, R.: On the characterization of survival distributions in reliability by properties of their renewal densities. Communications in Statistics-Theory and Methods A 8, 685–697 (1979)
7. Johnson, N., Kotz, S., Kemp, A.: Univariate Discrete Distributions, 2nd edn. Wiley, New York (1993)
8. Kervern, G.: Latest Advances in Cindynics, Economica (1994)
9. Konar, A.: Computational Intelligence: Principles, Techniques and Applications. Springer, Heidelberg (2005)
10. Medgyessy, P.: On the unimodality of discrete distributions. Periodica Mathematica Hungarica 2, 245–257 (1972)
11. Steutel, F.: Preservation of Infinite Divisibility under Mixing and Related Topics, Math. Centre Tracts 33, Math. Centre, Amsterdam (1970)
12. Steutel, F., van Harn, K.: Discrete analogues of selfdecomposability and stability. The Annals of Probability 7, 893–899 (1979)
13. Takács, L.: A generalization of the ballot problem and its application to queues. Journal of the American Statistical Association 57, 327–337 (1962)
14. Wimmer, G., Kalas, J.: A characterization of the geometric distribution. Tatra Mt. Math. Pub. 17, 325–329 (1999)

Usability Factors for an Intelligent Tutoring System on Computer Programming

Konstantina Chrysafiadi and Maria Virvou

Abstract In this paper we describe factors for designing a usable and eventually useful Intelligent Tutoring System (ITS) for teaching a programming language. Our model is called TIP-4U (Tutoring Intelligence on Programming - 4 components' Usability). We combine the specific knowledge that has been obtained about the Psychology of Programming (PoP) and the results of empirical studies with the principles of Human Computer Interactions (HCI) systems, in order to define usability factors for the four components of an ITS. The factors of our model are attached to the components of an ITS and underline the role of the modules in combination with the particular characteristics of tutoring a programming language.

1 Introduction

In the recent years there is a growing interest in the use of computer technology in the educational field. Successful teaching demands user-friendly interface, multimedia and hypermedia techniques, adaptivity to individual students' needs and reasoning abilities. Adaptivity is achieved using the technology of Intelligent Tutoring Systems (ITS). Indeed, a number of successful evaluations of ITSs [1][2] have shown that such systems can be effective in improving learning by increasing the students' motivation and performance in comparison with traditional instructional methods.

Taking into account the fact that the construction of an ITS is complex and time consuming, as well as the fact that the usability is recognized as an important quality factor of any software and system, we have to focus on factors that make an ITS usable and eventually useful. An ITS apart from the interface module, includes the knowledge module, the student module and the tutor module. So, a large number

Department of Informatics, University of Piraeus,
80 Karaoli & Dimitriou St., 18534 Piraeus, Greece
e-mail: mvirvou@unipi.gr;kchrysafiadi@yahoo.gr

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 339–347.
springerlink.com

of people with different professions, such as programmers, instructors, experts of a specific domain, has to participate in the development of an ITS. Therefore, the usability of an ITS includes factors from the fields of HCI, pedagogy and psychology. In this paper we describe the TIP-4U (Tutoring Intelligence on Programming - 4 components' Usability), that includes factors that make an ITS for teaching a programming language usable. We have selected this knowledge domain because there is a significant scientific interest in the effective tutoring of programming courses [3][4][5]. Our model aims to combine the specific knowledge that has been learned about the Psychology of Programming (PoP) and the results of empirical studies with the principles of Human Computer Interactions (HCI) systems.

2 Related Work

2.1 Intelligent Tutoring Systems

An intelligent tutoring system (ITS) is any computer system that contains some intelligence and can be used in learning [6], providing direct customized instruction or feedback to students, i.e. without the intervention of human beings, whilst performing a task [7]. Intelligent tutoring systems consist of four different subsystems or modules: the interface module which provides the means for the student to interact with the ITS, the domain module which contains a description of the knowledge or behaviours that represent expertise in the subject-matter domain the ITS is teaching, the student module which uses a student model containing descriptions of student knowledge or behaviours, including his or her misconceptions and knowledge gaps, and the tutor module which receives any mismatch between a student's behaviour or knowledge and subsequently takes corrective action, such as providing feedback or remedial instruction. To be able to do this, it needs information about what a human tutor in such situations would do. Thus, building an ITS needs careful preparation in terms of describing the knowledge and possible behaviors of experts, students and tutors.

2.2 Usability

Usability addresses the relationship between tools or systems and their users. It depends on a number of factors including how well the functionality fits user needs, how well the flow through the application fits user tasks, and how well the response of the application fits user expectations. Nielsen defines usability more specifically in five attributes [8]:

- **Learnability:** users must be able to learn the system easily and shortly.

- **Efficiency of use:** it examines how the system supports users to carry out tasks in order to increase productivity.
- **Memorability:** interfaces which facilitate users to recall how to carry out infrequent tasks and use the system after a certain period of time has elapsed since its implementation.
- **Few and non-catastrophic errors:** the system must prevent users from making critical errors and provide recovery from errors.
- **Subjective satisfaction:** the level of users' satisfaction.

Fyrthermore, Shneiderman sums up usability in the following "five measurable human factors" [9]: speed of task performance, user error rate, subjective satisfaction, time to learn and retention over time. Preece et al [10] supports that "usability is generally regarded as ensuring that interactive products are easy to learn, effective to use, and enjoyable from the user's perspective". Moreover, Dix et al [11] classify usability into three principles to apply in order to design an interactive system:

- **Learnability:** refers to the ease users can master system's potentialities and interact in maximum effectiveness.
- **Flexibility:** refers to numerous ways the user and the system exchange data and information.
- **Robustness:** refers to "the level of support provided to the user in determining successful achievement and assessment of goals."

2.3 Programming Psychology

Teaching programming involves more than just teaching the syntax and semantics of a programming language. In 1971, Gerald Weinberg [12] published the Psychology of Computer Programming, with the stated goal to trigger a new field that studies computer programming as a human activity. In the field of PoP, programming is often defined as a process of transforming a mental plan that is in familiar terms into one that is compatible with the computer [13][14].
Soloway and Ehrich [15] introduced the notion of a schema or "programming plan". Rist [16] extends and tightens the schema concept and relates it to a full theory of program development and comprehension in novices and experts. While novices may acquire an understanding of the syntax of a programming language, without appropriate mental models they are unable to successfully understand complex concepts or engage in more difficult aspects of problem solving [17]. In a survey of program understanding, von Mayrhauser and Vans [18] summarize studies noting that experts: have efficiently organized and specialized knowledge schemas; organize their knowledge according to functional characteristics such as the nature of the underlying algorithm; use both general problem solving strategies and specialized strategies; use specialized schemas and a top-down, breath-first approach to efficiently decompose and understand programs; and are flexible in their approach to program comprehension and their willingness to abandon questionable hypotheses.

On the other hand, studies collected in Soloway and Spohrer [19] outline deficits in novices' understanding of various specific programming language constructs, note shortcomings in their planning and testing of code, explore more general issues relating to the use of program plans, show how prior knowledge can be a source of errors, and more.
Although, a variety of factors affect learning to program, we are still far from identifying the full reasons why some students learn to program easily while others not. Among others, learning programming may be affected by self-efficacy and mental models [20], working memory space and field dependency [21].

3 TIP-4U: Tutoring with Intelligence on Programming - 4 components Usability

Except of user-friendly interface, an ITS has to meet many educational and pedagogical qualities. Such qualities are the individualization of tutoring, reasoning abilities, diagnostic and didactic support. In the field of programming, where it has been noticed that students meet many difficulties and have a low performance, we have to focus on the particular characteristics of students who attend a course of programming, on their psychological features and on pedagogical strategies. Taking these into account, we suggest the following model (figure 1) for designing a useful ITS for teaching a programming language, utilizing general HCI principles as well as the derived knowledge from the field of the PoP.

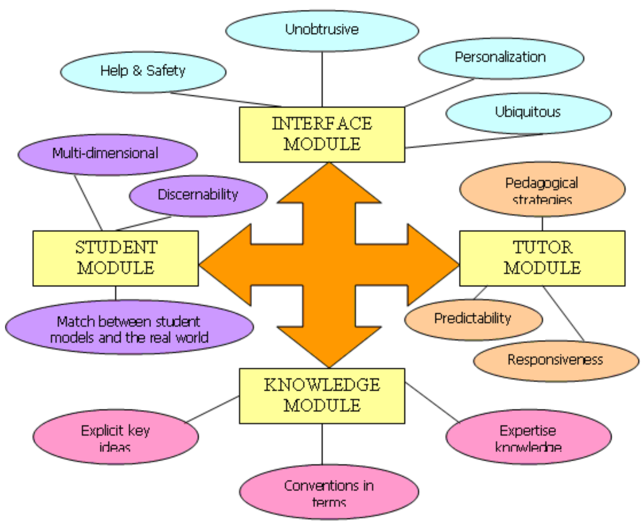


Fig. 1: TIP-4U Model.

3.1 Interface module

The interface module is responsive for the communication between a student and other modules of ITS. Research from the human factors and software design disciplines is applicable to designing friendly and usable interfaces. So the interface module of an ITS, has to comply with the following usability factors:

- **Personalization:** the user interface of an ITS, has to adapt to each individual student dynamically according to her/ his needs. That is to use different icons, font types, colors, structure following the needs of users.
- **Unobtrusive:** The interface has to be clear and natural. Colors and font types and sizes of texts have to be selected carefully in order the interface to be clear and easily readable. Also, the use of icons and graphics should become carefully without load the interface with 'heavy' and extra, usually unnecessary, information.
- **Ubiquitous:** Users must carry out easy and quick search to valuable info. Both novice and expert users should properly interact with the system according with their skills. They must be able to explore interfaces fast with the minimum mental fatigue.
- **Help & safety:** the system must prevent users from making errors and provide them with several means of recovery. Furthermore, effective help has to be provided.

3.2 Knowledge module

One of the most important components of an adaptive educational application is the representation of knowledge. To enable communication between system and learner at content level, the domain model of the system has to be adequate with respect to inferences and relations of domain entities with the mental domain of a human expert [22]. The design of the knowledge module has to meet the following factors.

- **Expertise knowledge:** the representation of the subject matter has to be a dynamic model of the domain knowledge and a set of rules by which the system can "reason." Thus, the domain knowledge of programming can be organized into a hierarchical tree, concepts maps, and semantic networks. A significant component of expertise knowledge is learning objectives, which determine the concepts that must be understood in each chapter-level [23]. Concerning programming language, the domain knowledge has to be organized taking into account the different components (syntax and semantics) of a programming language and the different kind and levels of knowledge and skills on programming. Linn and Dalbey [24] propose a "chain of cognitive accomplishments" that should arise from ideal computer programming instruction. This chain starts with the features of the language being taught, continues with design skills and ends with problem-solving skills. Davies [25] distinguishes between programming knowledge (of a declarative nature) and programming strategies (the way knowledge is used and applied). Therefore, learning objectives have to be categorized on three groups: a) declarative knowledge on programming language's features, b) designing skills and c) problem solving skills.

- **Conventions in terms:** the vocabulary, expressions and terms of domain knowledge have to be clear, concrete and follow the users' and programming culture.
- **Explicit key ideas:** Mancy & Reid [21] discovered that field dependency appears to be a critical skill in learning to program. Therefore, key ideas and important concepts have to be highlighted and content's structure has to be "clear". Moore & Dwyer [26] have suggested that use of color may help field-dependent students to see structure.

3.3 Student module

The student modeling component is responsible for dynamically representing the emerging knowledge and skills of the student [27] and for inferring the learner's knowledge and misconceptions for her/ his behavior [28]. Thus, a useful and effective student model should comply with the following factors.

- **Discernability:** The definition of the characteristics of a student model has to be discerned. In other words, it has to be clear limits between the different level of knowledge, skills and abilities, psychological and pedagogical states of a student. For the definition of the aspect of knowledge level and skills on programming of a student model we can advised the conceptual framework for analyzing students' knowledge of programming that was developed by McGill and Volet [29] and the evaluation method of knowledge of programming that was developed by deRaadt [30]. McGill and Violet discern three knowledge types in the view of cognitive psychology: declarative, procedural, strategic and three knowledge types in the view of educational research: syntactic (basic knowledge), conceptual (be able to combine knowledge, analytical thought) and strategic (integrated knowledge). De Raadt suggests five knowledge levels: a) No answer: learner has no knowledge of programming and cannot give an answer, b) Pre-structural: substantial lack of knowledge of programming structures, c) One-structural: learner is able to describe a part of code, d) Multi-structural: learner is able to describe a program line to line, e) Relational: learner is able toy describe the whole of a program.
- **Multi-dimensional:** A student model of an educational application for teaching of programming has to have many dimensions. Learning a programming language deals with many aspects. Such aspects are language's features, designing and problem-solving skills. Furthermore, the process of learning a programming language is affected from the student's background on programming skills or other programming languages and from their ability to solve problems. Also, it may be affected from the gender and the age of the learner.

- **Match between student models and the real world:** In the field of programming, designing of a student model has to focus on different characteristics and needs of novices and experts that have been discovered by many researches which took place in real classrooms [31]. McIver [32] states that trivial syntax errors may in fact impede learning, as they distract students from the fundamentals of programming and problem solving. Garner et al. [33] discovered that minor syntactical errors persisted as the most common cause of problems for novice programmers in both the stronger and weaker students throughout the duration of an introductory programming course. Furthermore, the same study indicates that program design come second only to the problems of syntax and are more prominent than problems relating to the comprehension of individuals language constructs. Sohporer & Soloway [34] state that students of a programming course have difficulty in putting all the pieces together.

3.4 Tutor module

According to Wenger [35], when learning is viewed as a successive transition between knowledge states, the purpose of teaching is accordingly to facilitate the student's traversal of the space of knowledge states. He claims that this requires that ITS alternate between diagnostic and didactic support. Thus, the following factors should be concerned while designing the tutor module.

- **Responsiveness:** the tutor model has to be informed about the progress and misconceptions of a student and to provide individualized instruction, support and advice. A recent study has shown that students' self-efficacy has a direct effect on overall success in an introductory course of programming [20]. Some ways to increase self-efficacy are persuasion, learning by peer modeling of tasks, social influences, and performance successes. For students to monitor their capableness, timely and sufficient feedback is necessary. Consequently, feedback and generally responsiveness to the actions, behavior and needs of students is a crucial factor.
- **Pedagogical strategies:** A tutor model has to follow pedagogical strategies in order to provide effective diagnostic and didactic support. As Winslow [36] supports, good pedagogy requires the instructor to keep initial facts, models and rules simple, and only expand and refine them as the student gains experience.
- **Predictability:** refers to the prediction of future outcomes on student's progress based on previous interactions and tutor experience. The tutor module has to be able to identify the misconceptions and needs of students and provide the appropriate support in order to guide them to reach the learning goal.

4 Conclusion

Our target in this paper was to define factors that contribute to designing a useful ITS for teaching a programming language. We combine the knowledge of the field of the PoP and the results of empirical studies of programming courses with the principles of HCI systems, in order to suggest an effective way to design an educational system for teaching programming, with the ability to present the teaching material in a flexible way and to provide learners with tailored instruction and feedback. In the future, an evaluation of these factors should be done.

References

1. Anderson, J.R., Boyle, C.F., Corbett, A., Lewis, M.: Cognitive modeling and Intelligence tutoring. Artificial Intelligence 42, 7–49 (1990)
2. Koedinger, K.R., Anderson, J.R.: Intelligent Tutoring Goes to School in the Big City. International Journal of Artificial Intelligence in Education 8, 30–43 (1997)
3. Byckling, P., Sajaniemi, J.: A role-based analysis model for the evaluation of novices' programming knowledge development. In: Proceedings of the 2006 international workshop on Computing education research, Canterbury, United Kingdom, pp. 85–96 (2006)
4. Grunbacher, P., Seyff, N., Briggs, R., Hoh Peter, I., Kitapsi, H., Port, D.: Making every student a winner: The WinWin approach in software engineering education. Journal of Systems and Software 80(8), 1191–1200 (2007)
5. Woszczynski, A., Haddad, H.M., Zgambo, A.F.: Towards a model of student success in programming cources. In: Proceedings of the 43rd annual Southeast regional conference, Kennesaw, Georgia, pp. 301–302 (2005)
6. Feedman, R.: What is an Intelligent Tutoring System? Intelligence 11(3), 15–16 (2000)
7. Psotka, J., Mutter, S.A.: Intelligent Tutoring Systems: Lessons Learned. Lawrence Erlbaum Associates, Mahwah (1988)
8. Nielsen, J.: Usability Engineering. Academic Press, London (1993)
9. Shneiderman, B.: Designing the User Interface: Strategies for Human-Computer Interaction, 3rd edn. Addison-Wesley, Reading (1998)
10. Preece, J., Rogers, Y., Sharp, H.: Interaction Design: Beyond Human Computer Interaction. John Wiley & Sons, Inc., Chichester (2006)
11. Dix, A., Abowd, G., Beale, R., Finlay, J.: Human-Computer Interaction, 2nd edn. Prentice Hall Europe, United Kingdom (1998)
12. Weinberg, G.M.: The Psychology of Computer Programming. Van Nostrand Reinhold Company, New York (1971)
13. Lewis, C., Olson, G.M.: Can Principles of Cognition Lower the Barriers to Programming? In: Olson, G.M., Sheppard, S., Soloway, E. (eds.) Empirical Studies of Programmers: Second Workshop, pp. 248–263. Ablex, Norwood (1987)
14. Green, T.R.G., Petre, M.: Usability Analysis of Visual Programming Environments: A 'Cognitive Dimensions' Framework. Journal of Visual Languages and Computing 7(2), 131–174 (1996)
15. Soloway, E., Erlich, K.: Empirical studies of programming knowledge. IEEE Transactions on Software Engineering SE-10(5), 595–609 (1984)
16. Rist, R.S.: Program structure and design. Cognitive Science 13, 507–562 (1995)

17. Carlisle, G.: Experiences with Novices: The Importance of Graphical Representations in Supporting Mental Models. In: 12^{th} Workshop of the Psychology of Programming Interest Group, Cozenza Italy, pp. 33–44 (2000)
18. Von Mayrhauser, A., Vans, A.M.: Program understanding - A survey (Tech. Rep. CS-94-120). Department of Computer Science, Colorado State University (1994)
19. Soloway, E., Spohrer, J.C. (eds.): Studying the novice programmer. Lawrence Erlbaum, Hillsdale (1989)
20. Weidenbeck, S., LaBelle, D., Kain, V.: Factors Affecting Course Outcomes in Introductory Programming. In: 16^{th} Workshop of the Psychology of Programming Interest Group, Carlow, Ireland, pp. 97–110 (2004)
21. Mancy, Reid: Aspects of Cognitive Style and Programming. In: 16^{th} Workshop of the Psychology of Programming Interest Group, Carlow, Ireland, pp. 1–9 (2004)
22. Peylo, C., Teiken, W., Rollinger, C., Gust, H.: An ontology as domain model in a web-based educational system for prolog. In: Etheredge, J., Manaris, B. (eds.) Proceedings of the 13^{th} International Florida Artificial Intelligence Research Society Conference, pp. 55–59. AAAI Press, Menlo Park (2000)
23. Driscoll, M.: Web- based Training. In: Creating e-Learning Experiences, 2nd edn. John Wiley & Sons, Inc., San Francisco (2002)
24. Linn, M.C., Dalbey, J.: Cognitive consequences of programming instruction. In: Soloway, E., Spohrer, J.C. (eds.) Studying the novice programmer, pp. 57–81. Lawrence Erlbaum, Hillsdale (1989)
25. Davies, S.P.: Models and theories of programming strategy. International Journal of Man-Machine Studies 39, 237–267 (1993)
26. Moore, D.M., Dwyer, F.M.: Effect of color coded information on students' levels of field dependence. Perceptual and Motor Skills 72, 611–616 (1991)
27. Chin, D.: Empirical evaluation of user models and user adapted systems. User Modeling and User-Adapted Interaction 11, 181–194 (2001)
28. Nwana, H.: User Modelling and User Adapted Interaction in an Intelligent Tutoring System. User Modeling and User-Adapted Interaction 1(1), 1–32 (1991)
29. McGill, T.J., Volet, S.E.: A Conceptual Framework for Analyzing Students' Knowledge of Programming. Journal of Research on Computing in Education 29, 276–297 (1997)
30. de Raadt, M.: A Review of Australasian Investigations into Problem Solving and the Novice Programmer. Computer Science Education 17(3), 201–213 (2007)
31. Robins, A., Rountree, J., Rountree, N.: Learning and Teaching Programming: A Review and Discussion. Computer Science Education 13(2), 137–172 (2003)
32. McIver, L.: The Effect of Programming Language on Error Rates of Novices. In: 12^{th} Workshop of the Psychology of Programming Interest Group, Psychology of Programming Interest Group, Cozenza Italy, pp. 181–192 (2000)
33. Garner, S., Haden, P., Robins, R.: My Program is Correct But it Doesn't Run. In: Proceedings of the $7th$ Australasian Computing Education Conference, pp. 173–180. Australian Computer Society, Newcastle NSW (2005)
34. Soloway, E., Spohrer, J.: Studying the Novice Programmer. Lawrence Erlbaum Associates, Hillsdale (1989)
35. Wenger, E.: Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge. Morgan Kaufmann Publishers, Inc., Los Altos (1987)
36. Winslow, L.E.: Programming pedagogy - A psychological overview. SIGCSE Bulletin 28, 17–22 (1996)

Emotional Intelligence in Multimodal Object Oriented User Interfaces

Efthymios Alepis and Maria Virvou

Abstract In this paper, we describe a multimodal emotion recognition system that is structured according to the object oriented method. The system uses the OO approach that combines evidence from multiple modalities of interaction and data from emotion stereotypes and classifies them into well structured objects with their own properties and methods. The system's architecture can be adopted in future emotion recognition systems with multiple modalities and improved emotion detection algorithms, while the resulting emotion detection server is capable of using and handling transmitted information from different sources of data during human-computer interaction.

1 Introduction

One of the major scientific challenges is the exploration of how humans interact with their environment and with each other. Perceiving, learning and adapting to the world around us are commonly labeled as intelligent behavior [1]. In many situations human-computer interaction may be improved by multimodal emotional interaction in real time [2], [3]. Affective computing has recently become a very important field of research because it focuses on recognizing and reproducing human feelings within human computer interaction. Human feelings are considered very important but only recently have started being taken into account in user interfaces. Thus, the area of affective computing is not yet well understood and needs a lot more research to reach maturity.

As Picard claims in [4], one of the major challenges in affective computing is to try to improve the accuracy of recognizing people's emotions. During the last decade, the visual and the audio channel of human-computer interaction were considered

Department of Informatics, University of Piraeus,
80 Karaoli & Dimitriou St., 18534 Piraeus, Greece
e-mail: talepis,mvirvou@unipi.gr

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 349–359.
springerlink.com

as most important in human recognition of affective states [5]. Yet, research in psychophysiology has produced firm evidence that affective arousal has a range of somatic and physiological correlates, such as heart rate, skin clamminess, body temperature, etc. [6]. Correspondingly, a reasonable suggestion for the improvement of accuracy in affect recognition is the combination of more than one modes of interaction in user interfaces. It is hoped that the multimodal approach may provide not only better performance, but also more robustness [1].

Similar views about the benefits of the combination of modalities have been supported by many researchers in the field of human-computer interaction [7, 8, 9, 10, 11]. However, progress in emotion recognition based on multiple modalities has been rather slow. Although several approaches have been proposed to recognize human emotions based on facial expressions or speech unimodally, relatively limited work has been done to fuse these two and other modalities to improve the accuracy and robustness of the emotion recognition systems [12].

The issue of combining multiple modalities raises the problem of how these modalities may be combined. Correspondingly, this problem consists of the determination of a general architecture of a multi-modal emotion recognition system, as well as of the sophisticated mechanisms that will fuse this system's available data in order to utilize the emotion recognition functions. In fact, the mathematical tools and theories that have been used for affect recognition can lead to a classification of affect recognizers. Such a classification has been made in [13] where affect recognizers have been classified into two groups on the basis of the mathematical tools that these recognizers have used: 1. The first group using traditional classification methods in pattern recognition, including rule-based systems, discriminate analysis, fuzzy rules, case-based and instance-based learning, linear and nonlinear regression, neural networks, Bayesian learning and other learning techniques. 2. The second group of approaches using Hidden Markov Models, Bayesian networks etc. Indeed, a recent piece of research that uses the above approaches for the integration of audio-visual evidence is reported in [14]. Specifically, for person-dependent recognition, Zeng and his colleagues [11] apply the voting method to combine the frame-based classification results from both audio and visual channels. For person-independent tests, they apply multi-stream hidden Markov models (HMM) to combine the information from multiple component streams.

The fore mentioned approaches indicate methodologies and algorithms that try to combine multi-modal information efficiently, and all of them present advantages in the field of affective computing. However, after a thorough investigation in the related scientific literature we found that there is a shortage of studies that aim at the theoretical data structures of multimodal information systems, aside from their algorithmic models. Furthermore, the well-known Object Oriented Method (OOM) has not been found to be incorporated in current emotion recognition systems as architecture for modeling them efficiently and reliably.

In view of the above, in this paper we present an object-oriented model for emotion recognition purposes. The object-oriented model is used to represent significantly the available information from the modalities of interaction, as well as information about the users and their actions. The proposed model incorporates all common object oriented concepts, which are described explicitly in the next section, trying to provide exploitable information, more robustness for the algorithms that are going to use the available data and easiness in the addition of potential new modalities for emotion recognition.

In section 2 we make a sort overview of the OO-Method and the tools that have been used for the generation and the analysis of object oriented models. In section 3 we present the general architecture of the object oriented emotion recognition system, while in section 4 we describe how emotion recognition data may be incorporated into the system. Finally in section 5 we come up with conclusions and future work.

2 The object oriented approach

Object-oriented programming can trace its roots to the 1960s, but was not commonly used in mainstream software application development until the early 1990s [15]. Object-oriented programming provided researchers with ways to maintain software quality and to develop object oriented applications in part to address common problems by emphasizing discrete, reusable units of programming logic. An object-oriented program may be considered as a collection of cooperating objects, as opposed to the conventional model, in which a program is seen as a list of tasks (subroutines) to perform. In OOP, each object is capable of receiving messages, processing data, and sending messages to other objects and can be viewed as an independent mechanism with distinct roles or responsibilities.

Object oriented approaches have been already widely used in software development environments [16], [17]. An improved object oriented model for Java is presented in [18]. In [19], an object-oriented analysis is adopted for the implementation of remote sensing imagery to GIS. The authors of this paper argue that there is a large gap between theoretically available information and used information to support decision making. As a proposed strategy to bridge this gap, these authors suggest the extension of their signal processing approach for image analysis by exploration of a hierarchical image object network to represent the strongly linked real-world objects.

Additionally to the OO method, the UML [20] approach has been developed to standardize the set of notations used by most well known object oriented methods. In order to support these approaches, CASE tools like Rational Rose [21] and Paradigm Plus [22] have been developed.

According to [23], a number of fundamental concepts are found in the strong majority of definitions of object oriented programming:

- Class Classes define the abstract characteristics of objects, including their characteristics (attributes, fields and properties) and their functions.
- Objects Objects are patterns of a class.
- Instances An instance is the actual object created at runtime.
- Methods Methods illustrate objects' abilities. In programming languages, methods are referred to as functions.

- Message passing Message passing represent the general process by which an object sends data to another object or asks the other object to invoke a method.
- Inheritance Subclasses are more specialized versions of a class, which inherit attributes and behaviours from their parent classes, and can introduce their own.
- Abstraction Abstraction is simplifying complex reality by modelling classes appropriate to the problem, and working at the most appropriate level of inheritance for a given aspect of the problem.
- Encapsulation Encapsulation conceals the functional details of a class from objects that send messages to it.
- Polymorphism Polymorphism allows the programmer to treat derived class members just like their parent class' members. Polymorphism in object-oriented programming is the ability of objects belonging to different data types to respond to method calls of methods of the same name, each one according to an appropriate type-specific behaviour.
- Decoupling Decoupling allows for the separation of object interactions from classes and inheritance into distinct layers of abstraction.

3 General architecture of the object oriented emotion recognition system

In this section, we describe the general architecture of the emotion recognition system. Data that can be considered as relational to the general emotion recognition process are distinguished by two individual categories. The first category comprises of the multimodal data that may be collected by each individual modality. The emotion detection server consists of one or more modalities and each of them are factually the emotion detection server's properties. Correspondingly, emotional states constitute attributes and methods of each modality. In each implementation of the system, there is a set of available modalities. The authors have presented in their previous works [24, 25], emotion recognition systems that consist of two or three modalities. As it is illustrated in figure 1, the emotion detection server may read and exchange information with the available modalities. Each modality may provide the system with information about the recognition of one or more emotional states, with specific degrees of certainty. Additionally, each modality may provide supplementary information concerning the emotional states of users, which is associated with specific user actions during the human-computer interaction. Such actions include correct or wrong browsing, answers in tests in educational software environments, etc.

Stereotypic information is also very important for emotion recognition purposes and a complete study considering their incorporation into the emotion recognition system is shown in [26]. In this paper we suggest an object oriented structure of all the available stereotypic information. As it is illustrated in figure 2, the main emotion stereotype class consists of three subclasses. The first class stores and administers data that are associated with users' characteristics. These characteristics derive from each user's personality, such as the user's educational level, the user's sex and age, etc. and help the system improve its emotion recognition capabilities. The second subclass models stereotypic information concerning user actions during their interaction, while the third subclass represents pre-stored information about each modality's ability of recognizing each one of the available emotional states.

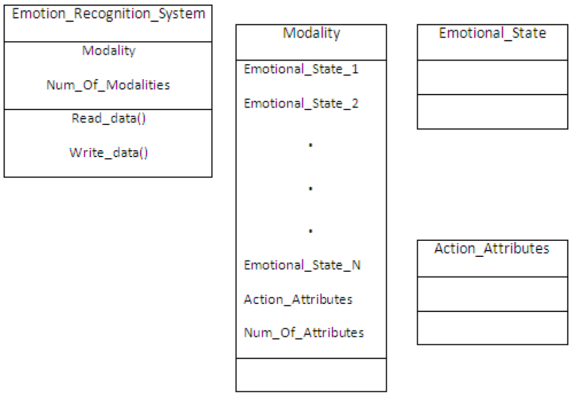


Fig. 1: Object model for the Emotion Recognition System.

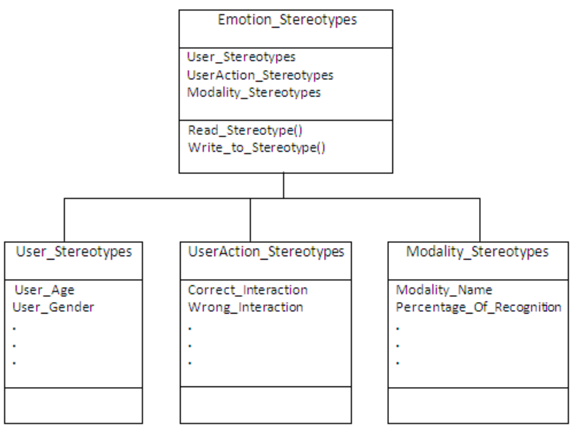


Fig. 2: Object model for the construction of the emotional stereotypes.

One of the major contributions of the object oriented architecture of the emotional recognition system is the great easiness in adapting the system to new or more roles that may lead to better emotion recognition capabilities (such as the incorporation of new modalities). Moreover, this architecture provides a framework for emotion recognition systems in different computerized environments and laboratory installations. As it is illustrated in figure 3, the resulting emotion detection server can exchange emotional interaction data from multiple sources, such as personal computers, mobile devices, or even integrated laboratory installations with multiple devices as input modalities.

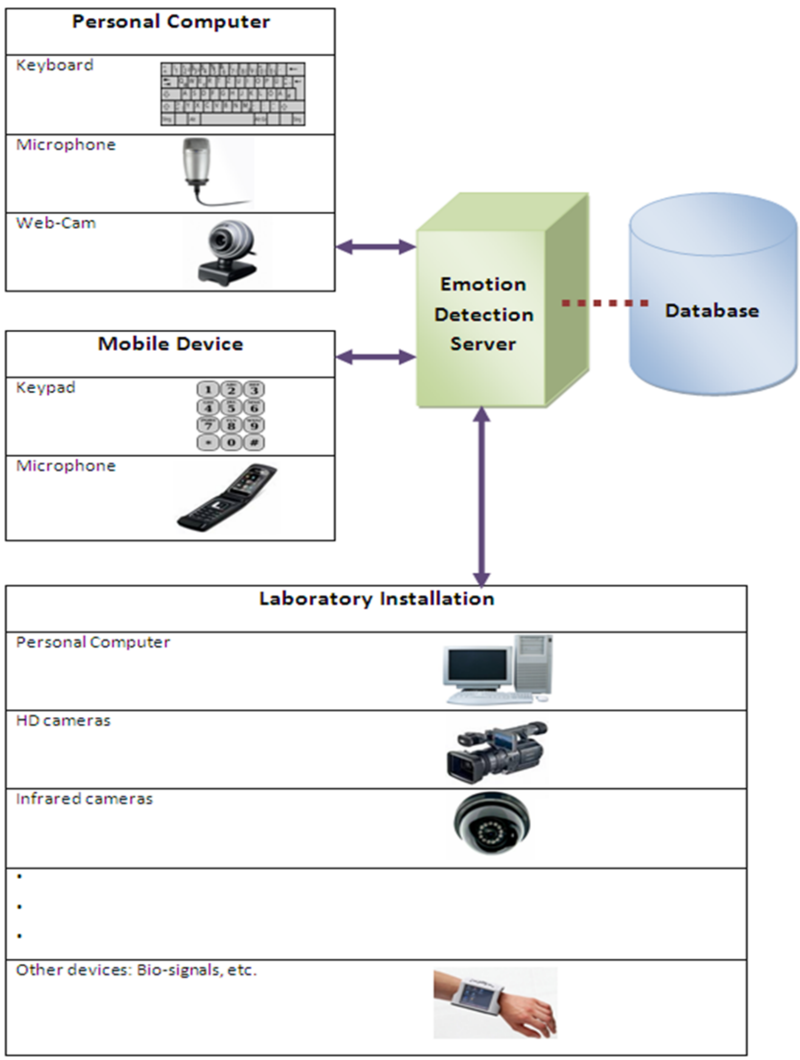


Fig. 3: General Architecture of the Emotion Detection Server.

4 Emotion recognition data into objects

In section 3 we have shown how the object oriented method may be used in order to provide a reliable model structuring the information that is used in emotion recognition systems. Furthermore, this approach can be adopted in future emotion recognition systems with more modalities and improved emotion detection mechanisms. In this section we present actual emotion recognition data that are classified according to the fore mention object oriented architecture. The aim of this section is to illustrate the variety of different data, in order to indicate the necessity of a well structured approach which can classify and manipulate them.

4.1 Data from multiple modalities

As is it shown in [27], different computer modalities have distinguishable capabilities in recognizing humans' emotional states. For example, some emotions may give significant visual evidence of their existence to any observer, while in other situations audio-lingual data may be preferable. Furthermore, in human-computer interaction, specific categories of users may use different modalities for their interaction. In an empirical study that we conducted in previous research work we have noticed that a very high percentage (85%) of young people (below 30 years old) who are also inex-perienced with computers reported to have preferred expressing themselves through the oral mode rather than the keyboard during their interaction with a computer. On the contrary participants who were computer experts did not give us considerable data for the affect perception during the oral communication with their computer. Another important issue concerning the combination of data from multiple modalities is the fact that different modalities may give evidence for different emotions or emotional states. A modality may be able to provide a system with information about six discrete emotions, while another modality may only decide whether positive or negative feelings are detected.

4.2 Data from user input actions

The emotion recognition system incorporated a user monitoring component that captured all users' actions concerning two modalities of interaction, namely the personal computer's keyboard and the computer's microphone. After processing the collected data we also came up with statistical results that associated user input actions through the keyboard and microphone with emotional states. More specifically, considering the keyboard we have the following categories of user actions: k1) user types normally k2) user types quickly (speed higher than the usual speed of the particular user) k3) user types slowly (speed lower than the usual speed of the particular user) k4) user uses the "delete" key of his/her personal computer often k5) user presses unrelated keys on the keyboard k6) user does not use the keyboard. These actions were also considered as criteria for the evaluation of emotion with respect to the user's action in the keyboard. Considering the users' basic input actions through the microphone we come up with 7 cases: m1) user speaks using strong language m2) users uses exclamations m3) user speaks with a high voice volume (higher than the average recorded level) m4) user speaks with a low voice volume (low than the average recorded level) m5) user speaks in a normal voice volume m6) user speaks words from a specific list of words showing an emotion m7) user does not say anything. These seven actions were also considered as criteria for the evaluation of emotion with respect to what the user says.

4.3 Stereotypic information

Considering the important problem of which mode gives better results or in what extent should the evidence from each mode should be taken into account, the authors have proposed a novel approach for calculating the weight of significance of each mode based on stereotypes and a multi-criteria theory [26,27]. Stereotype-based reasoning takes an initial impression of the user and uses this to build a user model based on default assumptions [28]. Stereotypes constitute a powerful mechanism for building user models [28]. This is due to the fact that stereotypes represent information that enables the system to make a large number of plausible inferences on the basis of a substantially smaller number of observations [29]. The stereotype inferences are used in combination with a decision theory, namely Simple Additive Weighting (SAW) ([30], [31]) for estimating weight of significance of each mode in the affective reasoning of the system for a particular user.

In previous research work of the authors [26], we have classified our users into stereotypes concerning their age, their educational level, their computer knowledge level and their gender. For each user there is a value that corresponds to a four-dimensional stereotypic vector of the form:

(User_Name, Stereotypic Characteristic1, Stereotypic Characteristic2, Stereotypic Characteristic3, Stereotypic Characteristic4).

Stereotypic Characteristic1 refers to the user's age and is an element of the following set concerning ages: [(10-16), (16-24), (24-36), (36-50), (over 50)]. *Stereotypic Characteristic 2* refers to the user's computer knowledge level and is an element of the following set concerning the user's computer experience in months (using a personal computer): [(less than 1), (1-6), (6-12), (over 12)]. Similarly we have defined *stereotypic characteristics 3* and *4* that refer to the user's educational level and to the user's gender respectively. The inferences of this stereotype vector provide information about the weights of importance of each mode for the users belonging to that stereotype. For example if a user belongs to the stereotype [(16-24)] is inferred to have a tendency to express his/her feelings through the oral mode of interaction. Stereotypes can provide inferences concerning hypothesis about users' feelings and which modality should be more important for providing evidence about users' feelings. More specifically, in many cases, data from the vocal interaction or the interaction through the keyboard gives evidence of different emotions with quite similar degrees of certainty. For example, the system may have evidence that a user is either angry while saying or typing something or stressed or even confused. The incorporation of stereotypes in the system provides inferences concerning people belonging to the same category with the user that may help in recognizing an emotion that is more common for the users of this category among others or even distinguishing emotions. Evidence for the character or the personality of a particular user may raise the degree of certainty for a particular emotion recognized.

4.4 Conclusions

In this paper, we described a multimodal emotion recognition system that is structured according to the object oriented method. The system uses the OO approach that combines evidence from multiple modalities of interaction and data from emotion stereotypes and classifies them into well structured objects with their own properties and methods. Advantages of the proposed approach include the well-known conveniences and capabilities of object oriented structures, such as easiness in the system's maintenance, great extensibility, better communication through different modalities, good cooperation with different object oriented programming languages, easiness in code debugging, as well as code reusability.
The present system's architecture can be adopted in future emotion recognition systems with multiple modalities and improved emotion detection algorithms. Furthermore the resulting emotion detection server is capable of using and handling transmitted information from different sources of human-computer interaction. Independent user interfaces may send wirelessly or wired information about users' interaction to the emotion detection server and the server can respond with information about possibly recognized emotional states of the users. As for future work we plan to improve our system and test the system's structure by exploiting additional modalities of interaction.

References

1. Pantic, M., Rothkrantz, L.J.M.: Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE 91, 1370–1390 (2003)
2. Jascanu, N., Jascanu, V., Bumbaru, S.: Toward emotional E-commerce: The customer agent. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part I. LNCS (LNAI), vol. 5177, pp. 202–209. Springer, Heidelberg (2008)
3. Bernhardt, D., Robinson, P.: Interactive control of music using emotional body expressions. In: Conference on Human Factors in Computing Systems - Proceedings, pp. 3117–3122 (2008)
4. Picard, R.W.: Affective Computing: Challenges. Int. Journal of Human-Computer Studies 59(1-2), 55–64 (2003)

5. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.: Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine 18(1), 32–80 (2001)
6. Cacioppo, J.T., Berntson, G.G., Larsen, J.T., Poehlmann, K.M., Ito, T.A.: The Psychophysiology of Emotion. In: Lewis, M., Haviland-Jones, J.M. (eds.) Handbook of Emotions, pp. 173–191 (2000)
7. Chen, L.S., Huang, T.S., Miyasato, T., Nakatsu, R.: Multimodal Human Emotion/Expression Recognition. In: Proceedings of the 3rd International Conference on Face & Gesture Recognition. IEEE Computer Society, Los Alamitos (1998)
8. De Silva, L., Miyasato, T., Nakatsu, R.: Facial emotion recognition using multimodal information. In: IEEE Int. Conf. on Information, Communications and Signal Processing (ICICS 1997), pp. 397–401 (1997)
9. Huang, T.S., Chen, L.S., Tao, H.: Bimodal emotion recognition by man and machine. In: ATR Workshop on Virtual Communication Environments, Kyoto, Japan (1998)
10. Oviatt, S.: User-modeling and evaluation of multimodal interfaces. Proceedings of the IEEE, 1457–1468 (2003)
11. Zeng, Z., Tu, J., Liu, M., Huang, T., Pianfetti, B., Roth, D., Levinson, S.: Audio-Visual Affect Recognition. IEEE Transactions on Multimedia 9, 424–428 (2007)
12. Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C., Kazemzadeh, A., Lee, S., Neumann, U., Narayanan, S.: Analysis of emotion recognition using facial expressions, speech and multimodal information. In: Proceedings of the 6th international conference on Multimodal interfaces State College. ACM, PA (2004)
13. Liao, W., Zhang, W., Zhu, Z., Ji, Q., Gray, W.D.: Toward a decision-theoretic framework for affect recognition and user assistance. Int. J. Hum.-Comput. Stud. 64, 847–873 (2006)
14. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications, vol. 186. Springer, Heidelberg (1981)
15. Wikipedia. The free encyclopedia, `http://en.wikipedia.org/wiki/Object_oriented`
16. Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., Lorensen, W.: Object Oriented Modeling and Design. Prentice Hall, Englewood Cliffs (1991)
17. Jacobson, I., Christerson, M., Jonsson, P., Overgaard, G.: Object-oriented software engineering. A use case driven approach. ACM Press/Addison-Wesley (1992)
18. Pastor, O., Gómez, J., Insfrán, E., Pelechano, V.: The OO-Method approach for information systems modeling: From object-oriented conceptual modeling to automated programming. Information Systems 26(7), 507–534 (2001)
19. Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M.: Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS Journal of Photogrammetry and Remote Sensing 58(3-4), 239–258 (2004)
20. Booch, G.: The unified modeling language. Performance Computing/Unix Review 14(13), 41–48 (1996)
21. Rational Software Corporation, Rational Rose User's Manual (1997)
22. Platinum Technology, Inc., Paradigm Plus: Round-Trip Engineering for JAVA, White Paper (1997)
23. Armstrong, D.J.: The Quarks of Object-Oriented Development. Communications of the ACM 49(2), 123–128 (2006)
24. Alepis, E., Virvou, M., Kabassi, K.: Requirements Analysis and Design of an Affective Bi-Modal Intelligent Tutoring System: The Case of Keyboard and Microphone. Studies in Computational Intelligence, Intelligent Interactive Systems in Knowledge-Based Environments, vol. 104, pp. 9–24 (2008)

25. Tsihrintzis, G., Virvou, M., Stathopoulou, I.O., Alepis, E.: On improving visual-facial emotion recognition with audio-lingual and keyboard stroke pattern information. In: Web Intelligence and Intelligent Agent Technology, WI-IAT 2008, vol. 1, pp. 810–816 (2008)
26. Alepis, E., Virvou, M.: Emotional Intelligence: Constructing user stereotypes for affective bi- modal interaction. In: Gabrys, B., Howlett, R.J., Jain, L.C. (eds.) KES 2006, Part I. LNCS (LNAI), vol. 4251, pp. 435–442. Springer, Heidelberg (2006)
27. Virvou, M., Tsihrintzis, G., Alepis, E., Stathopoulou, I.O., Kabassi, K.: Combining Empirical Studies of Audio-Lingual and Visual-Facial Modalities for Emotion Recognition. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part II. LNCS (LNAI), vol. 4693, pp. 1130–1137. Springer, Heidelberg (2007)
28. Moriyama, T., Ozawa, S.: Measurement of Human Vocal Emotion Using Fuzzy Control. Systems and Computers in Japan 32(4) (2001)
29. Kay, J.: Stereotypes, student models and scrutability. In: Gauthier, G., VanLehn, K., Frasson, C. (eds.) ITS 2000. LNCS, vol. 1839, pp. 19–30. Springer, Heidelberg (2000)
30. Rich, E.: Users are individuals: individualizing user models. International Journal of Man-Machine Studies 18, 199–214 (1983)
31. Fishburn, P.C.: Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments, Operations Research (1967)
32. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications. Lecture Notes in Economics and Mathematical Systems, vol. 186. Springer, Heidelberg (1981)

A Mobile Music Recommender System based on a Two-level Genre-Rating SVM Classifier enhanced by Collaborative Filtering

P.S. Lampropoulou, A.S. Lampropoulos, and G.A. Tsihrintzis

Abstract In this paper, we present a mobile recommender system for music. The system acts as middleware that facilitates the recommendation in push technology-based mobile services and combines content-based retrieval and collaborative filtering. Specifically, a mobile user is provided with the ability to query for music files that belong to the same genre by simply sending an example music file from his/her mobile device. The system recommends to the user music files not only on the basis of acoustic content similarity, but also incorporates collaborative filtering techniques in order to offer more personalized recommendations.

1 Introduction

In recent years there is a rapid growth of mobile technologies which is further boosted by the evolution in mobile device technologies and mobile network technologies. The modern mobile devices have inherent abilities to provide multimedia services as they are equipped with camera, radio, music players and other modules that create and handle multimedia data. These advances have led users to consider their mobile devices not only as simple communication devices, but rather as multimedia entertainment and storage equipment [1].

Moreover, a serious limitation of mobile devices, that is, their limited storage memory, was eliminated in the recent years. Furthermore, recent increasing amount of available digital storage space have led to the creation of large repositories of music files for use by broad classes of mobile users. It is obvious, that mobile users can not manage these resources of music files properly, as the huge amount of data exceeds the time, they spend time to find these. In turn, this fact gives rise to a need for systems that have the ability, to recommend users appropriate set of music files.

University of Piraeus, Department of Informatics,
80 Karaoli and Dimitriou St, Piraeus 18534, Greece.
e-mail: vlamp,arislamp,geoatsi@unipi.gr

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 361–368.
springerlink.com

The recommender systems represent the similar social process of recommendation and help into the pressure of information overload.

Recommender systems take into account users personal needs and interests and provide information in a way that will be most appropriate and valuable to them.Recommender systems are also particularly useful for mobile users as they have to decision in a short time and the effort required for interacting with the system must be limited as much as possible.

A mobile recommender system not only supply a user with search results to a specific user's query but it is also attempting to predict items that a user might be interested in. So, a recommender system attempt to result music pieces from a collection according to their similarity to the user's query by also embedding the individualization and personalization of the user.

Consequently, our work emphasizes in creation of a mobile music recommender system which will be able to be adapted according to user's change of preferences, to use a small or none amount of user's preferences, to offer quick results without sending unwanted content. In order, our system overcome the main drawbacks of the main recommendation technique, the Collaboration Filtering, 1)non-association, 2)user bias, and 3)cold start we used a combination of content-based retrieval techniques and item-based Collaboration Filtering.[2]

The paper is organized as follows: Section 2 reviews related works, while Section 3 presents an overview of our system as it has been developed. Section 4 describes the recommendation technique incorporated into the back-end of system. Finally, conclusions are drawn and future research directions are illustrated in Section 5.

2 Related Work

In this section, we review briefly the most relevant works on mobile recommender systems and on music recommender systems. Collaborative filtering and content-based recommending are two fundamental techniques that have been used for performing recommendation. Both techniques have their own advantages however they cannot perform well in many situations.

Collaborative Filtering technique recommends [3][4] recommend pieces to a user by considering someone else's ratings of those pieces. For example, suppose that there is a target user who likes pieces A and B. If there are many other users who like A, B, and C, C will probably be recommended to the target user.

A target user the preferred content of the group whose content consumption mind is similar to that of the user. Because of the mature of the technique, Collaborative Filtering has been attractive for predicting various preference problems such as net-news, e-commerce, digital libraries, etc. Several papers applied Collaborative Filtering to mobile devices. MobiTip is a system that Collaborative Filtering used for predicting rating changes of movie data based on users new or updated ratings in their system [5].

One of the first music recommendation systems that used Collaborative Filtering was presented at [6] that utilized filtered data from the Web to support a playlist-based or so-called ”collaborative filtering-based recommendation engine”. Their system generated recommendations based on the contents of the user’s playlists.

On the other hand Content-based recommendation systems are based on finding the objects of interest are by their associated features. These systems learns a profile of the user.s interests based on the features that present in items the user has rated and they make recommendations without having to match their interest to someone else.

To improve performance, various hybrid techniques have been considered. An hybrid approach of recommendation is the combination of content-based prediction and collaborative filtering. The content-based prediction part of this approach can solve the sparse matrix and the first-rater problem of pure collaborative filtering by predicted ratings of unrated items using prior knowledge about user preference from rated items. Since the performance of the hybrid approach usually depends on the accuracy of the content-based prediction, in this approach Support Vector Machines used in content-based prediction tasks[7].

3 System Overview

The recommender system is developed as a *middleware* system that offers to a mobile user the ability to submit a query by sending an example music file with a grade rating from his/her mobile device to a digital music library and request for recommended music files that belong to the same musical genre (“are similar to the query”). As our previous work ALIMOS [8, 9] the system was developed to alleviate the limitations of existing mobile recommender systems. Specifically, the system does not require an operating system in the mobile device, nor do any APIs or specific client need to be installed. Our system uses Push technology which is supported by all mobile devices. Additionally, the recommendation procedure followed is quite simple. Indeed, the user initiates the service by sending a Multimedia Mobile Service(MMS) to the server containing the audio file, that used to initiate the recommendation process.

The recommendation of the system is not *content delivery* that would be a result in unnecessarily high costs to the user. Instead, the system returns simple push messages. The first push message contains a link with a list of the recommended music files. The first list contains the names of the music files that are contented- based similar to the audio query file. The link redirects the user into a web-form where the user is able to rate each music file with a grade from one to five. By submitting the ratings, the system sends a second push message that contains a list of appropriate links of the recommended music files as they updating in order to take into account the user’s rating. The links are named so that the user can decide from the link name whether to proceed to selective downloading of several of these music files from the web server to his mobile device.

In our application, the corresponding system architecture is a multi-tier application with a *front-end* and a *back-end* level. The front-end level includes all those modules that implement the communication between the user, i.e., the mobile network and the application, while the back-end level refers to all those modules that implement the recommendation mechanism. Even though the music databases may be distributed and separate from the back-end and the Push Proxy Gateway (PPG), the Push Initiator (PI) and the Wireless Access Protocol (WAP) servers in the front-end, the design of our system is such that these various modules are not visible to the users and all communication needs among the modules are handled automatically by the system[10],[11].

The system operation can be summarized by the following steps.

1. The user submits a MMS containing the query audio file
2. The server receives the MMS, extracts the query audio file and stores the necessary user information (e.g., sender's telephone number).
3. The server submits the audio file to the content-based retrieval (CBR) sub-system of the back-end level.
4. The CBR sub-system of the back-end level returns the list of the five most relevant results.
5. The system push a message with the list and a link to a web-page where the user rates from one to five each music file from the list of the results and submits the results to the back-end.
6. The back-end uses the average of user ratings for returning a list of results with the same rating.
7. The system forwards the new list to the user by updating the push message. The results in the push message are not the actual music files, but rather links to the WAP server named after the music piece title and performing artist name.
8. The user may download from the WAP server the audio file(s) that he/she selects.

4 Back-end recommendation enhanced with Content based and Collaborative Filtering

As we described in the previous section, the back-end level consists of all these modules that implement the recommendation process. In this paper, the back end of our middleware system is constituted with a two level RBF-SVM classifier which enhances content-based information and ratings of music files to recommend into a personalized way music files to the user. The goal of the recommendation mechanism is to allow the service to return audio/music search results that are not only similar in content and of the same genre as the query audio/music file, but have also taken into account user preferences that represented as ratings in each music file.

Specifically, the user profile is created by training a two level of RBF-SVM classifier, which are later fine-tuned via the information about the files downloaded by

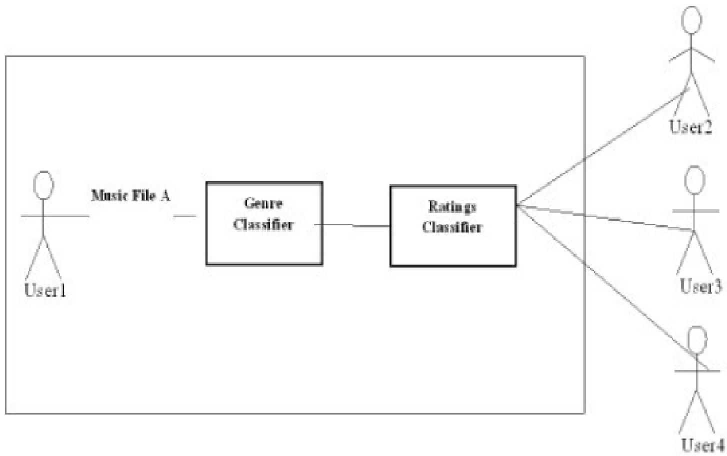


Fig. 1: Recommendation Technique

the user from those returned by our system in response to previous queries by the same user. In steps, the procedure is as follows:

1. The user seeds the search, by submitting a query (target music piece). The feature extractor extracts a set of values for a corresponding set of 30 audio features.
2. In the first level the classifier answers the Genre of query music file and the second level classifies this song into a rating class. More specifically, the first level of the classifier identifies the musical genre as *confidence* of the query file. The decision value (confidence) is the distance of the query from the margin of the specific class. The class is selected which corresponds to the classifier with the highest decision value.
3. The second level of the classifier return a possible rating for the user's query (target music piece). Each music file in our database is associated with one of 5 rating degree. In other words, each musical genre class is sub-divided by 5 rating classes. The rating of each music file is updated through the average ratings of users for the specific music file.
4. The classifier returns 5 songs with similar decision value from the margin of the specific class and with the same rating with the predicted rating.
5. Then the user evaluates the results giving a ratings to the results. The average of these ratings updates the rating for the query (target music file).
6. If the user selects to download a music/audio from the results of previous search, the system updates the query using the confidence value of the selected song and rating and searches for songs which are close to the new decision value and new rating degree of the music file. Also, the ratings of the retrieved files are updated taking into account the ratings of the new user.

In this process, the user preference is captured and stored using the class ID and the latest decision value from the margin of the specific class as well as the corresponding ratings of the queries and retrieved files. This allows the adjustment of the user preference according to a specific class, while attention is paid to a subspace of this class according to the decision value from the margin of the class and the ratings in the neighborhood of the decision value.

An example of the recommendation process of the system is presented below. In the initial step the user submits the query. The query corresponds to the file ”Led Zeppelin -The Rain Song”. This file is processed by the content-based retrieval process in the back-end level of the system. This file is classified to the ”Rock” class and in rating 2.

The system returns five songs from the selected class with the highest decision value and with rating 2. This is presented in Table 1. Then the user rates the results as shown in Table 2.

Table 1: Content-Based Retrieval Results for the ”Led Zeppelin -The Rain Song”

Title	**Decision Value**
1.Pink Floyd - Money	0.97
2.U2 - Whose gonna ride your wild horses	0.95
3.Led Zeppelin - In The Evening	0.92
4.Pink Floyd- The Dogs Of War	0.87
5. Pink Floyd - Run Like Hell	0.86

Table 2: User’s Rating

Title	**User Rating**
1.Pink Floyd - Money	4
2.U2 - Whose gonna ride your wild horses	2
3.Led Zeppelin - In The Evening	2
4.Pink Floyd- The Dogs Of War	3
5. Pink Floyd - Run Like Hell	3

The system uses the average rating of the user ($2.8 \simeq 3$) to return a new list with results from the same class, from ”Rock”, but also with the new rating 3. The new recommendations are presented in the table Table 3.

5 Conclusions

Our work develops a mobile recommendation middleware system. Our system recommends facilitates the access to digital audio/music libraries in push technology-based mobile services. Specifically, our system provides a semi-automatic interface

Table 3: Results from the ”Rock” class with rating 3 and Decision Value $\in [0.95, 0.85]$

Title	Rating
1.Led Zeppelin -In The Evening	3
2.Pink Floyd - Nobody Home	3
3.U2 - Running to stand still	3
4.Led Zeppelin-All My Love	3
5.U2 - I Will Follow	3

that allows mobile users to interact with a digital audio/music library and find/retrieve audio/music files in a flexible way that combines mobile technologies and content-based retrieval techniques. Thus, a mobile user is provided with the ability to query for audio/music files that belong to the same genre by simply sending an example music file from his/her mobile device.

Future work in this area will follow the directions of further evaluation of the recommendation mechanism of system. Additionally, we will explore the development of alternative recommendation mechanisms and effect a comparative evaluation of the advantages and disadvantages of each approach. This and other related work is currently in progress and will be reported shortly.

References

1. Yoon, W.J., Oh, S.H., Park, K.S.: Robust music information retrieval on mobile network based on multi-feature clustering. In: Li, X., Zaïane, O.R., Li, Z.-h. (eds.) ADMA 2006. LNCS, vol. 4093, pp. 279–283. Springer, Heidelberg (2006)
2. Paolo, M., Paolo, A.: Trust-aware recommender systems. In: RecSys 2007: Proceedings of the 2007 ACM conference on Recommender systems, pp. 17–24 (2007)
3. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: Proc. Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, 1998, pp. 43–52 (1998)
4. Cohen, W.W., Fan, W.: Web-collaborative filtering: recommending music by crawling the web. Computer Networks 33(1–6), 685–698 (2000)
5. Rudstrom, A., Svensson, M., Coster, R., Hook, K.: Mobitip: Using bluetooth as a mediator of social context. In: Ubicomp 2004 Adjunct Proceedings (2004)
6. Cohen, W.W., Fan, W.: Web-collaborative filtering: recommending music by crawling the web. Computer Networks 33(1–6), 685–698 (2000)
7. Rojsattarat, E., Soonthornphisaj, N.: Hybrid recommendation: Combining content-based prediction and collaborative filtering, pp. 337–344 (2003)
8. Lampropoulou, P.S., Lampropoulos, A.S., Tsihrintzis, G.A.: Alimos: A middleware system for accessing digital music libraries in mobile services. In: Gabrys, B., Howlett, R.J., Jain, L.C. (eds.) KES 2006. LNCS, vol. 4251, pp. 384–391. Springer, Heidelberg (2006)

9. Lampropoulou, P.S., Lampropoulos, A.S., Tsihrintzis, G.A.: Evaluation of a middleware system for accessing digital music libraries in mobile services. In: 5th International Conference on Information Technology: New Generations, Las Vegas, Nevada, USA (April 2008)
10. Push architectural overview, wap-250-pusharchoverview-20010703-p, wap forum
11. Push access protocol specification, wap-247-pap-20010429-a, wap forum

Video Deinterlacing Algorithm Based on Fuzzy Reasoning with Angle Extraction Approach

Gwanggil Jeon, Yong Fang, Kangjun Lee, Min Young Jung,
Rokkyu Lee, and Jechang Jeong

Abstract In this paper, we present the outline and the formulation of an interpolation technique for designing video deinterlacing by employing a modified Sobel mask integrated with a simple fuzzy edge direction detection algorithm. A linear interpolation is used to interpolate the non-direction-designated regions and a weighted-average method is used to interpolate the direction-designated regions along edge directions. The fuzzy if-then rules are utilized to carry out the determining edge direction. Experimental results have shown a low interpolation error, in comparison with other broadly used video deinterlacing algorithms.

1 Introduction

Interlaced scanning format is used in conventional broadcast to avoid large area flicker, while maintaining a good vertical resolution. With the continuing development of HDTV and the convergence of video formats, the displays nowadays required to allow many different formats and transform them to a standard display format. This is named video format conversion. The format conversion includes

Gwanggil Jeon · Kangjun Lee · Jechang Jeong
Department of Electronics and Computer Engineering, Hanyang University, Korea
e-mail: {windcap315,ee9627,jjeong}@ece.hanyang.ac.kr

Yong Fang
College of Information Engineering, Northwest A&F University, China
e-mail: yfang79@gmail.com

Min Young Jung
Department of Computer Science and Engineering, Hanyang University, Korea
e-mail: jmini@hanyang.ac.kr

Rokkyu Lee
Telechips Inc., 1000-12 Daechi-dong, Gangnam-gu, Seoul, Korea
e-mail: gunboy83@gmail.com

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 369–379.
springerlink.com

changing following attributes of video information: frame size, line size, refreshes rate or image aspect [1], [2]. Furthermore it involves converting an interlaced field to a progressive one.

Deinterlacing (or line doubling) is the conversion of video from an interlaced image to a progressive one. The aim of deinterlacing is to reconstruct the missing pixels in interlaced sequences, decrease aliasing, and increase the vertical resolution of the sequences. As many video formats used for broadcasting, the need for format conversion is increasing critically. The deinterlacing with high performance is, therefore, becoming a significant theme in the field of video technology. There are number of different deinterlacing methods to realize this. These methods can be roughly split into three groups, linier filtering methods [3],[4],[5], adaptive (or nonlinear) methods [6], and motion-compensated methods [7],[8],[9]. Some of these methods have been analyzed in [2].

In this paper, we propose a new video deinterlacing method that is based on the *fuzzy reasoning* and *angle extraction* algorithm. The rest of this paper is composed as follows. In Section 2, we present some definitions and representations. The proposed method is introduced in Section 3. In Section 4, simulation results are obtained to show the feasibility of the proposed design. Finally, Section 5 presents our conclusions.

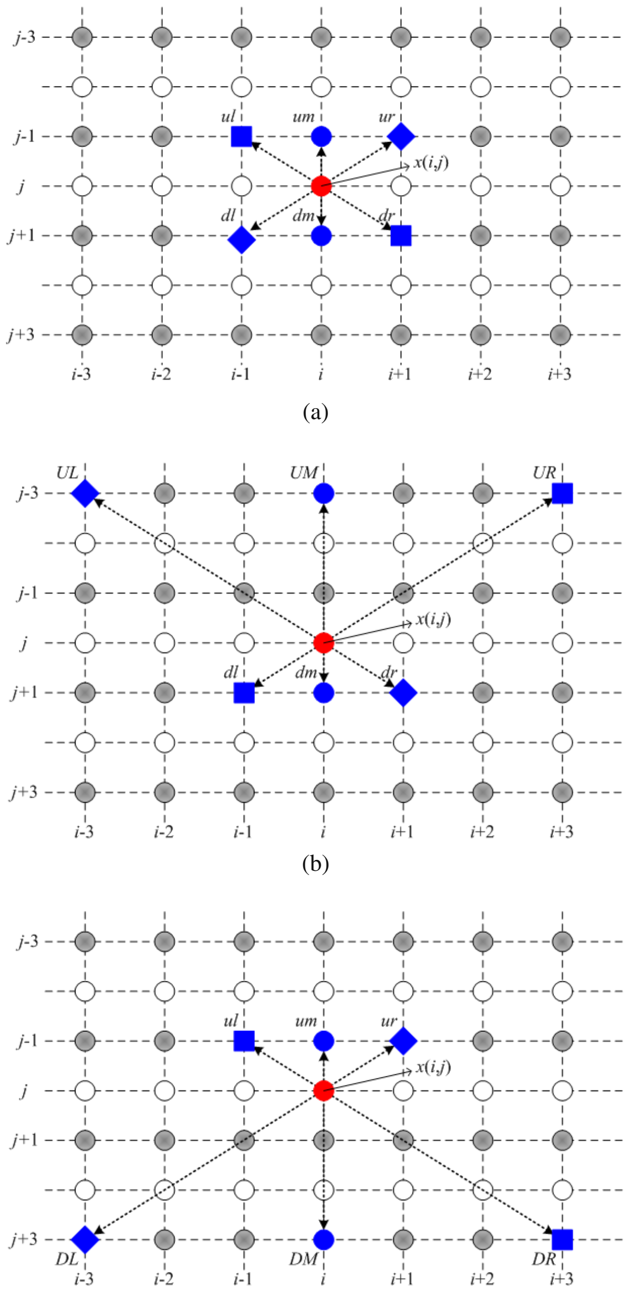
2 Preliminaries

Let $x(i,j)$ denote the signal to be interpolated where i refers to the column number and j to the line number. A 2D localized window is used to calculate directional correlations and to interpolate the current pixel, as shown in Fig. 1, where u, d, l, m, and r represent *up*, *down*, *left*, *mid*, and *right*, respectively. They works as the displacement parameter that the filter is a shifted -1 and +1 in vertical direction and -1 and +1 in horizontal direction. In addition, we use following parameter UL, UM, UR, DL, DM, and DR which are defined as Equation 1.

$$\begin{aligned}
ul &= x(i-1,j-1) & dl &= x(i-1,j+1),\\
UL &= x(i-3,j-3) & DL &= x(i-3,j+3),\\
um &= x(i,j-1) & dm &= x(i,j+1),\\
UM &= x(i,j-3) & DM &= x(i,j+3),\\
ur &= x(i+1,j-1) & dr &= x(i+1,j+1),\\
UR &= x(i+3,j-3) & DR &= x(i+3,j+3).
\end{aligned} \tag{1}$$

The differences in intensity functions can be easily determined as

$$\begin{aligned}
f_{45} &= |ur-dl|, & g_{45} &= |UR-dl|, & h_{45} &= |ur-DL|,\\
f_{90} &= |um-dm|, & g_{90} &= |UM-dm|, & h_{90} &= |um-DM|,\\
f_{135} &= |ul-dr|, & g_{135} &= |UL-dr|, & h_{135} &= |ul-DR|,
\end{aligned} \tag{2}$$



(a) (b) (c)

Fig. 1: A window for the direction-based deinterlacing: (a) f_θ, (b) g_θ, (c) h_θ

where θ={$45^o, 90^o, 135^o$} represents the edge direction. We define minimum and maximum values of f_θ, g_θ, and h_θ, as Equation 3.

$$\begin{aligned} f_{min} &= min(f_{45}, f_{90}, f_{135}), & f_{max} &= max(f_{45}, f_{90}, f_{135}) \\ g_{min} &= min(g_{45}, g_{90}, g_{135}), & g_{max} &= max(g_{45}, g_{90}, g_{135}) \\ h_{min} &= min(h_{45}, h_{90}, h_{135}), & h_{max} &= max(h_{45}, h_{90}, h_{135}) \end{aligned} \tag{3}$$

The candidate deinterlaced pixel (CDP), $x_\theta(i, j)$, where $\theta = 45^o, 90^o, 135^o$ is the average value between any two pixels in the edge direction, which is calculated as follows,

$$x_\theta(i,j) = \begin{cases} \dfrac{ur+dl}{2} & \text{if} \quad f_{45} = f_{min} \\ \dfrac{um+dm}{2} & \text{if} \quad f_{90} = f_{min} \\ \dfrac{ul+dr}{2} & \text{if} \quad f_{135} = f_{min} \end{cases} \tag{4}$$

3 Fuzzy Reasoning, Angle Extraction, and Implementation

To calculate the weight values which will be used to determine the missing pixel value, we derive the nine luminance intensity difference values, i.e., f_θ, g_θ, and h_θ. The differences in luminance intensity of the input pixels are represented as the input variables f_θ, g_θ, and h_θ. The system contains a membership function SMALL for the weight output, denoted as μ_S, is shown in Fig. 2, which is given by Equation 5.

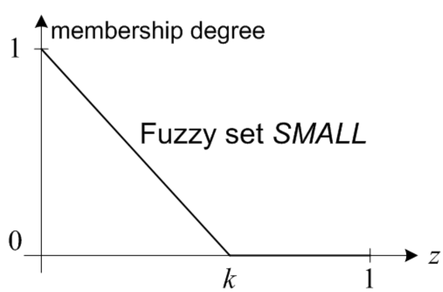


Fig. 2: Membership function SMALL denoted as μ_S and given by Equation 5.

$$\mu_S(Z) = \begin{cases} \dfrac{k-z}{k} & \text{if} \quad z \le k \\ 0 & otherwise \end{cases} \tag{5}$$

where k is preferably set as the value of 0.75 by experiments.

In this proposed fuzzy decision system, there are three output variables, *strongly consistent* (SC), *weakly consistent* (WC), and *inconsistent* (IC). According to the properties of input variables, 27 fuzzy decision rules are used in the proposed fuzzy reasoning system which is designed as Table 1.

Table 1: Extracted rules for the interpolation missing pixel

rule number	if θ of $f_\theta = f_{min}$ is	and if θ of $g_\theta = g_{min}$ is	and if θ of $h_\theta = h_{min}$ is	then the direction is
1	45°	45°	45°	SC
2	45°	45°	90°	SC
3	45°	45°	135°	SC
4	45°	90°	45°	SC
5	45°	90°	90°	WC
6	45°	90°	135°	IC
7	45°	135°	45°	WC
8	45°	135°	90°	IC
9	45°	135°	135°	IC
10	90°	45°	45°	WC
11	90°	45°	90°	SC
12	90°	45°	135°	IC
13	90°	90°	45°	SC
14	90°	90°	90°	WC
15	90°	90°	135°	WC
16	90°	135°	45°	IC
17	90°	135°	90°	SC
18	90°	135°	135°	WC
19	135°	45°	45°	WC
20	135°	45°	90°	IC
21	135°	45°	135°	SC
22	135°	90°	45°	IC
23	135°	90°	90°	WC
24	135°	90°	135°	SC
25	135°	135°	45°	SC
26	135°	135°	90°	SC
27	135°	135°	135°	SC

The weights are calculated as follows,

$$\omega_\theta = \mu_S\left(\frac{f_\theta}{f_{max}}\right) \tag{6}$$

where ω_θ represents the weight obtained from the fuzzy reasoning system for each contributing $x(i,j)$ CDP.

The angle extraction is performed to determine the leading orientation of the edge direction on missing pixel. When angle extraction is operating, the orientation angle of each neighborhood original image pixel is computed. According to Fig. 1, when the orientation angle of $x(i,j)$ denoted as $ang(i,j)$ is computed, the luminance values of the original pixels nearby $x(i,j)$ are used for the following computations:

$$\Delta_X(i,j) = (2ul + 2dl) - (2ur + 2dr) \tag{7}$$

$$\Delta_Y(i,j) = (ul + 2um + ur) - (dl + 2dm + dr) \tag{8}$$

$$ang(i,j) = -\frac{180}{\pi} tan^{-1}\left(\frac{\Delta_X(i,j)}{\Delta_Y(i,j)}\right) \tag{9}$$

The obtained orientation angle of each neighborhood original image pixel is quantized into five sectors such as $ang(i,j) > 157.5^o$, $157.5^o > ang(i,j) > 112.5^o$, $112.5^o > ang(i,j) > 67.5^o$, $67.5^o > ang(i,j) > 22.5^o$, and $22.5^o > ang(i,j)$.

The edge direction based deinterlacing methods are not effective when a false edge direction is determined. The filter output could finally be illustrated, where the operated image is denoted as $x_{prop}(i,j)$, which means weight multiplied results, i.e.,

$$\begin{array}{ll}
\text{if} & (\text{direction is IC}) \\
 & x_{prop}(i,j) = x_{90^o}(i,j) \\
\text{else if} & \left[\begin{array}{l}(\text{direction is IC})\,OR \\ [(\text{direction is WC})\,AND\,(\theta - 22.5^o \le ang(i,j) \le \theta + 22.5^o)]\end{array}\right] \\
 & x_{prop}(i,j) = x_\theta(i,j) \\
\text{else} & x_{prop}(i,j) = x_\Omega(i,j) = \frac{\sum_\theta \omega_\theta \cdot x_\theta}{\sum_\theta \omega_\theta}
\end{array} \tag{10}$$

where Ω represents the final normalized weights extracted from the fuzzy inference system for each contributing $x(i,j)$ CDP.

4 Simulation Results

The experimental results and related analyses of the proposed algorithm are described in this Section. We conducted our experiments on seven real world sequences with a field size of 352×288 for both of objective and subjective performances. They are Akiyo (AK), Flower (FL), Foreman (FO), Mobile (MO), News (NE), Stefan (ST), and Table Tennis (TT). As a measure of objective performance between an interpolated sequence and the original one, we employed the peak signal-to-noise ratio (PSNR) in decibels (dB) which is given in Equation 11.

$$PSNR(x_{org}, x_{rec}) = 10\log_{10}\frac{{I_{max}}^2}{MSE(x_{org}, x_{rec})} \tag{11}$$

where x_{org} and x_{rec} are the original and interpolated sequences of size $width \times height$, respectively. The symbol I_{max} is the maximum possible luminance intensity value (with 8-bit integer values, I_{max} will be 255). This similarity measure is based on another performance measure, the mean-square error (MSE),

$$MSE(x_{org}, x_{rec}) = \frac{\sum_{i=1}^{width} \sum_{j=1}^{height} [x_{org}(i,j) - x_{rec}(i,j)]^2}{width \times height} \tag{12}$$

The video sequences were sub sampled by a factor of two in the vertical direction without anti-aliasing filtering. The original progressive video sequences were used as a reference point to which we compared our deinterlacing method. All the video sequences were changed from progressive format to interlaced one as shown in Fig. 3 according to the different deinterlacing methods [2].

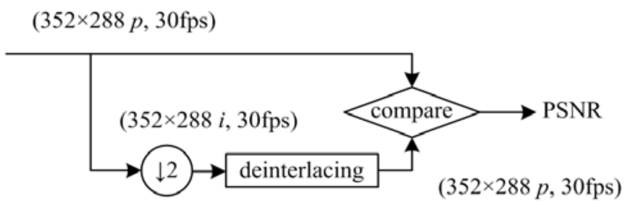


Fig. 3: Performance measurement method.

The set of conventional methods used in the comparison includes ELA [10], EELA [11], MELA [12], DOI [13], NEDD [14], LCID [15], LABI [16], FDED [17], and FDOI [18].

4.1 *Objective Performance Comparisons*

Table 2 and Table 3 show the average PSNR (dB) and computational CPU time (ms) comparisons of each deinterlacing method, over the corresponding video sequences, for seven test sequences.

From the results, we see the proposed method outperforms the other methods in all of the selected sequences in terms of PSNR. In addition, the proposed method needs only 10%, 43%, and 31% average computational CPU time as compared to that of DOI, FDED, and MADLSD methods with a 0.3775dB, 0.0067dB, and 0.7216dB average PSNR gain.

Table 2: PSNR comparison for CIF sequences (dB/frame)

	AK	FL	FO	MO	NE	ST	TT	Average
ELA	37.6815	21.9325	31.3965	23.3409	31.5308	25.9657	31.2361	29.012
EELA	36.7604	21.9588	30.3962	23.2673	31.9229	26.3585	30.8843	28.7926
DOI	39.6058	22.2521	30.2166	24.9101	33.384	26.6883	31.5472	29.8006
NEDD	38.6852	22.3364	30.6379	24.7633	32.687	26.9959	31.6351	29.6773
FDED	39.9255	22.27	31.1383	25.1571	33.5498	27.2761	31.8828	30.1714
MADL.	38.6215	22.3717	30.4243	24.5538	32.5933	26.4629	31.1682	29.4565
LABI	33.8914	20.4972	28.0312	21.3759	28.4069	23.0753	27.5959	26.1248
LCID	39.742	22.4388	31.2011	25.311	33.1898	27.0937	32.0628	30.1485
FDOI	38.868	22.4428	30.7809	24.8739	32.8404	27.1242	31.7855	29.8165
Prop	39.6552	22.4057	30.6353	25.4538	33.6424	27.4818	31.9724	30.1781

Table 3: Average results of CPU time for CIF sequences (ms/frame)

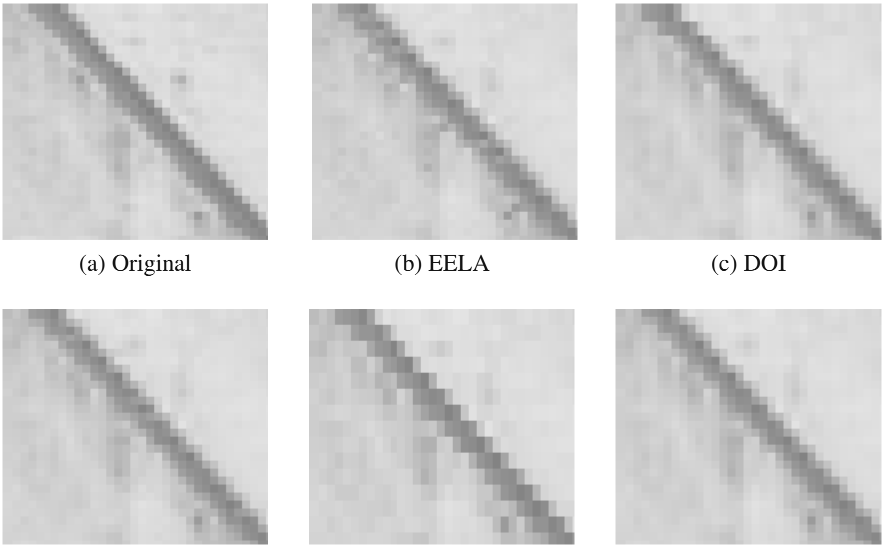
	AK	FL	FO	MO	NE	ST	TT	Average
ELA	23.2974	23.2974	30.9966	19.2979	27.297	23.7974	23.4974	24.4973
EELA	22.6975	18.198	36.796	32.0965	22.8975	27.397	23.1974	26.1828
DOI	171.9811	538.9407	312.4656	648.4287	363.26	663.927	812.4106	501.6305
NEDD	39.2957	35.4961	50.7944	34.9962	42.7953	50.4944	51.2944	43.5952
FDED	93.9897	93.4897	92.9898	109.288	148.7836	132.7854	175.7807	121.0153
MADL.	134.9395	146.2831	138.5892	146.4804	185.2459	188.4024	254.2939	170.6472
LABI	15.7983	15.4983	15.7983	19.9978	19.2979	27.497	38.9957	21.8405
LCID	31.2966	19.7978	31.2966	15.2983	23.2974	30.9966	31.2966	26.1828
FDOI	97.7892	109.488	97.7892	124.9863	109.488	168.2815	175.4807	126.1861
Prop	76.752	31.488	46.248	46.248	46.248	61.992	61.992	52.9954

4.2 Subjective Performance Comparisons

In this section, we show the performances of the five algorithms (i.e., EELA, DOI, NEDD, LABI, and the proposed method) applied to the Foreman sequence. Figure 4(a) shows the original image and the reconstructed images of 60th frame in the Foreman sequence using the five deinterlacing methods. It can be clearly seen that the proposed method generates an image that preserves diagonal edges better than that of other methods. We also found that our proposed method yields more effective visual quality with smoother edges. It reduces staircase artifacts, giving relatively satisfactory image quality. Figure 5 shows the different images between the original and the reconstructed images.

5 Conclusion

In this paper, we proposed a new deinterlacing method that based on fuzzy reasoning and angle extraction approach. The edge direction based deinterlacing methods



(a) Original (b) EELA (c) DOI

(d) NEDD (e) LABI (f) Prop

Fig. 4: The reconstructed images of the 60th frame in the Foreman sequence: (a) the original image, (b) EELA, (c) DOI, (d) NEDD, (e) LABI, and (f) the proposed method

are not effective when a false edge direction is determined. To alleviate this issue, we first differentiate a pixel into non-direction-designated region and direction-designated region. The fuzzy if-then rules are utilized to carry out the determining edge direction. After that we use linear interpolation to interpolate the non-direction-designated regions and use weighted-average method to interpolate the direction-designated regions. The proposed method was tested on several video sequences and fond to provide visually better images than the traditional deinterlacing methods.

Acknowledgments

This research was supported by Seoul Future Contents Convergence (SFCC) Cluster established by Seoul R&BD Program.

References

1. Jack, K.: Video demystified - A handbook for the digital engineer, 4th edn. Elsevier, Jordan Hill (2005)
2. Bellers, E.B., De Haan, G.: De-interlacing: a key technology for scan rate conversion. Elsevier, Amsterdam (2000)

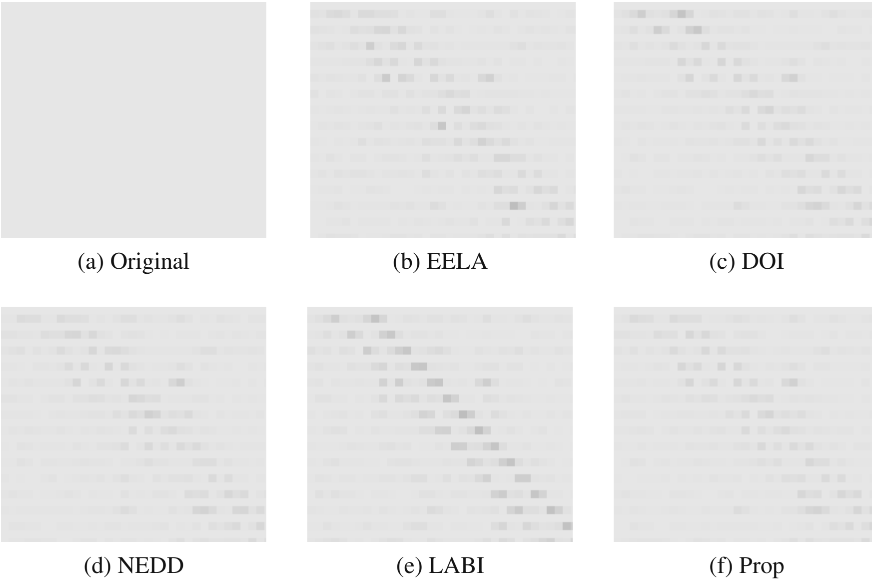


Fig. 5: The difference images of the 60th frame in the Foreman sequence: (a) the original image, (b) EELA, (c) DOI, (d) NEDD, (e) LABI, and (f) the proposed method

3. Thomas, G.: A comparison of motion-compensated interlace-to-progressive conversion methods. Signal Process.: Image Commun. 12, 209–229 (1998)
4. Tekalp, A.M.: Digital Video Processing. Prentice-Hall, Englewood Cliffs (1995)
5. Wang, Y., Ostermann, J., Zhang, Y.-Q.: Video Processing and Communications. Prentice-Hall, Englewood Cliffs (2002)
6. Bhatt, B., Templin, F., Cavallerano, A., Gornstein, V., Hogstrom, B., Derovanessian, H., Lamadrid, S., Mailhot, J.: Grand alliance HDTV multi-format scan converter. IEEE Trans. Cons. Elect. 41(4), 1020–1031 (1995)
7. De Haan, G., Bellers, E.B.: De-interlacing of video data. IEEE Trans. Cons. Elect. 43(3), 819–825 (1997)
8. Mohammadi, H.M., Langlois, P., Savaria, Y.: A five-field motion compensated deinterlacing method based on vertical motion. IEEE Trans. Cons. Elect. 53(3), 1117–1124 (2007)
9. Sugiyama, K., Nakamura, H.: A method of de-interlacing with motion compensation interpolation. IEEE Trans. Cons. Elect. 45(3), 611–616 (1999)
10. Doyle, T.: Interlaced to sequential conversion for EDTV applications. In: Proc. 2nd Int. Workshop Signal Processing of HDTV, Feburary 1990, pp. 412–430 (1990)
11. Chen, T., Wu, H.R., Yu, Z.H.: Efficient deinterlacing algorithm using edge-based line average interpolation. Opt. Eng. 39(8), 2101–2105 (2000)
12. Kim, W., Jin, S., Jeong, J.: Novel intra deinterlacing algorithm using content adaptive interpolation. IEEE Trans. Cons. Elect. 53(3), 1036–1043 (2007)
13. Yoo, H., Jeong, J.: Direction-oriented interpolation and its application to de-interlacing. IEEE Trans. Cons. Elect. 48(4), 954–962 (2002)

14. Park, M.K., Kang, M.G., Nam, K., Oh, S.G.: New edge dependent deinterlacing algorithm based on horizontal edge pattern. IEEE Trans. Cons. Elect. 49(4), 1508–1512 (2003)
15. Chen, P.-Y., Lai, Y.-H.: A low-complexity interpolation method for deinterlacing. IEICE Trans. Inf. and Syst. E90-D(2) (Feburary 2007)
16. Lee, D.-H.: A new edge-based intra-field interpolation method for deinterlacing using locally adaptive-thresholded binary image. IEEE Trans. Cons. Elect. 54(1), 110–115 (2008)
17. Michaud, F., Le Dinh, C.T., Lachiver, G.: Fuzzy detection of edge-direction for video line doubling. IEEE Trans. Circuits and Syst. Video Technol. 7(3), 539–542 (1997)
18. Jeon, G., Jeong, J.: Fuzzy rule and Bayesian network based line interpolation for video deinterlacing. IEICE Trans. Commun. E90-B(6), 1495–1507 (2007)

Semantic Similarity Based Video Retrieval

Min Young Jung and Sung Han Park

Abstract In this paper, we propose semantic similarity measure to overcome semantic gap in video retrieval. In particularly, our method is feature selection for the video ontology construction. Video ontology is aimed at bridging of the gap between the semantic nature of user queries and raw video contents using scene keyword. Moreover, results of semantic retrieval show not only the concept of topic keyword but also a sub-concept of the topic keyword using semantic query extension. Through this process, recall is likely to provide high accuracy results in our method. The experiments compared with keyframe-based indexing have demonstrated that this proposed scene-based indexing presents better results in several kinds of videos.

1 Introduction

Over the last decade, multimedia content indexing and retrieval has been influenced by the important progress in numerous fields, such as digital content production, archiving and information retrieval [1]. Content-based research has steadily advanced after the development of query by image and video content (QBIC) systems that explore multimedia data efficiently [2]. However, early content-based algorithms, such as QBIC systems, focused on automatic information extraction. And the metadata are generated in advance to speed up the retrieval of the lower-level features. The MPEG-7 standard provides a multimedia description scheme (MDS) that describes low-level features and high-level features [3]. However, the relationship between these types of features has not been described so far. In addition, the overall scene context cannot be accessed because the scene is indexed with a single keyword or content. There have been several investigations related to ontology building using hierarchical tree structures of terms. A structured ontology can determine the meaning of certain terms and their actual meaning in relation to

Min Young Jung · Sung Han Park
Department of Computer Science and Engineering, Hanyang University,1271, Sa-3-dong, Sangnok-gu, Ansan, Gyeonggi-do, Korea e-mail: {jmini,shpark}@hanyang.ac.kr

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 381–390.
springerlink.com

vocabulary. They can also infer alternative meanings of terms. On the basis of ontology, several studies have been made on sets of high-level semantic concepts. The first set is generally referred to as LSCOM-Lite [4]. It is composed of 39 concepts and was released by NIST in conjunction with the TRECVID 2005 [5] data set. The second sets of concepts was created by the MediaMill group in the Netherlands [6]. Recently, the LSCOM consortium finished a manual annotation effort for 1,000 concepts [4]. This brings concept detection within reach of research in ontology engineering. Ontology provides background knowledge about various topics. Examples of these types of ontology are the medical subject headings (MeSH) ontology and the gene ontology [7], which are confined mainly to a limited domain. In addition, domain-independent knowledge structures such as WordNet [8] and Cyc [9] can be found as well, yet their knowledge-based structure has been largely crafted after leaning upon the interaction with users. The concept of knowledge-based structure which grants the semantic relationship has just begun to be employed for video retrieval. Ontologies have been used with the same purpose as well. Hoogs describes low-level and high-level features using extended WordNet [10],[11]. But WordNet creates a structure that is expensive and complicated. Moreover, the indexing structure characterizes only the keyframe image so that the image solely, not the video, is retrieved. Furthermore, all processes rely upon manual handling. As a method for supplementing it, Bertini concentrates on the sports domain, particularly soccer, and his method automatically extracts and features information on the events occurred in a soccer match [12]. The drawback here is the limited applicability of such technique. To solve this problem, the semantic gap between low-level and high-level features must be bridged. Therefore, in this paper a multimedia retrieval system that employs back-ground knowledge structured by means of ontology or a thesaurus is introduced and evaluated. This paper means to propose a general semantic indexing structure by using ontology. In Section 2, we present an overview of a video retrieval system. The proposed method is introduced in Section 3. Section 4 shows semantic retrieval. A simulation study aimed at comparing the performance of the different methods is presented in Section 5. Finally, conclusions are derived in Section 6.

2 The Video Ontology System

In this section, let us introduce video ontology system to overcome semantic gap in video retrieval. As shown in Fig. 1, the novel video ontology system has four major components: the semantic scene creation, the video analysis stage, the video ontology creation and the semantic scene search. First, creating the semantic scene requires having the video in the scene unit for modeling and searching purposes. Namely; the semantic scene uses the basic unit of ontology construction and retrieval. Subsequently, the video is analyzed by using the multimedia feature information provided in the MPEG-7 Standard. Then, the video ontology is built on the basis of the analyzed video of the feature information. Finally, the semantic

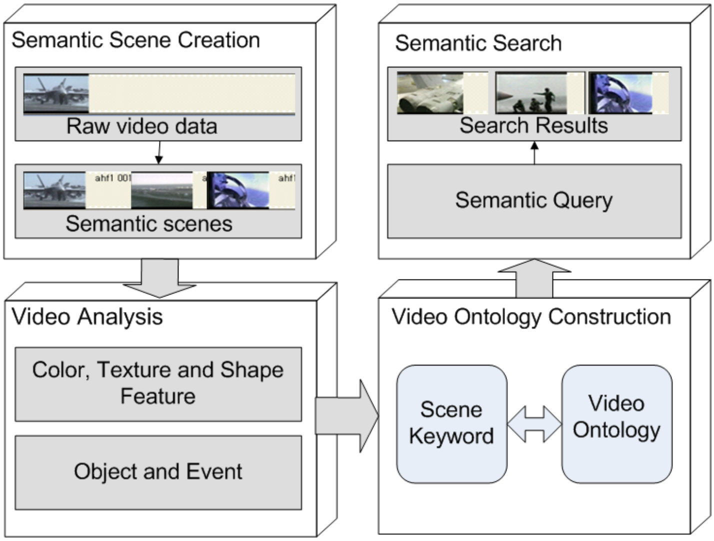


Fig. 1: Ontologies system for semantic video retrieval.

search displays the result of related semantic scenes via the implemented ontology-based approach. In this paper, we assumed that proposed system had performed semantic scene creation and video analysis. In this paper, we propose video ontology construction and semantic search method.

3 Proposed Video Ontology Construction

The video ontology is aimed at defining relationships between the visual features and the semantic concepts. The creation stage is pictured in Fig. 2. As I mentioned before, the semantic scene includes the shot-level information. Each shot extracts key frames. The low-level feature is drawn out of the key frame image, which contains the object and the event defines the relation to key frame. But our system don't describe visual feature of all object. We store only visual feature of main object. Namely, main object and event information stored with the video ontology describe the low-level feature. The object itself it characterized by low-level features such as color, shape and texture. The event is described by the direction and motion of the object. The semantic scene is defined as the Scene keyword(SK). The Scene keyword is defined as Eq.(1):

$$SK(i) = \{C_i, S_i, T_i, M_i\} \tag{1}$$

Index i represent semantic scene. Where C_i, S_i, and T_i denotes color, shape and texture feature vector, respectively. M_i represent motion direction. In particular, the

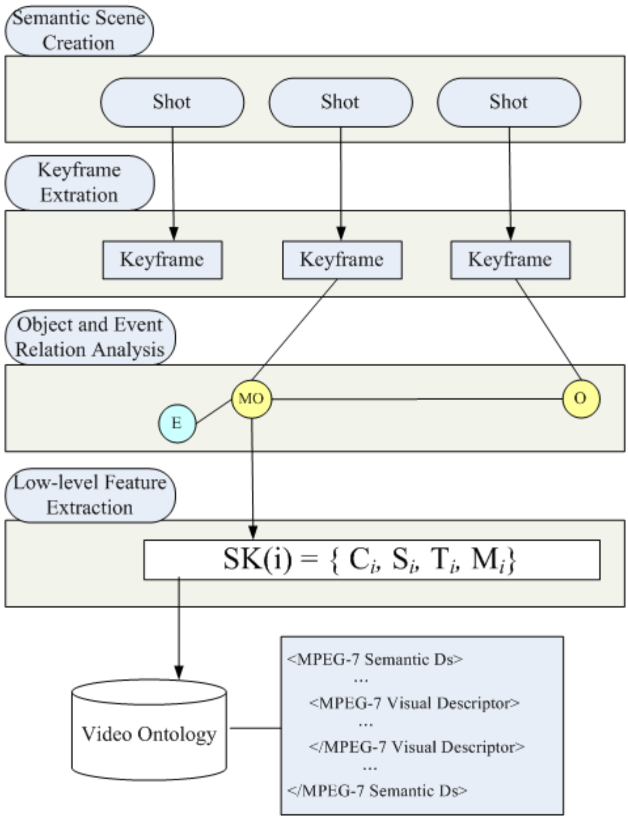


Fig. 2: Video ontology construction.

event can not be entirely defined with one key frame, but with the relationship among objects which exist in each key frame image. That is, the direction of the event is defined as the movement of the same object between shots whereas the velocity is defined based on the time interval between shots. The metadata structure of the video ontology not only describes the relationship between the object and the event using MPEG-7 Semantic Ds but also stores low-level feature by means of MPEG-7 visual descriptor.

Structure of video ontology is constructed bridging of semantic word and visual feature as shown Fig.3. To do this, video ontology store scene keyword of main object. First of all, the video ontology system has the dictionary of term. Through the hierarchical structure of term, the semantic concepts are defined. Then, the new concept can be derived by defining the hierarchical relationships of these terms. Term is defined by scene keyword. The scene keyword internal structure provides the storage for the list of term (subjects) to be used during the semantic scene indexing. Fig. 3(A) shows how an event relates to the main object. The left child node

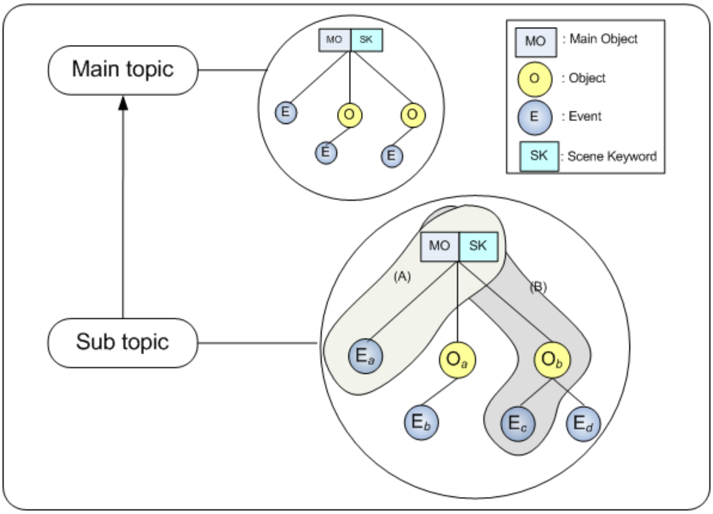


Fig. 3: The tree structure of video ontology.

becomes an event of the main object whereas its right sibling is an object which relates to the main object. The shaded region labeled as (B) depicts the relationship between another object O_b and the main object. The left child of O_b stands for an event in relation to its parent node while its right child represents an event of object(O_b) or to node of object relate with object(O_b).

4 Semantic Retrieval

In this section, we propose semantic query creation and semantic similarity using proposed structure in the previous chapter.

4.1 Semantic Query Creation

We proposed a system for text query based searching. Video ontology is indexed by scene keyword in scene unit. Semantic query creation appears in Fig 4. User input the query of text type. We search matching text of main-concept in video ontology. And then, Scene keyword query is generated. Scene keyword is main object of main concept.

Those semantic scenes with a matching result include not only the concept of the topic keyword but also a sub-concept of it. After a user query is executed, a semantic query will not be created unless the retrieved results do not exist. This is more clearly explained in Table 1.

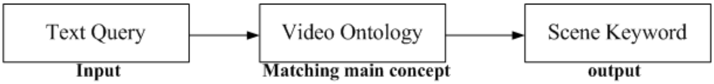
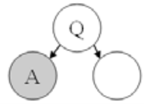
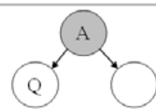
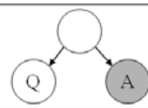


Fig. 4: Procedure of semantic query creation.

Table 1: Extended semantic query

Query	Location of search node	
In case of existing child node	child	
In case of leaf node	parent	
	sibling	

A group of semantic queries give rise to an extended user query in the video ontology. Using user query, the semantic query is created in place of the standard query in the video ontology. If the child node exists in the video ontology, we create the semantic query that is able to retrieve it. In case of a leaf node, a semantic query for retrieving both the parent node and the sibling node is created.

4.2 Semantic Similarity Measure

In this section, we propose similarity measure for semantic retrieval. It is difficult to similarity measure using conventional method. Previous method is not suitable for similarity measure using multiple features. Semantic similarity is important in semantic retrieval. To compute semantic similarity, we extract scene information from keyframe of shot and construct a scene keyword. We assume that all features were tied to the extraction. We extract color information from main object i of keyframe and construct a color vector $C_i(c_{1,i}, c_{2,i}, \ldots, c_{p,i}, \ldots c_{w,i})$ .We could evaluate the degree of color similarity between object i of scene keyword and object j of scene keyword, That is,

$$sim_{color}(i,j) = \frac{\overrightarrow{i_c} \bullet \overrightarrow{j_c}}{|\overrightarrow{i_c}| \times |\overrightarrow{j_c}|} = \left\{ \frac{\sum_{p=1}^{k} c_{p,i} \times c_{p,j}}{\sqrt{\sum_{p=1}^{k} c_{p,i}^2} \times \sqrt{\sum_{p=1}^{k} c_{p,j}^2}} \right\} \tag{2}$$

We could calculate the shape, texture and motion features. The computation of shape, texture and motion similarity is using MPEG-7 visual descriptor which is defined as Eq.(3)-(5). Formula is similar to Eq. (2).

$$sim_{shape}(i,j) = \frac{\overrightarrow{i_s} \bullet \overrightarrow{j_s}}{|\overrightarrow{i_s}| \times |\overrightarrow{j_s}|} = \left\{ \frac{\sum_{p=1}^{k} s_{p,i} \times s_{p,j}}{\sqrt{\sum_{p=1}^{k} s_{p,i}^2} \times \sqrt{\sum_{p=1}^{k} s_{p,j}^2}} \right\} \tag{3}$$

$$sim_{texture}(i,j) = \frac{\overrightarrow{i_t} \bullet \overrightarrow{j_t}}{|\overrightarrow{i_t}| \times |\overrightarrow{j_t}|} = \left\{ \frac{\sum_{p=1}^{k} t_{p,i} \times t_{p,j}}{\sqrt{\sum_{p=1}^{k} t_{p,i}^2} \times \sqrt{\sum_{p=1}^{k} t_{p,j}^2}} \right\} \tag{4}$$

$$sim_{motion}(i,j) = \frac{\overrightarrow{i_m} \bullet \overrightarrow{j_m}}{|\overrightarrow{i_m}| \times |\overrightarrow{j_m}|} = \left\{ \frac{\sum_{p=1}^{k} m_{p,i} \times m_{p,j}}{\sqrt{\sum_{p=1}^{k} m_{p,i}^2} \times \sqrt{\sum_{p=1}^{k} m_{p,j}^2}} \right\} \tag{5}$$

Along with increasing of k, accuracy will increase also. In our case, we choose k=10. Semantic retrieval is based on scene similarity. To calculate scene similarity, we compute weight value which is defined as Eq.(6). Each weight value is determined by the impact of semantic retrieval.

$$sim_{scene} = \{w_c \times sim_{color}\} + \{w_s \times sim_{shape}\} + \{w_t \times sim_{texture}\} + \{w_m \times sim_{motion}\} \tag{6}$$

5 Simulation Results

In this section, we present some experimental results. We compared our proposed method, called semantic-based video retrieval (SBVR) system with another well-known ontology-based video retrieval (OBVR) system [11]. Conventional OBVR system indexes the low-level and the high-level features of the key frame. Our video retrieval experiments were conducted on 100 "use case" queries to uncover the relationship between the numbers of concepts, detection accuracy, and retrieval performance. These 100 queries were used because truth annotation for relevant shots was available for each query on the same TRECVID 2005 video development collection, which had also been annotated with truth for the concept lists. The experiments were conducted in such a way that there is an increase in the number of constructed data. The metadata of the implemented ontology are parsed to be easily retrieved and then stored at the database with a structured table format. The effectiveness of retrieval is usually measured by the following two indicators, recall and precision. There are few differences in terms of precision and test results when the amount of tested data is not enough. But if data are gathered more and more, the tests can yield a better precision favoring the semantic-based retrieval approach. This is due to the fact that the SBVR method defines not only a key frame but also a

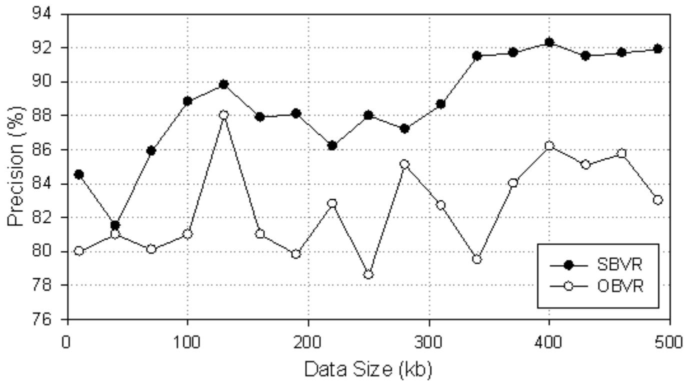


Fig. 5: Precision of video retrieval.

relationship of shots included in a scene. Hence the SBVR method is able to provide more precise results than the OBVR method as shown in Fig. 5.

The recall measure denotes the actually retrieved rate among all related experiment data concerning video retrieval. The OBVR algorithm is interesting in the sense that it indexes only the key frame while the proposed SBVR method describes the relation of frame in the scene. The SBVR method can retrieve related scenes which are constructed by way of the ontology. Therefore the proposed method exhibits higher recall rates, as reported in Fig. 6. After adopting SBVR, we observe that the text is matched precisely. That is to say, even though a keyword of the semantic scene does not match, it can be searched similarly provided it is included as one topic category. Put differently, though a keyword of the semantic scene does not match exactly, only if the scene to be probed is categorized into a single topic, it will be included into the search results.

6 Conclusions

In this paper, we have proposed a system for ontology-based semantic indexing and retrieval of content. When the low-level feature is well extracted, it shows the ontology unit as the semantic retrieval results as well as the scene unit. The video ontology method supplements the structural disadvantages of metadata contained in the existing MPEG-7 standard. Current MPEG-7 structure is the storage structure in which the semantic information and low-level information become independent. As it is proposed a new structure, the problem was supplemented. The precision of retrieval was improved through the domain restriction by the scene name ontology. Also, the video can be semantically recognized as the scene unit via the scene unit ontology. If large video ontologies are constructed, videos will seem to be analyzed

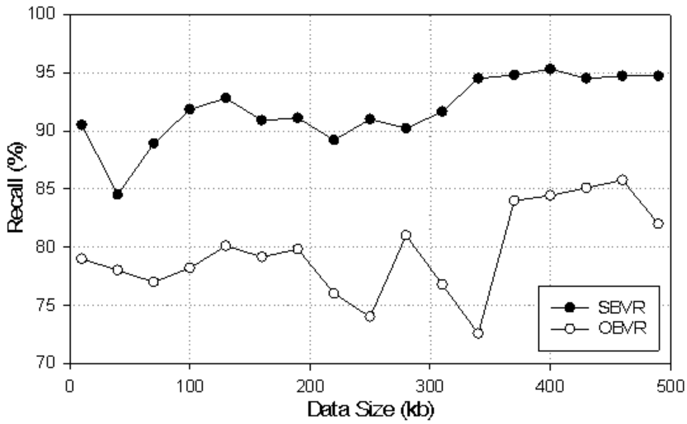


Fig. 6: Recall of video retrieval.

automatically. We believe that this approach is in the right direction for bridging the current semantic gap of content interpretation between humans and computers, which is the main hurdle for a wide expansion of multi-media in the Semantic Web.

References

1. Baeza-Yates, R.A., Ribeiro-Neto, B.A.: Modern information retrieval. ACM Press/Addison-Wesley (1999)
2. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. on Pattern Analysis and Machine Intelligence 22(12), 1349–1380 (2000)
3. Chang, S.-F., Sikora, T., Purl, A.: Overview of the MPEG-7. IEEE Trans. on Circuits and Systems for Video Technology 11(6), 688–695 (2001)
4. Naphade, M., Smith, J.R., Tesic, J., Chang, S.-F., Hsu, W., Kennedy, L., Hauptmann, A., Curtis, J.: Large-scale concept ontology for multimedia. IEEE Multimedia 13(3), 86–91 (2006)
5. Over, P., Laneva, T., Kraaij, W., Smeaton, A.F.: Trecvid 2005-An overview. In: Proc. TRECVID 2005 (2005)
6. Snoek, C.G.M., Worring, M., van Gemert, J.C., Geusebroek, J.-M., Smeulders, A.W.M.: The challenge problem for automated detection of 101 semantic concepts in multimedia. ACM Multimedia (2006)
7. Tveit, H., Mollestad, T., Lægreid, A.: The alignment of the medical subject headings to the gene ontology and its application in gene annotation. In: Tsumoto, S., Słowiński, R., Komorowski, J., Grzymała-Busse, J.W. (eds.) RSCTC 2004. LNCS, vol. 3066, pp. 798–804. Springer, Heidelberg (2004)
8. Fellbaum, C.: WordNet: An Electronic Lexical Database. Bradford Book (1998)
9. Lenat, D., Guha, R.V.: Building Large Knowledge-Based Systems: Representation and Inference in the Cyc. Project. Addison-Wesley, Reading (1990)

10. Hoogs, A., Rittscher, J., Stein, G., Schmiederer, J.: Video content annotation using visual analysis and a large semantic knowledgebase. In: Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 327–334 (2003)
11. Hollink, L., Worring, M., Schreiber, A.T.: Building a visual ontology for video retrieval. In: Proc. of the ACM Multimedia, November 2005, pp. 479–482 (2005)
12. Bertini, M., Cucchiara, R., Del Bimbo, A., Grana, C.: MOM: Multimedia Ontology Manager. A Framework for Automatic Annotation and Semantic Retrieval of Video Sequences. In: Proc. of the ACM Multimedia, S. Barbara, CA, USA, October 23-27 (2006)

Fast and low cost 3d object recognition

Cesare Bornaghi, Beniamino Ferrari, Ernesto Damiani

Abstract This paper presents a 3D object recognition method aimed to industrial applications. The proposed method compares any object represented as a set of 3D polygonal surfaces through their corresponding normal map, a bidimensional array which stores local curvature (mesh normals) as the pixels RGB components of a color image. The recognition approach, based on the computation of a difference map resulting from the comparison of normal maps, is simple yet fast and accurate. First results show the effectiveness of the method on a database of 3D models of sanitary equipments.

1 Introduction

The term 3D object recognition usually refers to a recognition methodology operating on dataset representing object shapes as range data or polygonal meshes. For sanitary equipment, in particular, the 3D approach is interesting for many reasons, including a more accurate representation of object features that leads to a potentially higher discriminating power and promises to better cope with the large amount of variations typical of industrial manufacturing of toilets. Early researches on 3D object recognition (expecially faces set) were conducted over a decade ago as reported from Bowyer et al. [1] in their recent survey on this topic and many different approaches have been developed over time. Indeed, while some approaches work best

Cesare Bornaghi
University of Milan, DTI, via Bramante 65, 26013 Crema, Italy
e-mail: cbornaghi@crema.unimi.it

Beniamino Ferrari
via Gorizia 20, 26013 Crema, Italy
e-mail: sunch1ld@hotmail.com

Ernesto Damiani
University of Milan, DTI, via Bramante 65, 26013 Crema, Italy
e-mail: damiani@dti.unimi.it

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 391–399.
springerlink.com

on medium-sized objects, i.e. faces showing a standard relaxed expression, others aim to cope with shape variations due to random object positions. These methods are based on feature extraction and they represent surface curvature and metric size properties as a point in feature space, whose distance to other points (faces) measures model similarity [2]. Various extensions of 2D recognition techniques to range images have been proposed, such as those based on eigenfaces [3] or Hausdorff distance matching [4]. Other works compare objects through a spherical correlation of their Extended Gaussian Image [5], or through Principal Component Analysis (PCA) [6, 7, 8], or even measure the distance between any two 3D surfaces by the Iterative Closest Point (ICP) method [9, 10]. To increase recognition rate in case of deformed surface (expression variation) Bronstein et al. [11] apply canonical images to 3D face analysis, while other authors combine 3D and 2D similarity scores obtained comparing 3D and 2D objects profiles [12], or extract a feature vector combining Gabor filter responses in 2D and point signatures in 3D [13]. Here, we present a 3D object recognition method aimed at industrial applications and based on normal map [15] , a bidimensional array representing local curvature of a 3D polygonal mesh in terms of RGB color data. This novel approach features a high recognition precision, a good robustness to a broad range of expression variations and a low comparison time allowing to efficiently work on a very large object database. This paper is organized as follows. In section 2. the proposed methodology is presented in detail. In section 3. experimental results are shown and briefly discussed. The paper ends with section 4. outlining directions for future research.

2 Proposed method

A normal map is simply an RGB color image providing a bidimensional representation of a 3D surface, in which each normal to each polygon of a given mesh is represented by a RGB color pixel. More precisely, each pixel in a normal map encodes the three scalar components of the normal to a particular polygon in the mesh surface using the three RGB channels usually associated to a color image. To this aim we project the 3D geometry onto 2D space through medium object mapping. The result is a bidimensional representation of original object geometry which retains spatial relationships between object features. It allows the comparison of two 3D surfaces through corresponding 2D color images in a fast and accurate way. As shown in the flow-chart in Figure 1, after each object has been enrolled and its surface mesh generated, the preprocessing pipeline provides the corresponding normal map through mesh 2D projection and sampling. The comparison between any two object is therefore performed by computing a new array called the difference map, which is a gray scale image obtained subtracting pixel by pixel the two corresponding normal maps, multiplied by a previously built possible feature weighting mask. The whole recognition process is discussed in detail in the following section.

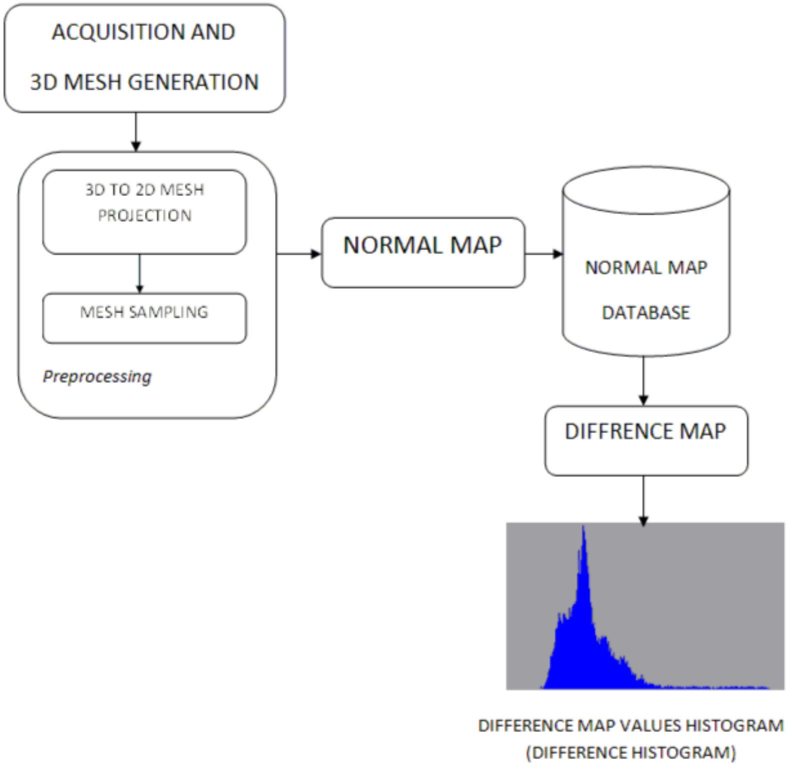


Fig. 1: Algorithm execution flow chart.

2.1 *Object Acquisition and Mesh Generation*

As the proposed method works on 3D polygonal meshes we firstly need to acquire objects and represent them as polygonal surfaces. Laser or structured light scanning technology could be used to this purpose, capturing range data and then converting them in 3D vertices and polygons. Stereoscopic imaging is another feasible option for 3D object digitizing which relies on 2D images shot by known angles and reconstructs a 3D mesh conform to a set of feature points previously correlated between the different views. Even if any of these technologies could provide valid 3D data for the presented method, we opted for a feature based mesh warping technique because it requires a much simpler equipment more likely to be adopted in a real application (a digital cameras shooting from side position) and, though the resulting object shape accuracy is inferior compared to real 3D scanning, it proved to be sufficient for recognition. A more detailed description of enrollment procedure adopted during the experiments of this study is given in the next section.

2.2 *Laser scanning*

Object database is generated by a set of sample objects, which are acquired using a laser scanning system based on laser reflection. We used a calibrated laser scanning system based on reference point calibration. We prepare acquisition scenario in this way: a camera in front of the object to recognize laser reflection. A laser ray is located in front of the object and a panel for preventive laser calibration behind the object.

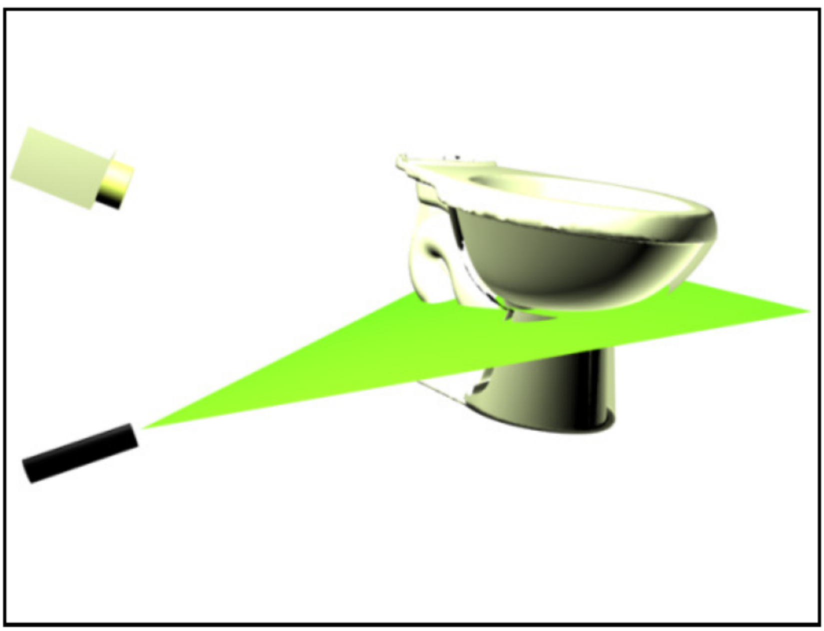


Fig. 2: Simple scheme of laser scanner acquisition method.

Laser scanner used during acquisition phase provides noise management algorithm. It provides quality mesh models. This models are used without change in recognition chain.

2.3 Mesh 3D to 2D Projection

Object features affect local curvature on a shape, so they could be considered as a signature of the object. As local curvature of a polygonal mesh is faithfully represented by polygon normals, we intend to represent these normals by a color image transferring object's 3D features in a 2D space. To this aim, we first need to project the vertex coordinates onto a 2D space, a task which can be thought as the inverse of the well known texture mapping technique. We used a planar projection (opportunely adapted to mesh size). Formally, given an arbitrary mesh M , we want associate to each mesh vertex v_i with coords (x_i, y_i, z_i) in R3 the corresponding RGB values. We make a matrix containing correspondent values: x=R, y=G, z=B. After that we normalize values on 8 bit, that maintain each coordinate value between 0 and 255. We decide to normalize each value using an ordered couple for each matrix cell (s_i, t_i), with $0 \leq s, t \leq 1$ The resulting object geometry is represented by a generic mesh, projected onto the domain (S, T)[14] .

2.4 Mesh Sampling

At this point we can store normals of mesh in a bidimensional matrix with dimension , to represent object geometry. To this purpose we have to sample the previously mapped geometry and quantize the length of the three scalar components of each normal. Therefore we assign to each pixel (i, j) in N, with and , the three scalar components of the normal to the point of the mesh surface with mapping coordinates $(i/l, j/m)$. The resulting sampling resolution is $1/l$ for the s range and $1/m$ for the t range. The normal components are stored in pixel (i, j) as RGB color components. We refer to the resulting matrix N as the normal map of mesh M. A normal map with a standard color depth of 24 bit allows 8 bit quantization for each normal component, providing $360^\circ/256 = 1.4^\circ$ of angular precision which proved to be adequate to the recognition task. As the 24 bit format is the industry standard for color images, there is a practical advantage using it for normal maps storing three 8 bit normal components for each image pixel. Anyway we tested the method with both 4 and 16 bit quantization. While 4 bits are clearly not enough for the purpose (with $360^\circ/16 = 22.5^\circ$ of angular precision), 16 bits did not provide substantially better performances in terms of recognition accuracy.

2.5 Classification and comparison Methodology

When the sampling phase is completed, we can register the new object, i.e. its normal map, in the reference database, or perform a search through it to find a

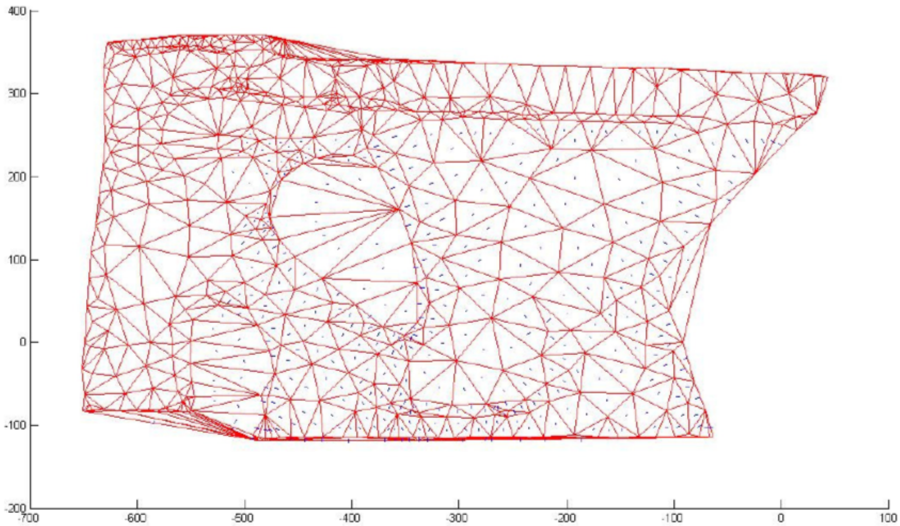


Fig. 3: Mesh example.

matching object. To compare any two object meshes M_A and M_B we compute the angle included between each pairs of normals represented by colors of pixels with corresponding mapping coordinates, and store it in a new map. As each pixel (xN_A, yN_A) in N_A has corresponding color components (rN_A, gN_A, bN_A) and each pixel (xN_A, yN_A) in N_B has corresponding components (rN_B, gN_B, bN_B) the angle included between the normals represented by each pair of pixel with $xN_A = xN_B$ and $yN_A = yN_B$ is given by:

$$\theta = arcos(rN_A \cdot rN_B \cdot gN_A \cdot gN_B \cdot bN_A \cdot bN_B) \tag{1}$$

with components opportunely normalized from color domain to spatial domain, so $0 \leq rN_A, gN_A, bN_A \leq 1$ and $0 \leq rN_B, gN_B, bN_B \leq 1$. The angle θ with $0 \leq \theta \leq \pi$ normalized within [0-255] range, is stored as 8 bit value in a bidimensional $m \times n$ array D which can be viewed as a gray-scale image.

3 Experiments and discussions

As one of the goals of our experiments was to test the performance of the proposed method in a realistic operative environment, we decided to build a 3D object database in which insert 2D color pictures instead of range data. Objects used building database are several kind of toilettes that have to be recognized. More precisely, every object model in the database has been created translating and rotating each singular polygonal model to closely fit a set of features extracted from 3D laser acquisition model of each individual enrolled in the system. We adopted the mesh warping technique described in [15], so for each enrolled object a set of corresponding features extracted by a structured snake method from the two views are correlated first and then used to guide the prototype mesh warping performed through a Dirichlet Free Form Deformation. The prototype model mesh used in the dataset has about 3K (post clustering) triangular facets, and even if it is possible to use mesh with higher level of detail, we found this resolution to be adequate to the recognition purpose. This is mainly due to the optimized tessellation which privileges key areas, produced by 3D scanner features almost evenly spaced vertices. We chosen a set of different toilettes that have to be recognized in Industrial application in glaze automation. The full database includes 7 different toilets in 20 different positions (140 models in total). We made up an array of weighted points for feature selection and recognition. Than we tested the proposed recognition method on two sets, with and without an interest point weighting mask, evaluating the recognition rate through a one-tomany comparison performed on a probe set of objects with various features and positions. The results are generally better than those obtained by many 2D algorithms but an effective comparison would require a dataset featuring both 2D and 3D data, better if acquired during the same enrolment session. It is important to underline that 3D recognition system are better than 2D systems because they extract much features. Using this approach we can use 2D data structures for 3D features make robust recognition in lightweight computational way. More precisely, on the 1st gallery (feature based warping, 140 toilettes) the experiments showed a recognition rate and a cumulative match score (cms) reaching 100%, while recognition success on the second gallery (laser scanning, 6 objects jects) were slightly inferior with cms=97.8, probably due to residual artifacts and noise in the captured geometry. We used normal maps with a fixed size which proved to be adequate for recognition purposes while the Gaussian function is parameterized by fixed σ and fixed k , depending by heuristic data coming from tests [15]. The second testing session is meant to verify recognition performance in case of feature variations. For this experiment the probe set is medium toilette while the gallery set includes all the available variations of each toilette in the gallery database. Precision/recall results without the feature points weighting mask showed that the method picked a variation of the probe model up to the 4th retrieval. The proposed recognition algorithm proved to be very fast, requiring approx. 5 seconds on a P4/3.4 Ghz based PC for a single one to one comparison, a figure allowing to use the method in a one to many application even on a large gallery. Calculating the gap from alignment algorithm, we found the score approximation for each object in the database as shown in Figure4.

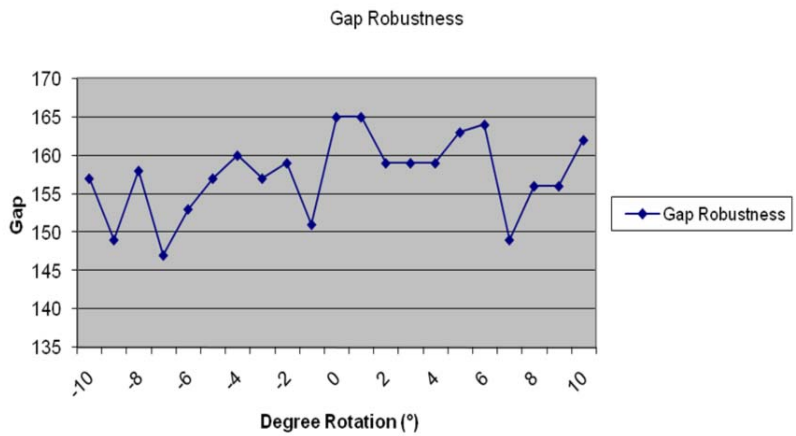


Fig. 4: Gap robustness Graph related to Object form and positions in database.

4 Concluding remarks

We presented a novel 3D object recognition approach based on normal map, a 2D array representing object surfaces local curvature, aimed to industrial applications. It proved to be simple, robust to object variations, fast and with a high average recognition rate. Preliminary results show that wavelet compression of normal image could greatly reduce the size of object descriptor (the normal map) not significantly affecting the recognition precision. As the normal image is a 2D mapping of mesh features, ongoing research will integrate additional 2D color info (texture) captured during the same enrollment session. Implementing a true multi-modal version of the basic algorithm which correlates the texture and normal image could further enhance the discriminating power even for complex 3D recognition issues. We propose a quite fast recognition algorithm that could be used in different industrial application, expecially in glaze automation. Future work suggest to make distributed implementation of this algorithm to improve real time responses.

References

1. Bowyer, K., Chang, K., Flynn, P.: A survey of approaches to three-dimensional face recognition. In: Proc. of 17th International Conference on Pattern Recognition (ICPR 2004), August 2004, vol. 1, pp. 358–361 (2004)
2. Gordon, G.: Face Recognition from Depth Maps and Surface Curvature. In: Proc. of SPIE, Geometric Methods in Computer Vision, vol. 1570, pp. 1–12 (July 1991)
3. Achermann, B., Jiang, X., Bunke, H.: Face recognition using range images. In: Proc. of International Conference on Virtual Systems and MultiMedia (VSMM 1997), Geneva, Switzerland, September 1997, pp. 129–136 (1997)

4. Achermann, B., Bunke, H.: Classifying range images of human faces with the hausdorff distance. In: Proc. of 15th International Conference on Pattern Recognition (ICPR 2000), Barcelona, Spain, vol. 2, pp. 813–817 (2000)
5. Tanaka, H.T., Ikeda, M., Chiaki, H.: Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition. In: Proc. of 3rd International Conference on Automated Face and Gesture Recognition (FG 1998), Nara, Japan, April 1998, pp. 372–377 (1998)
6. Hesher, C., Srivastava, A., Erlebacher, G.: A novel technique for face recognition using range images. In: Proc. of Seventh Int'l Symp. on Signal Processing and Its Applications (ISSPA 2003), Paris, France (July 2003)
7. Chang, K., Bowyer, K., Flynn, P.: Face recognition using 2D and 3D facial data. In: Proc. of Multimodal User Authentication Workshop, Santa Barbara, CA, December 2003, pp. 25–32 (2003)
8. Guo, K., Yang, X., Zhang, R., Zhai, G., Yu, S.: Face super-resolution using 8-connected Markov Random Fields with embedded prior. In: 19th International Conference on Pattern Recognition, ICPR 2008 (2008)
9. Medioni, G., Waupotitsch, R.: Face recognition and modeling in 3D. In: Proc. of IEEE Int'l Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003), Nice, France, October 2003, pp. 232–233 (2003)
10. Tangelder, J.W.H., Veltkamp, R.C.: A survey of content based 3D shape retrieval methods. Journal of Multimedia Tools and Applications 39(3), 441–471 (2008)
11. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Expression-invariant 3D face recognition. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 62–69. Springer, Heidelberg (2003)
12. Beumier, C., Acheroy, M.: Face verification from 3D and grey level cues. Pattern Recognition Letters 22(12), 1321–1329 (2001)
13. Wang, Y., Chua, C., Ho, Y.: Facial feature detection and face recognition from 2D and 3D images. Pattern Recognition Letters 23(10), 1191–1202 (2002)
14. Colombo, A., Cusano, C., Schettini, R.: 3D face detection using curvature analysis. Journal of Pattern Recognition 2006 (3), 444–445 (2006)
15. Gu, X., Gortler, S., Hoppe, H.: Geometry images. In: Proc. of SIGGRAPH 2002, July 2002, pp. 355–361. ACM, New York (2002)

Emotional State Inference Using Face Related Features

Marco Anisetti and Valerio Bellandi

Abstract Obtaining reliable and complete systems able to extract human emotional status from streaming videos is of paramount importance to Human Machine Interaction (HMI) applications. Side views, unnatural postures and context are challenges. This paper presents a semi-supervised fuzzy emotional classification system based on Russell's circumplex model. This emotional inference system relies only on face related features codified with the Facial Action Coding System (FACS). These features are provided by a morphable 3D tracking system robust to posture, occlusion and illumination changes.

1 Introduction

Streaming videos or images containing faces are essential to advanced vision-based HMI [1]. The ability to detect, track and recognize facial motion is useful in applications such as people identification, expression analysis and surveillance. Inferring an emotional meaning from facial expression is still a challenging task due to i) high variability of the context and its emotional evaluation by the subject (as a psychological issue) and ii) from changes in pose, facial expression deformations and illumination modifications. From a psychological point of view, the context in which a particular facial expression is displayed is crucial. Several works have used facial features alone to ascribe emotional meaning to facial expressions (e.g. smile means happiness), without any or enough consideration of the context.

To solve this context dependency we propose a context aware system based on expert evaluation of emotional reactions inside a specific environment. In particular, this work is focused on HMI context exploitation and refining our previous

Università di Milano
Dipartimento di Tecnologie dell'Informazione
Via Bramante 65, 26013, Crema (CR), Italy
e-mail: {anisetti,bellandi}@dti.unimi.it

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 401–411.
springerlink.com

method [2][3] for obtaining precise 3D face posture from a monocular video sequence and FACS (proposed by Ekman and Friesen [4]) coding-related expression parameters [5].

Our aim is to extract information regarding the emotional experience or the emotional information that a subject conveys about the context, in terms of hedonic value, activation, and control of Russell circumplex space [6] [7] using only face related features.

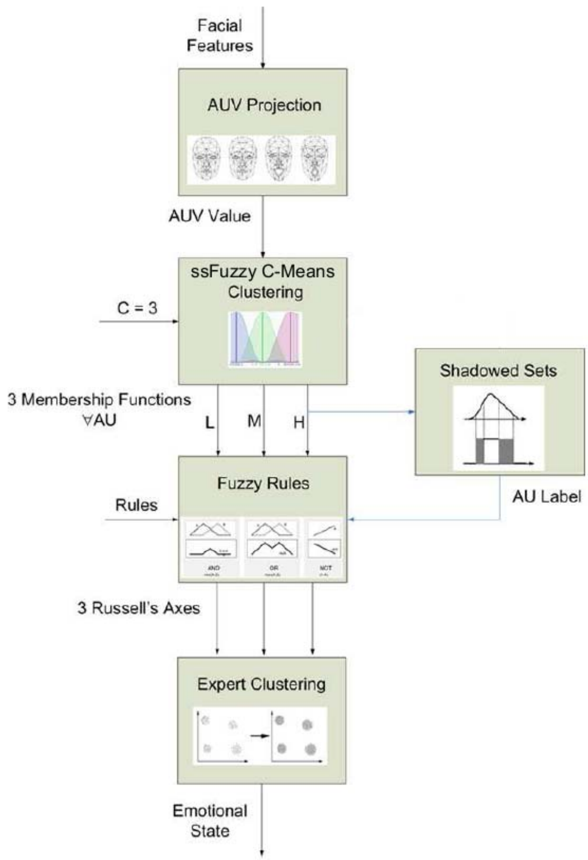


Fig. 1: Emotional classifier architecture: AU classification and intensity extraction together with Russell's space mapping.

Figure 1 shows our architecture for an emotion inference engine including: i) Action Unit (AU) classification and intensity extraction system and, ii) Russell's space mapping system. The contribution of this architecture is threefold:

- *Native FACS classification with intensity*. To classify our expression related features we select a FACS coding system so facial muscular movements need to be classified as Action Units (AUs). Thanks to our tracking approach, every muscular movement is associated with an intensity so that each AU can be associated with a continuous intensity value. Every AU has a different range of motion intensity. A fuzzification approach is used to define three intensity classes: Low, Medium, and High for each AU so that different muscular movements can be compared in terms of a rule based system.
- *Emotion representation in Russell's circumplex model*. The circumplex model allows identification of regions in 3D space not only related to the prototypic emotions (fear, joy, sadness and so forth) but also to other physical states (tense, calm, serene, comfortable, fatigued, bored, drowsy). Our tracking system is able to identify a point inside this 3D space by mapping only muscle and head movement over each axis using a fuzzy rule system. Expert-based clustering on this space is used to obtain the emotional classification.
- *Context adaptability*. The system can be simply adapted to every context scenario via ad hoc expert clustering for each scenario. The AU related system is still valid for every context and AU mapping over Russell's space axis.

To evaluate the quality of the AU classification system we use the Cohn-Kanade database [8], and the MMI Database [9], for Russell's space mapping system. Regarding emotion recognition, we perform our test with the Hammal-Caplier database [10] for simple emotion evaluation and our own interaction database for Human-Machine interaction preliminary tests.

2 Related work: Emotion recognition

Automatic emotion recognition is a challenge. Ekman's description of the six prototypical facial expressions of emotion is linguistic and thus ambiguous. A general agreement about the emotional meaning of facial feature configurations has not been reached. Recent theories [11], [12] suggest that emotions are complex processes, mediated by the presence of many hidden variables such as evaluation of the context which depends on an individual's hopes, plans, desires and so forth. The previous work is divided into holistic vs. analytic and static vs. dynamic approaches.

- **holistic vs. analytic.** Holistic methods attempt to recognize a small set of prototypical emotional expressions starting from the whole face [13] [14]. Alternatively, analytic methods would decompose the face into a set of different features and classify these features to find the expressed emotion [15]. The shortcoming of almost all automated facial expression analysis research to date is that it is

limited to deliberate facial expressions recorded under controlled conditions thus omitting significant head motion and other factors which would complicate the analysis.
- **static vs. dynamic.** The classifiers used so far can either be static or dynamic. Static classifiers use feature vectors related to a single frame in order to perform classification (e.g., neural networks, Bayesian networks, linear discriminant analysis). Temporal classifiers try to capture the temporal pattern in the sequence of feature vectors related to each frame, such as the Hidden Markov Model (HMM) based methods [16] [17]. "Static" classifiers are easier to train and implement, but when used on a continuous video sequence, they can be unreliable, especially for frames that are not at the peak of the expression. From these studies it is clear that the study of emotional expression dynamics is very helpful when dealing with spontaneous emotions because they are more subtle than the prototypical ones.

3 Novel approach: Circumplex model

The psychological foundation of our work follows Grammer's approach [18] where AU configurations have been found to correlate with the values shown in the Russell's model coordinates. Emotion can be communicated by several non-verbal behavioral systems. In particular, we deal with face expressions, eyes and head movements. The face is a complex system whose function is signaling and regulating interactions between individuals and their social/physical environment. Because of its complexity, two distinct levels of analysis [19] need to be taken into account:

- *Molecular level*: The molecular level concerns minimal and distinct movements of the large number of facial muscles, which allow high mobility and expressiveness of the face in order to define a kind of alphabet for objective coding of face movements
- *Molar level*: The molar level concerns final configuration of facial movements.

In order to achieve the first level of analysis we use FACS, a coding system of all the visible facial movements and their anatomic-physiological components. We focus our attention on the molar level as a promising approach, referring to Russell's circumplex model where each emotional state can be described as a point in a three dimensional space defined with three axes:

- *Pleasure-Displeasure*: This axis represents the hedonic dimension. The range of variability starts from one extreme (agony) through a neutral point (adaptation level) to its opposite extreme (ecstasy). Several studies on newborn babies [20] have shown that there are some typical facial configurations which correlate with hedonic reactions to positive and negative stimuli. These reactions are independent of social context and from any cognitive appraisal of the stimulus. We separate each facial configuration in terms of pleasure and displeasure signaling AUs as shown in Table 1.

	Pleasure	**Displeasure**
Upper Face	AU6, AU1+2	AU1, AU4, AU1+4
Middle Face	Null	AU9, AU10, AU11
Lower Face	AU12	AU15, AU17, AU20

Table 1: Facial muscular movement mapping for Hedonic axis.

- *Arousal*: This axis represents the activation dimension. Range of variability starts from sleep, then drowsiness, through various stages of alertness to frenetic excitement defined by [21] and [7]. Generally this measure is correlated with physiological activation. Many works suggest that the activation dimension can be correlated with the amplitude of the various facial units, especially the jaw drop and the eyes opening.
- *Coping*: According to [22] we can infer the coping/control dimension from actions like avoidance (low control)/approach (high control). We evaluate these actions from head and eye movements (Table 2).

	Avoidance → low control	**Approach → high control**
Eyes	Eye distraction	Gaze toward stimulus
Head	Head down, moving back/shrinking	Head forward
Gestures	Self-contact	Hands, arms toward stimulus

Table 2: Movement mapping for Coping axis.

We are able to define Russell's space using only the tracking feature of our algorithm. In the following sections, we describe in detail our emotion inference engine, starting from AU classification and an intensity extraction system.

4 Russell's axes classification

In this section we summarize the process of axes definition for creating an emotional inference space. In order to perform an emotional classification using Russell's approach, we need to map AU to axes values. This mapping can exploit some well defined rules. In this section we describe: i) the pre-processing step for AU clustering, ii) the shadowing step for AU classification, and iii) the rules definition for mapping into axes.

4.1 AU classification

First of all, in order to outline the rules needed for each axis definition, we construct a classifier in terms of range for every AU signal. Even though AU coding is only a label coding system, we use the intensity of each AU to create an axes-oriented emotion classification system. We adopt a classification for every AU into Low, Medium and High intensity. This classification is fundamental for Russell axes creation through a set of fuzzy rules. By considering a fuzzy approach, we can design 3 membership functions manually for Low, Medium and High classification, and use the same membership functions for every feature signal. The classification of every AU is different since each AU is related to a different muscular movement with a specific intrinsic characteristic. For instance, small eyebrow movements are as important as bigger ones detectable on the jaw. This consideration coupled with the one on non linearity underlined earlier makes the unique manual definition of membership functions not efficient enough for our classification purposes. An expert evaluation is required. On the other hand an a priori definition of membership functions for every AU is quite difficult even for an expert. Furthermore, a fully supervised cluster requires analyzing numerous videos. For that reason we decide to use a semi-supervised approach, in particular ssFCM (semi-supervised Fuzzy C-Means). Using this approach, a subset of data $X = [X^l|X^u]$ are labeled (X^l) by an expert and then used in ssFCM coupled with unlabeled data (X^u). U is the $c \times n$ matrix of membership functions for every class (row) and every set of data X (column) where c c is the number of classes and n is the cardinality of X. The membership matrix U includes the labeled part U^l as the unlabeled U^u. . The difference between the normal FCM and the ssFCM obviously resides in the use of this labeled data. In particular, the initialization of the center of the cluster is defined using only the labeled data. The rest of the FCM algorithm remains the same, involving both labeled and unlabeled data and updating the membership values for both cases. In general, when the labeled data is the smallest part of the entire data set, this approach is reinforced by using a weight factor for every labeled element. These weights, introduced by [23] allow us to tailor the computation to agree with any expert knowledge about the importance of each set of data as a training example. In particular we use the cluster parameter $c = 3$ (Low, Medium and High) and fuzzification coefficient $m = 2$. We perform this membership function creation for each feature using all signals extracted from the Cohn-Kanade database. With this training we obtain 3 membership functions for every AU signal. These membership functions are generic for each AU. It is easy to obtain a personalized version of membership functions for every subject, to better deal with different ways of expressing a certain AU. With our tests, we prove that this generic AU membership definition is sufficient for good performance. After this classification we need to obtain an AU label classification to perform a comparison with our AU-labeled database. We also use this label to define the axes rules. In order to obtain this label, we define a shadow set using the alpha cut approach [24]. Using our function $A(x)$, we need to discover the alpha cut parameters α.

$$V = \left| \int_{-\inf}^{a_1} A(x)dx + \int_{a_2}^{+\inf} (1 - A(x))dx - \int_{a_1}^{a_2} A(x)dx \right| \quad (1)$$

The thresholds for α should lead to $V(\alpha) = 0$. Following this minimization, we obtain the best α cut for our shadowed set. This approach classifies the three states into presence, no presence, and uncertain. Using this classification, we could perform some comparisons with our labeled database. After this process we obtain an AU classifier that is the basis for our emotional inference system.

4.2 Axes creation

In this section we describe the definition of each axis for emotion classification using a circumplex model. We assert that the previous classification of AU values becomes very useful for defining the axis crosswise rules-based system. In this section, for the sake of conciseness, we present the rules used for the positive part of hedonic axis. As already described, the hedonic axis represents the pleasure or displeasure emotional valence. For the pleasure/positive part of the hedonic axes (Table 1), we analyze the value of several AUs (AU1+AU2, AU12, AU6)[1]. We use a shadowed set for AUs that are simultaneously present. We call an AU that is shadowed AUS.

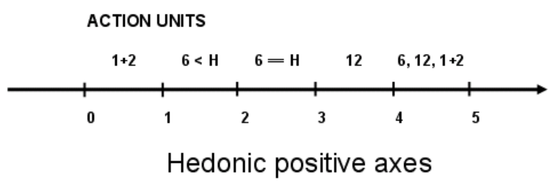


Fig. 2: Hedonic axis: Positive part and AU mapping. H means High AU classification.

The mapping of the positive part of the hedonic axis follows a well defined set of rules that splits this part of the hedonic axis into five segments (Figure 2). We present the set rules for these segments S in equation 2.

$$\begin{aligned}
S_1 &= (AUS1 \ \wedge \ AUS2) \ \wedge \ ((AU1 \ \textit{is High}) \ \wedge \ (AU2 \ \textit{is High})) \\
S_2 &= (AU6 \ \textit{is not High}) \\
S_3 &= AUS6 \ \wedge \ (AU6 \ \textit{is High}) \\
S_4 &= (AU12 \ \textit{is not High}) \\
S_5 &= (AUS6 \ \wedge \ AUS12) \ \wedge \ (AU12 \ \textit{is High}) \ \wedge \ (AU6 \ \textit{is High})
\end{aligned} \quad (2)$$

[1] AU1+AU2 means that AU1 and AU2 need to be contemporary valid. For this reason we use boolean logic to discover when AU1 and AU2 are simultaneously active.

We define these rules for segment classification when we need to map this onto a hedonic axis. Segment rules for axes are as defined in Figure 2 (S_1 After this mapping, it becomes obvious that the value of each segment defines the offset of each axis segment ($[0\cdots1]$). Therefore the values of axes $E_+ \in [0\ldots5]$ are:

$$
\begin{aligned}
Segn = \; & \begin{array}{l} 0 = (S_1 > 0) \wedge (S_2 < 0) \wedge (S_3 < 0) \wedge (S_4 < 0) \wedge (S_5 < 0) \\ 1 = (S_2 > 0) \wedge (S_3 < 0) \wedge (S_4 < 0) \wedge (S_5 < 0) \\ 2 = (S_3 > 0) \wedge (S_4 < 0) \wedge (S_5 < 0) \\ 3 = (S_4 > 0) \wedge (S_5 < 0) \\ 4 = (S_5 > 0) \end{array} \\
E_+ = \; & Segn + S_{Segn+1}
\end{aligned} \tag{3}
$$

The approach is the same for the negative part of the hedonic axes: i) segmentation into 5 parts, ii) one rule for every part involving AUs for displeasure described in Table 1,and iii) mapping.

The approach is different for the arousal axis. Since the arousal axis represents the activation dimension, the range of variability starts from sleep, then drowsiness, through various stages of alertness to frenetic excitement. This measure is correlated with physiological activation. To describe this activation, we evaluate the facial muscles with special emphasis on eyes lids to create a scale between sleep and alertness. The arousal axis is obviously divided into positive and negative parts, around zero we have neutral status. Different from the hedonic axes, we do not use any rules system since we need only the intensity of all AU movement. To differentiate the positive and negative axes we use the eye lid aperture [25] ((particularly: i) positive from relaxed to wide open, and ii) negative from relaxed to closed) and facial muscle movements. We consider positive any kind of facial muscular movement that produces a significant change in facial expression. In fact we consider positive any kind of facial muscular movement that produce a significative changing in facial expression respect to neutral ones. For negative ones we consider any kind of muscular movement that tender to a total relaxed face. For negative expressions, we focus our attention on eyes lids (or rather the intensity of the AU related to eye lids) since the change in expression does not differ from the positive expression.

Regarding the coping axis, we can infer the coping dimension from avoidance/approach. Using Table 2 , we can use the intensity of AU related to head movement and iris movement to infer the value along the coping axes. This axis differs from others in that coping assumes only two values: positive or negative. Using these three axes and Russell-based classification, we are able to classify the emotional state of the subject. In Section 5 we describe our classifier for emotion.

5 Emotional classification

Emotional classification using Russell's space is an expert base clustering problem. Considering the definition of our 3 facial feature-based axes, the problem can be simplified into a two-dimensional Russell's space, one for positive coping and the other for negative. The main advantage of using the Russell space is that it is well clustered into the region with the relative emotional label. If the facial features are well tracked and the axis is well defined, the emotional recognition can be reduced to expert based clustering in a psychological well known 2D space. The first assumption is largely proved in our previous work, regarding the definition of the Russell's facial based axes. The psychological theory confirms our definition, and our preliminary experiment reinforces this confirmation. The complexity of the classification system depends upon the dimension of the cluster or rather the number of emotions that we want to identify. Depending on the number of emotional states, we can simplify our system into positive vs negative reactions to a certain stimulus. In some contexts, this simplification is enough, as in HMI. For further testing of the applicability of our emotional inference system, we apply ssFCM for clustering into the Russell's space of emotion labeled the Hammal-Caplier database. Figure 3 shows a ssFCM classification of the three emotional status of the Hammal-Caplier database (15 subjects). The joy space is well defined but not the disgust one. Two samples that belong to disgust are classified as surprise (evidenced in the figure with the circle and square at the same time). These two samples are wrongly classified mainly because tracking does not determine the subject's correct AU.

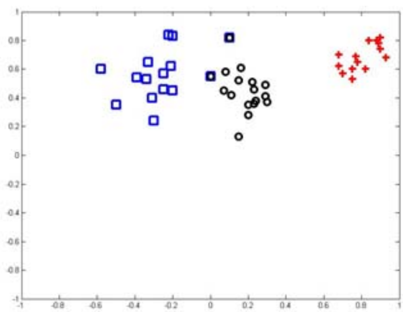


Fig. 3: Emotional classification with the three class: *i)* joy (red +), *ii)* surprise (black circle), *iii)* disgust (blue square) of Hammal-Caplier database with emotional axis normalized from -1 to 1.

Regarding the HMI scenario, a simple positive vs. negative reaction classification is enough to infer the emotional situation generated by the interaction.

6 Conclusion and future work

Concluding, this paper presents an semi supervised fuzzy emotional classification system based over Russell's space theory. This emotional inference system relies only on face related features. This features are provided through our tracking system. The system guarantee context awareness by simple expert tuning. More experimentation need to be carried out to obtain a more efficient tuning of emotional classification in more complex scenario. Some interesting future works is represented by analysis of temporal plot into Russell space. Figure 4 shows a neutral to joy back to neutral plot over Russell's space in Hammal-Caplier database.

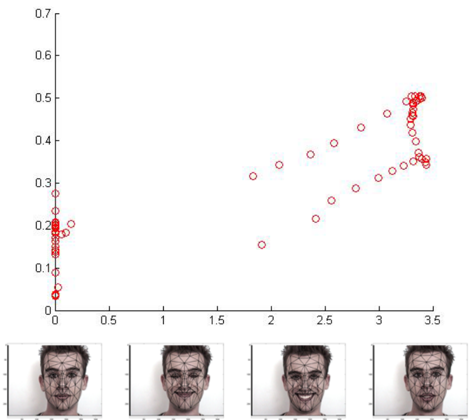


Fig. 4: Expert clustering: experimental plot in circumplex Hedonic-Arousal space in continuous time (classifying each frame) using Hammal-Caplier database.

References

1. Andreoni, G., Anisetti, M., Apolloni, B., Bellandi, V., Balzarotti, S., Beverina, F., Campadelli, P., Ciceri, M.R., Colombo, P., Fumagalli, F., Palmas, G., Piccini, L.: Emotional interfaces with ambient intelligence. In: Vasilakos, A., Pedrycz, W. (eds.) Ambient Intelligence, Wireless Networking and Ubiquitous Computing, ch. 7. Artech House, USA (2006)
2. Anisetti, M., Bellandi, V., Damiani, E., Beverina, F.: 3d espressive face model-based tracking algorithm. In: Signal Processing, Pattern Recognition, And Applications (SPPRA), Innsbruck (2006)
3. Anisetti, M., Bellandi, V., Beverina, F.: Face tracking algorithm robust to pose, illumination and face expression changes: a 3d parametric model approach. In: International Conference on Computer Vision Theory and Applications, Setubal, Portugal (2006)
4. Ekman, P., Friesen, W.: Facial action coding system: A technique for the measurement of facial movement. Consulting Psychologists Press (1978)
5. Damiani, E., Anisetti, M., Bellandi, V., Beverina, F., Ciceri, M.R., Balzarotti, S.: Psychology-aware video-enabled workplace. In: 3rd International Conference on Ubiquitous Intelligence and Computing, Whuan and Three Georges, China (2006)

6. Russell, J., Mehrabian, A.: Evidence for a three-factor theory of emotions, vol. 11, pp. 273–294 (1977)
7. Russell, J.: Core affect and the psychological construction of emotion. Psychological Review 110, 145–172 (2003)
8. Kanade, T., Cohn, J., Tian, Y.: Comprehensive database for facial expression analysis. In: Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000), Grenoble, France, pp. 46–53 (2000)
9. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-based database for facial expression analysis. In: Proc. IEEE Int'l Conf. Multimedia and Expo. (ICME 2005), Amsterdam, The Netherlands (2005)
10. Hammal, Z., Caplier, A., Rombaut, M.: A fusion process based on belief theory for classification of facial basic emotions. In: Proc. Fusion 2005 the 8th International Conference on Information fusion (ISIF 2005) (2005)
11. Scherer, K.: "J". In: Strongman, K. (ed.) International reviews of studies of emotion, vol. 2 (1992)
12. Frijda, N.: The place of appraisal in emotion. In: Cognition and Emotion, pp. 357–387 (1993)
13. Essa, I., Darrell, T., Pentland, A.: Tracking facial motion (1994)
14. Abboud, B., Davoine, F., Dang, M.: Expressive face recognition and synthesis (2003)
15. Mase, K.: Recognition of facial expression from optical flow. IEICE Transaction E74, 3474–3483 (1991)
16. Lien, J.: Automatic recognition of facial expressions using hidden markov models and estimation of expression. Ph.D. dissertation, Carnegie Mellon University (1998)
17. li Tian, Y., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), 97–115 (2001)
18. Grammer, K., Tessarek, A., Hofer, G.: From emoticons to avatars: simulation of facial expression. In: Proc. European Science Foundation Workshop on emotions and new media
19. Ricci, G.P., Bitti, P.E., Caterina, R.: Different behavioural markers in different smiles. In: Proceedings of the IXth Conferences of the International Society for Research on emotion, Isre, Toronto (1996)
20. Oster, D.H., Nagel, L.: Adult judgment and finegrained analysis of inant facil expressions: Testing the validity of a priori coding formulas. In: Developmental Psychology (1992)
21. Knez, Hygge: The circumplex structure of affect: A swedish version. Scandinavian Journal of Psychology 42(5) (December 2001)
22. Davidson, R.J.: Anterior cerebral asymmetry and the nature of emotion. Brain and Cognition (1992)
23. Bensaid, A.M., Hall, I.L.O., Bezdek, J.C., Clarke, L.P.: Partially supervised custering for image segmentation. Pattern Recognition 29(5), 859–871 (1996)
24. Pedrycz, W.: Shadowed sets: representing and processing fuzzy sets. IEEE Transactions on Systems, Man and Cybernetics, Part B 28, 103–109
25. Bellandi, V., Anisetti, M., Beverina, F.: Upper-face expression features extraction system for video sequences. In: Proc. of International Conference on Visualization, Imaging, and Image Processing (VIIP 2005), September 2005, pp. 83–88 (2005)

UBIAS Systems supporting interpretation of radiological images

Lidia Ogiela, Ryszard Tadeusiewicz, and Marek R. Ogiela

Abstract This publication presents a new class of cognitive systems designed for analysing visual data. Such systems have been created as the next step in the development of classical intelligent systems DSS (*Decision Support Systems*), which are currently the most widespread tools providing computer support for medical data. This publication defines and presents a new category of cognitive categorization systems. Our paper presents a new ideas and possibilities of using mathematical linguistic formalisms, in the form of pattern analysis and image understanding, in the development of a new class of intelligent medical diagnosis-support and image recognition systems i.e. Cognitive Vision Systems. Such ideas are aimed at facilitating the cognitive and semantic analysis of the meaning of some selected medical patterns based on the presence of cognitive resonance. When trying to use this cognitive phenomenon and methods to the automatic analysis, interpretation and understanding of image-type data, we propose a new class of cognitive categorization systems: UBIAS (Understanding Based Image Analysis Systems). The procedures proposed are based on the cognitive resonance and categorization model. The application and ideas presented in this paper will show how great the opportunities are for automatically detecting lesions in the analysed structures for eg. in the foot bone structure.

Lidia Ogiela
Faculty of Management of AGH University of Science and Technology,
Al. Mickiewicza 30, PL-30-059 Krakow, Poland
e-mail: logiela@agh.edu.pl

Ryszard Tadeusiewicz, Marek R. Ogiela
Institute of Automatics of AGH University of Science and Technology,
Al. Mickiewicza 30, PL-30-059 Krakow, Poland
e-mail: {rtad,,mogiela}@agh.edu.pl

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 413–419.
springerlink.com

1 Introduction

The human mind processes taking place in the human brain have been the subject of theoretical and practical analyses and empirical research for many years. However, although we increasingly frequently feel we can explain almost everything that takes place in the human brain, a closer study always reveals something to undermine this certainty we have. Among all the processes taking place there, key ones include thinking, cognitive, interpretation, analysis and reasoning processes as well as processes for the varied, complex and indepth analysis of information received from the senses or recalled from memory. These very types of human cognitive processes were used by the authors of this publication to attempt to combine human information analysis processes with automatic, computer data analysis. The process of data analysis and interpretation that has been continuously developing for a number of years is now moving to jobs consisting in not just simply interpreting the analysed data, but mainly focusing on reasoning about and understanding that data [1, 4, 9]. And it is for those activities aimed at understanding the analysed data that a special class of intelligent information systems has been developed, which are aimed not just at the simple analysis, but mainly strive to reveal the semantic information carried by that data, try to understand that data and to reason based on the semantic information it contains. Such a process was possible thanks to using formalisms of linguistic perception and understanding combined with the purely human process of interpreting, analysing, understanding and reasoning which occurs in the human mind.
This process is based on the cognitive resonance phenomenon which occurs during the analysis process, and which becomes the starting point for the process of data understanding which consists not just of the simple data analysis, but is based on its semantic meaning, which making it possible to reason (Fig. 1).

Fig. 1: Cognitive resonance phenomenon and the process of analysing, reasoning, interpretation and understanding of data.

The essence and ideas of cognitive resonance is indicating the similarities that appear in the analysed dataset and in the generated set of expectations as to the possible results of the knowledge acquired by the system [2]. These similarities are revealed during the comparative analysis conducted by the system, in the course of which the analysed data is subjected to the phenomenon of understanding. The reasoning process which is the result of the understanding process is an indispensable factor for the correct analysis of data, because if it did not occur, forecasting and reasoning as to the future of the phenomenon being studied would be impossible.

So conducting the analysis without the reasoning process could lead to impoverishing the entire analysis process, as it would be limited only to understanding the reasons for the occurrence of the analysed phenomenon, but without a chance of determining its further development.
Among the great variety of information systems, the class of UBIAS systems (Understanding Based Image Analysis Systems) is very popular like DSS systems (Diagnosis Support Systems) due to their wide application in clinical practice. In this paper we try to show an example of a system that was prepared not only for the simple diagnostics, but was also oriented towards cognitive interpretation leading to extracting the features, and understanding pathological diseases.

2 Semantic analysis and interpretation of medical patterns

Cognitive analysis used in IT systems is very often based on the syntactic approach and syntax analysis [7, 10]. For the purpose of interpreting and description the image meaning, it first uses a pre-processing operation usually composed of image coding using terminal symbols, shape approximation, as well as some kind of filtration or segmentation [7, 9]. As a result of executing these stages it is possible to obtain a new image representation in the form of hierarchic semantic tree structures and subsequent production steps of this representation from the initial grammar symbol [9].
An cognitive categorization system distinguishing image data at the stage of preprocessing must, in the majority of cases, perform image segmentation, identify primitive components and determine spatial as well as semantic relations between them. An appropriate classification is based on the recognition of whether a given representation of the actual image belongs to a class of images generated by languages defined by one of the possible number of grammars. Such grammars can be considered to belong to sequential, tree and graph grammars while recognition with their application is made in the course of a syntactic analysis performed by the automatic system analysis [10].
The most important difference between all traditional methods of automatic image processing and the new automatic understanding paradigm is that there is one directional flow of information in the traditional methods while in the new paradigm there are two-directional interactions between signals (represents features) extracted from the image analysis and expectations resulting from the knowledge of image content.
A cognitive method of interpreting disease units and pathological lesions forms the main element of a correctly functioning IT system supporting medical image diagnostics. Further down, such an interpretation of changes occurring in foot bone pathologies will be presented.

3 Intelligent information UBIAS systems for analysing foot bone images

An UBIAS system has been developed for analysing and interpreting foot bone images. The cognitive analysis of images showing foot bones was conducted using graph grammars for the recognition and intelligent understanding of X-ray images of foot bones under analysis. In order to perform a cognitive analysis aimed at understanding the analysed data showing foot bone lesions, a linguistic formalism was proposed in the form of an image grammar, the purpose of which is to introduce the appropriate definition of a language describing the possible cases of perceivable physiological norms and the possible lesions of foot bones. The proposed solution has been used for a special type of projections of foot bone images - the dorsoplanar projection, one of the possible projections used in foot diagnostics. Based on the analysis of foot bones in the presented dorsoplanar projection of the X-ray image, a graph was defined for describing the foot bone skeleton, which used the available descriptions and names for foot bones for the dorsoplanar projection. For the introduced graph describing the foot bone skeleton in the dorsoplanar projection, a graph of the appropriate relations was introduced which described the lesions occurring within the foot (Fig. 2).

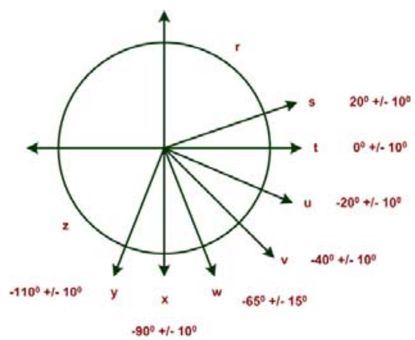


Fig. 2: A relation graph defining the spatial topology between foot bones.

The introduction of the above relation graph (Fig. 2) and the determination of a graph spanned on the skeleton of foot bones were used to define a graph in which all the adjacent foot bones were labelled as appropriate for the analysed dorsoplanar projection (Fig. 3).

The introduction of the above definitions was aimed at presenting the interrelations which may occur between particular elements of foot bone structure represented as a graph. Taking into account those relations and the method of labelling the apexes of the graph spanning the bones, we can create a linguistic description of the semantics of the lesions looked for and of the recognition of pathological situations. For the purposes of the analysis conducted, a formal definition of the

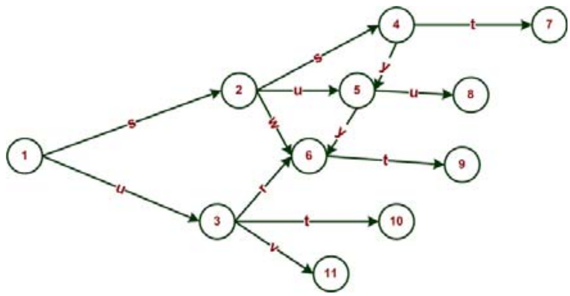


Fig. 3: A graph with numbers of adjacent bones marked based on the relation graph.

graph grammar was introduced, which takes into account the developed linguistic description of correct connections between foot bones. A set of non-terminal labels of apexes:
N = {ST, CALCANEUS, OS NAVICULARE, OS CUBOIDEUM, OS CUNEIFORME MEDIALE, OS CUNEIFORME INTERMEDIUM, OS CUNEIFORME LATERALE, M1, M2, M3, M4, M5}
A set of terminal labels of apexes:
$\sum$ = {c, on, oc, ocm, oci, ocl, m1, m2, m3, m4, m5},
Γ = {s, t, u, v, w, x, y} the set of edge labels shown in Fig. 2,
S start symbol = ST,
P - is a finite set of productions shown in Fig. 4.
Inputting the above grammar into UBIAS systems executing in-depth analyses of image-type data showing foot bone lesions demonstrated what the structure of a healthy foot should be like, what the foot bone structure should be like, and made it possible to define a set describing the correct structure of foot bones for the dorso-planar projection presented in this paper (Fig. 4).

The conducted analysis of the understanding of image-type data, aimed at an in-depth understanding of the images analysed (in this case specific lesions), goes towards a semantic meaning reasoning about the images being interpreted.
Figure 5 shows the possibilities describing cases of various pathologies - for e.g. the appearance of an additional instep bone and bone displacements. These enrich the set of derivation rules which define the proposed linguistic formalism used for meaning interpretation jobs.

The types of foot bone lesions shown above have been presented using a selected type of projection for foot bone imaging. Obviously, similar solutions can be proposed for the remaining projection types, i.e. the lateral projection (external and internal).

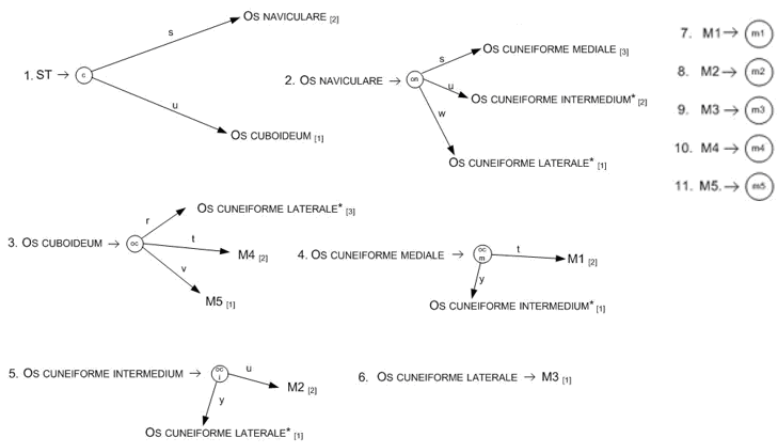


Fig. 4: A finite set of productions showing the healthy structure of foot bones.

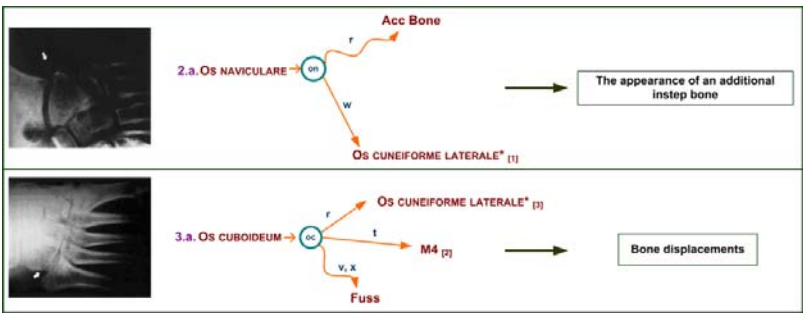


Fig. 5: The automatic understanding of foot bone lesions detected by an UBIAS system in a dorsoplanar projection for selected foot lesions.

4 Conclusions

The presented cognitive data analysis systems developed to analyse medical (visual) data can perform in-depth analyses, interpretations and reasoning with regard to various medical pathologies data, in particular different pathologies at different images. The research conducted by the author, based on the analysis of images with lesions, has demonstrated that cognitive data analysis can represent the factor that significantly enriches the capabilities of contemporary medical information systems. In particular, the described research has demonstrated that an appropriately built picture grammar makes it possible to conduct a precise analysis and description of medical images from which important semantic information can be extracted about the nature of pathological processes and lesions. It is worth noting that the results described in this paper have been obtained following the cognitive process, simulating an experts' method of thinking: if one observes a deformation of the organ shown by the medical image used, then one tries to understand the pathological process that was the reason for the appearance of deformations found.

Acknowledgments

This work has been supported by the Ministry of Science and Higher Education, Republic of Poland, under project number N519 007 32/0978

References

1. Albus, J.S., Meystel, A.M.: Engineering of Mind - An Introduction to the Science of Intelligent Systems. John Wiley & Sons Inc., Chichester (2001)
2. Branquinho, J. (ed.): The Foundations of Cognitive Science. Clarendon Press, Oxford (2002)
3. Burgener, F.A., Meyers, S.P., Tan, R.K., Zaunbauer, W.: Differential Diagnosis in Magnetic Resonance Imaging. Thieme (2002)
4. Davis, L.S. (ed.): Foundations of Image Understanding. Kluwer Academic Publishers, Dordrecht (2001)
5. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley & Sons, Inc., Chichester (2001)
6. Khan, M.G.: Heart Disease Diagnosis and Therapy. Williams & Wilkins, Baltimore (1996)
7. Meyer-Baese, A.: Pattern Recognition in Medical Imaging. Elsevier, Amsterdam (2003)
8. Ogiela, L.: UBIAS systems for cognitive interpretation and analysis of medical images. Opto-Electronic Review 17(2), 166–179 (2009)
9. Ogiela, L., Tadeusiewicz, R., Ogiela, M.R.: Cognitive Analysis In Diagnostic DSS-Type IT Systems. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J.M. (eds.) ICAISC 2006. LNCS (LNAI), vol. 4029, pp. 962–971. Springer, Heidelberg (2006)
10. Ogiela, L., Tadeusiewicz, R., Ogiela, M.R.: Cognitive Approach to Visual Data Interpreta-tion in Medical Information and Recognition Systems. In: Zheng, N., Jiang, X., Lan, X. (eds.) IWICPAS 2006. LNCS, vol. 4153, pp. 244–250. Springer, Heidelberg (2006)
11. Tadeusiewicz, R., Ogiela, M.R.: Medical Image Understanding Technology. Springer, Heidelberg (2004)

Multimedia Elements for Medical e-Learning Improvement

Liana Stanescu, Dan Burdescu, Gabriel Mihai, Marius Brezovan,
and Cosmin Stoica Spahiu

Abstract The paper presents a software tool based on topic maps dedicated to medical e-learning. The application allows the graphical visualization of the MeSH thesaurus (medical terms from Diseases and Drugs categories) with a topic map. With this graphical modality, the learner can view medical term description and its associative relationships with other medical descriptors. The paper presents also how to use the topic map for semantic querying of a multimedia database with medical images that are accompanied by diagnosis and treatment as crucial information. For retrieving the interest information this access path can be combined with another modern solution: the content-based visual query on the multimedia medical database using color and texture features automatically extracted. The paper presents the original algorithm for building and populating the topic map starting from MeSH thesaurus, mapping an xml file that can be downloaded free, to an xtm file that contains the topic map.

1 Introduction

The past few years have seen rapid advances in communication and information technology, and the presence of the worldwide web into everyday life has important implications for education [9]. The discipline of medicine has not been immune to these effects. The introduction of the multimedia components for the improvement of the learning content led to the necessity of the concept which precedes even the Internet, i.e. the multimedia learning. Both teachers and students consider that the multimedia learning improves the process of teaching and learning [9]. In

University of Craiova,
Faculty of Automation, Computers and Electronics,
Craiova, Romania
e-mail: stanescu_liana,burdescu_dumitru,brezovan_marius@software.ucv.ro

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 421–430.
springerlink.com

the medical domain, for establishing a correct diagnosis in a fast way, the imagistic investigations are important and large scale used because they represent less or non-invasive, quick and efficient methods. As a result, in the medical learning, in students or young specialists training, a special attention is pointed to raising their skills in interpreting the medical images and establishing a correct diagnosis and proper medications. In the medical learning process, the courses in traditional or electronic format are accompanied in many cases by a series of images. In general, the presented images number is minimal. Accordingly, the existence of a database with medical images accompanied by diagnosis and treatment, collected by teachers in the process of patients' diagnosis and investigation raises considerable the variety of communicated knowledge. Taking into consideration all these statements, we decided to design and implement a software tool dedicated to medical e-learning with the following functions:

- **Graphical visualization of the MeSH thesaurus (medical terms from Diseases and Drugs categories) with a topic map.** For example, in order to study the Ulcer diagnosis, the student can select this term in topic map. With the help of this graphical modality, the learner can view its description, the synonyms, but also the hierarchical and associative relationships of the selected term with other medical descriptors (for example, the associated drugs).
- **Semantic query of the multimedia medical database from topic map.** The student can select a topic that represents a diagnosis and launch on the database a simple query or a query that takes into consideration the term and all its synonyms. As a result, the system will retrieve the images from the database that belong to selected diagnosis. In a similar way, the query can use a drug name from topic map, and the system will retrieve the medical images representing cases where that drug was used.
- **Content-based visual query on the multimedia medical database.** In this case the query uses the characteristics extracted from images (for example color and texture). It will find in the database all the images that are significantly similar to the query region/image [11].

The student will be able to combine different access options, which is supposed to be the most successful approach in image retrieval. Using content-based visual query with other access methods to medical image database allows students to see images and associated information in a simple and direct manner. The student is stimulated to learn, by comparing similar cases or by comparing cases that are visually similar, but with different diagnostics [6,7]. The topic map can be used as a navigation tool. The user can navigate through topic map depending on its interest subject, bringing in this way big advantages. He does not have to be familiar with the logic of the database, he will learn about the semantic context, in which a collection and its single items are embedded and he may find useful items he would not have expected to find in the beginning. Another element of originality proposed by this paper is building and populating the topic map by automatic mapping the xml file that contains the MeSH thesaurus to the xtm file that represents the topic map. This database with medical images and the modern and combined access possibilities can be used in the education process by the medicine students and by any physician who wants to improve his knowledge for establishing a correct diagnosis based on imagistic investigation, which is frequently used in patient's diagnosis.

2 Related Work

Topic maps represent a new technology for the structuring and retrieval of information, based on principles used in traditional indexes and thesauri, with inspiration from semantic networks. We can mention papers that present interesting and modern modalities of using topic maps in e-learning. For example, TM4L is an e-learning environment providing editing and browsing support for developing and using topic maps-based digital course libraries. The TM4L functionality is enhanced by an interactive graphical user interface that combines a hierarchical layout with an animated view, coupled with context sensitive features [2,3]. Another author proposed topic map ontology, focusing on both students and teachers as active producers of learning resources. Topic maps customize the interface, and the interface should also provide possibilities for online students to share learning resources like "on campus" students do [5]. The examples can continue. Among the resources that can act as a source in topic map designing and population we find thesaurus. Kal Ahmed has created a standard topic map ontology that can be used to represent thesauri in topic maps [1, 4]. The technological development and the Internet contributed to the development of e-learning resources in the medical domain. Repositories and digital libraries for access to e-learning materials were established (MedEdPortal, Association of American Medical Colleges, End of Life/Palliative Resource Center, The Health Education Assets Library, Multimedia Educational Resource for Learning and Online Teaching, International Virtual Medical School) [9]. Most of the multimedia databases used in medical e-learning allow only simple text-based queries. Another modern query method is based on content, for example on image automated extracted features like color, texture, shape or regions. Medical teaching together with diagnostic aid and medical research represent the most important direction of using content-based visual query on multimedia medical databases [6, 7]. We mention MedGift, a valuable content-based visual retrieval system that works on a huge medical image database. The system is used in medical e-learning at the University Hospital in Geneva (`http://www.dim.hcuge.ch/medgift/01_Presentation_EN.htm`). Taking into consideration the above things about a thesaurus representation with a topic map, the paper presents our proposal of topic map design and population starting from MesH thesaurus with medical descriptors, as a first element of originality. We also present the topic map graphical view for educational purpose. Comparing with other similar tools already used in e-learning, our graphical view window is not only a navigation tool, but also a tool for semantic querying of the multimedia medical database. Then, on the records retrieved by the semantic query, the learner can launch the content-based visual query on color and texture features, at the image or region level.

3 The Medical Image Database

The system offers the possibility to insert new images in the database, together with their relevant information, namely: path and name of the image file, the diagnosis, treatment, evolution, etc. For realizing the content-based visual query, all the images loaded in the database are automatically processed in two steps for extraction of color feature and then of texture feature. For color feature extraction, the images are pre-processed, namely they are transformed from the RGB color space to HSV color space and quantized to 166 colors, being thus prepared for a future query [11]. The color information is stored in the database as a vector with 166 values and it is used furthermore in the content-based image query and content-based region query [11]. Together with color, texture is a powerful characteristic of an image, which is present in nature and in medical images also. Thus, a disease can be indicated by changes in the color and texture of a tissue [6]. There are many techniques used for texture extraction, but there is not any certain method that can be considered the most appropriate, this depending on the application and the type of images taken into account. The effectuated studies on medical images indicated that among the most representative methods of texture detection is the Gabor representation, reason for which it was chosen for extracting the colour texture feature from medical images in the database [15]. After the algorithm execution [13], the resulted image texture feature that is a 12-dimension vector is stored in the database.

4 MeSH Thesaurus

MeSH is the National Library of Medicine's controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. One or more terms, comprising one or more concepts, grouped together for important reasons, form a descriptor class. Relationships among concepts can be represented explicitly in the thesaurus. Hierarchical relationships are at the level of the descriptor class and often reflect important broader-narrower relationships between preferred concepts in descriptor classes. Other types of relationships present in the thesaurus include associative relationships, such as the pharmacological actions or see-related cross-references [8, 14]. We have selected from MeSH record only a number of data elements that are important in medical e-learning [14]: MESH HEADING (MH), ENTRY TERM, MESH TREE NUMBER, ANNOTATION, MESH SCOPE NOTE, PHARMACOLOGICAL ACTION, FORWARD CROSS REFERENCE. In our opinion, there are in MeSH three types of relationships between medical terms that are significant for medical e-learning [14]:

1. **Equivalence relationships**. This type of relationship is established between entry terms and main headings elements (entry terms has been thought of as synonyms and quasi-synonyms of the main heading).

2. **Hierarchical relationships**. These relationships are fundamental components in a thesaurus and represent powerful tools for retrieval. MeSH has an extensive tree structure, currently at nine levels, representing increasing levels of specificity.
3. **Associative relationships**. Many associative relationships are represented by the "see related" cross-reference. Other attributes, which can be thought of as an associative relationship are the Pharmacologic Action, Scope Note and Annotation. For example, this kind of relationship is established between a diagnosis and its cause or between a drug action and disease.

5 Topic Maps

Topic maps define a model for the semantic structuring of knowledge networks. They represent a solution for organizing and accessing large information pools, providing a bridge between knowledge management and information management domains. The basic concepts of the topic map paradigm are [10]:

- Topics that represent subjects and carry names
- Occurrences of a topic that point to relevant resources
- Associations that connect related topics
- Topic classes, occurrence classes and association classes that help to distinguish different kind of topics, occurrences and associations, respectively

Manual topic map population may require lots of resources. As a result it is considered the problem of the self-populating the topic maps. The available resources that can act as a source of input to auto-population are identified: ontology, relational or object-oriented database, metadata about resources, index glossary, thesaurus, data dictionary, document structures and link structures or unstructured documents [10]. It can be observed among these resources, the presence of thesaurus and XML, which are used for the design and self-population of topic maps, solution adopted by us also. When converting a thesaurus to a topic map, the thesaurus provides the topics and some basic associations: the part-whole and related-to association, synonyms and related terms.

5.1 Building the Topic Map

To generate the xtm file starting from the "Diseases" and "Chemicals and Drugs" categories of the MeSH thesaurus, two files that can be downloaded at no charge from MeSH site are used [8]:

- An xml file named desc2008 that gives information about medical descriptors.
- A txt file trees2008 that gives information about the organization of these medical descriptors in a tree structure.

The following steps should be followed:

1. **Selecting from desc2008.xml file only the tags considered to be useful to generate the topic map.**
   These tags are specified in a text file (AllowableXMLKeywords.txt) and they are: Xml, DescriptorRecordSet, DescriptorRecord, DescriptorUI, DescriptorName, String, TreeNumberList, TreeNumber, ConceptList, Concept, TermList ,Term, Annotation, ScopeNote, PharmacologicalActionList, PharmacologicalAction, SeeRelatedList, SeeRelatedDescriptor, and DescriptorReferredTo. For the reduction operation (tags filtering) the application needs the path to the desc2008.xml file and it will generate another xml file that will contains only the types of tags specified in the txt file.
2. **Processing the reduced xml file**
   During processing the reduced xml file, all the required information for topic generation is collected:

   - The content of the tag <DescriptorUI> will represent topic's ID.
   - The content of the tag <String> will represent the baseNameString for the topic
   - The content of the tag <String> from <TermList> will represent the resource data occurrence
   - The <TreeNumberList> tag's content will be used to generate associations

   The medical terms presented in the **desc2008** file are organized into a hierarchy and the **TreeNumber** tag indicates the sub-tree were the descriptor is included. The content of each **TreeNumber** tag (that is unique) will be used in generating the associations for the topic map [12].
3. **The xtm file generation**
   After all the information needed to generate topics and associations were collected, will be generated an xtm file according to standard xtm 1.0. This process is composed of [12]:

   - Topics generation process
   - Associations generation process that contains two phases:

   Phase 1) "Part-hole" associations generation
   Phase 2) "Related-to" associations generation

5.2 *Graphical View of the Topic Map*

Topic maps are very powerful due to their ability to organize information. When the amount of information is very large, intuitive user interfaces must be defined. Users need to know what the main subjects of the topic map are and how to access them. The creating process of the topic map and the visualization techniques must take account on these requirements. In figure 1 we can see a part from the topic map, built for a much coherent visualization of this medical terminology. The viewing interface for topic maps is organized in two panels: the left panel displays a list with topic map items (topics, topic types, associations, occurrences) and the right panel displays details about topics and associations as a tree structure. Certain topics represent types of topics: descriptor, EntryTerm, part, whole, part-whole, related-to and related. The list of topic map items can be seen as an index by the user and can be used to explore the content of the topic map. The user can select a topic from the list. After selection, the content of the selected topic is presented in details in the right panel. Based on the content provided by MesH descriptors, each topic is defined as an instance of the descriptor topic. For each topic there are defined occurrences of type EntryTerm. For the analyzed topic are defined two occurrences: “Ulcerative Colitis” and “Colitis, Ulcerative”. These occurrences are seen as synonyms and will be used for the search operation realized by a context menu with two options as in the figure 1. Based on the hierarchical structure of Mesh, a diagnosis category has several descriptors. This fact is reflected in the content of the topic map by the existence of associations that are of “part-whole” type. Each topic involved in association plays a certain role; the categories are of “whole” type and the descriptors of “part” type (figure 2). Labels and icons are used to understand better the relationships between the displayed topic maps objects. The “R” icon indicates the roles that are played by the topics in association. Each association is defined as an instance of a topic type. This fact is indicated by the “I” icon.

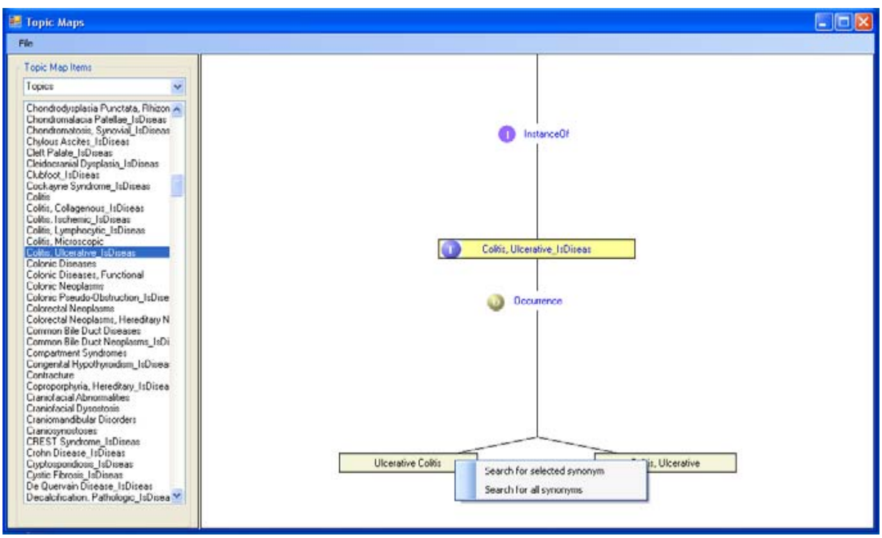


Fig. 1: Interface for topic map view.

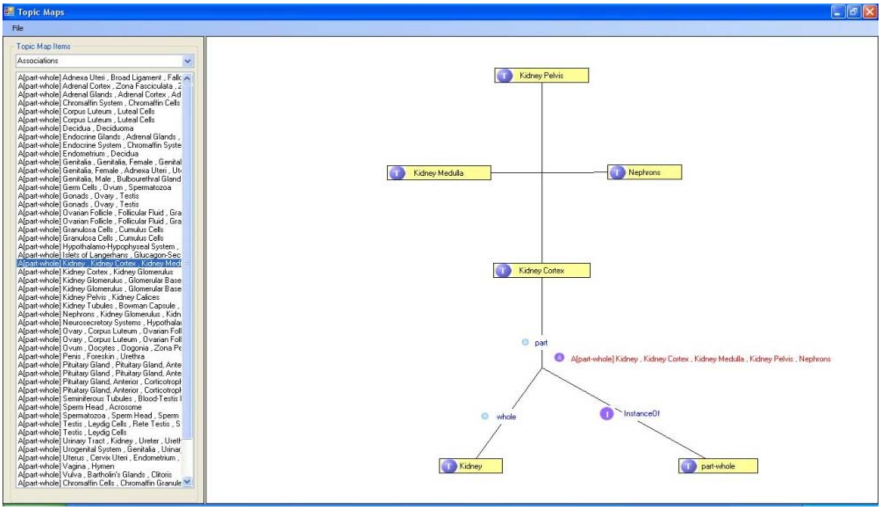


Fig. 2: A part-whole association example.

6 The Multimedia Medical Database Query

From the topic map graphical display window it is possible the semantic query of the multimedia medical database. In our topic map, since the occurrences are stored in the database, every synonym will be defined as a database query. The medical term “peptic ulcer”, for example, will initiate a query for “peptic ulcer” in the diagnosis field of the table containing images in the database. Consequently, every relevant image of “peptic ulcer” will be retrieved. The database search can be done in two ways:
Using a single synonym. In this case it is executed a Select command with the general form:
SELECT * FROM image where diagnosis = “synonym”
Using all the synonyms and the Select command has the following form:
SELECT diagnosis FROM image where diagnosis = “synonym1” or diagnosis = “synonym2” or . . .
The second query modality is very useful in the medical domain, because different specialists introduce the images in the database and for diagnosis they can use synonyms. The semantic query can be done in the same way for terms that represent drug name. This access path to the medical imagistic database can be combined with the content-based visual query, the results being useful to the doctor in the diagnosis process. For example, the student can select a certain diagnosis. There are displayed the corresponding images from that diagnosis and their afferent information. From the image set, the doctor can choose one and launch the content-based visual query on color and texture characteristics [13]. As result the first 10 images from the database that are most similar to the image query will be displayed. The

student can compare the visually similar images that belong to the same diagnosis, but also the images that are not visually similar from the point of view of color and texture, but that belong to the same diagnosis.

7 Conclusions

The paper presents an original application of the topic map concept in the medical domain: for browsing a thesaurus with medical terms and for semantic querying of a multimedia database containing medical images and their diagnosis and applied treatment. The new software tool was accepted by the teachers from Gastroenterology department of the Medicine and Pharmacy University from Craiova and appreciated as useful and original. During the year 2008, 60 students used the e-training module based on imagistic database and content-based visual query in study of the gastroenterology discipline. Each of them accessed the database for approximately 9 times, spending in average 200 minutes. The 60 students participated to the following experiment: they were asked to study the diagnosis from the Digestive System Diseases category, first using MeSH thesaurus and its organization as it is presented on the National Library of Medicine site(http://www.nlm.nih.gov/cgi/mesh/2008/MB_cgi? term=Digestive+System+Diseases&field=entry#TreeC06). After that they had to analyze the same diagnosis using the topic map created with this software tool. The result was that 75% from them considered that topic map is a much more intuitive alternative, especially because allows the graphical visualization of the associations between medical terms. 25% from the students considered that both alternatives are efficient. The students were also invited to query the multimedia medical database in two ways: using semantic querying available when the topic map is used and simple text based query that needs from the user to enter the diagnosis he is looking for. The results will present all the images that were categorized with that diagnosis. 90% of the students considered that is much more efficient the query available in topic map, especially the one that uses synonyms. This way the result is more detailed and the knowledge obtained are more complete.

References

1. Ahmed, K.: Topic Map Design Patterns For Information Architecture (2003), `http://www.techquila.com/tmsinia.html`
2. Dandan, W., Dicheva, D., Dichev, C., Akouala, J.: Retrieving information in topic maps: the case of TM4L. In: ACM Southeast Regional Conference, pp. 88–93 (2007)
3. Dicheva, D., Dichev, C., Dandan, W.: Visualizing topic maps for e-learning. In: Advanced Learning Technologies (ICALT 2005), pp. 950–951. IEEE Computer Society, Los Alamitos (2005)
4. Garshol, L.M.: Metadata? Thesauri? Taxonomies? Topic Maps! Journal of Information Science 30(4), 378–391 (2004) CILIP (Chartered Institute of Library and Information Professionals)

5. Koløas, L.: Topic Maps in E-learning: An Ontology Ensuring an Active Student Role as Producer. In: Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Ed/ITLib Digital Library, Association for the Advancement of Computing in Education (AACE), pp. 2107–2113 (2006)
6. Muller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A Review of Content based Image Retrieval Systems in Medical Application - Clinical Benefits and Future Directions. International Journal of Medical Informatics 73(1), 1–23 (2004)
7. Muller, H., Rosset, A., Garcia, A., Vallee, J.P., Geissbuhler, A.: Benefits of Content-based Visual Data Access in Radiology. Radio Graphics 25, 849–858 (2005)
8. National Library of Medicine, Medical Subject Headings, MeSH Tree Structures (2008), `http://www.nlm.nih.gov/mesh/2008/MeSHtree.C.html`
9. Ruiz, J., Mintzer, M.J., Leipzig, R.M.: The Impact of E-Learning in Medical Education. Academic Medicine 81(3), 207–212 (2006)
10. Rath, H.: The Topic Maps Handbook. Empolis GmbH, Gutersloh (2003)
11. Smith, J.R., Chang, S.F.: Automated image retrieval using color and texture. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) - Special Issue on Digital Libraries: Representation and Retrieval (1996), `http://www.ctr.columbia.edu/~jrsmith/html/pubs/PAMI/pami_final_1.html`
12. Stanescu, L., Burdescu, D.D., Mihai, G., Ion, A., Stoica Spahiu, C.: Topic Map for Medical E-Learning. In: Proceedings of the 2nd International Symposium on Intelligent Distributed Computing (IDC 2008). Studies in Computational Intelligence, vol. 162, pp. 305–310. Springer, Heidelberg (2008)
13. Stanescu, L., Mihaescu, M.C., Burdescu, D.D., Georgescu, E., Florea, L.: An Improved Platform for Medical E-Learning. In: Leung, H., Li, F., Lau, R., Li, Q. (eds.) ICWL 2007. LNCS, vol. 4823, pp. 392–403. Springer, Heidelberg (2008)
14. Stuart, J.N., Douglas Johnston, W., Humphreys, B.L.: Chapter 11: Relationships in Medical Subject Headings (MeSH), `http://www.nlm.nih.gov/mesh/meshrels.html`
15. Zhang, D., Wong, A., Infrawan, M., Lu, G.: Content-Based Image Retrieval Using Gabor Texture Features. In: Proc. IEEE Pacific-Rim Conference on Multimedia, pp. 392–395 (2000)

An Infrastructure for Pervasive Access to Clinical Data in eHospitals

Massimo Esposito, Luigi Gallo, Antonio Coronato, and Gennaro Della Vecchia

Abstract Ubiquitous computing technologies are being applied in many fields of business and institutions, varying from small intelligent spaces to large virtual enterprises. In particular, such technologies can be successfully used in health care facilities in order to reduce medical costs and improve quality of service. This paper presents an infrastructure for pervasively accessing Electronic Health Records (EHR) in a hospital. It relies on services which integrate Radio Frequency IDentification (RFID) and photosensor technologies for identifying, locating and tracking doctors and patients equipped with mobile devices and RFID tags, with the final aim of granting ubiquitous and transparent access to medical data stored into standard EHRs.

Massimo Esposito
Institute for High Performance Computing and Networking (ICAR) - Italian National Research Council (CNR),
Via Pietro Castellino 111, 80131, Naples, Italy
e-mail: massimo.esposito@na.icar.cnr.it
University of Naples "Parthenope", Centro Direzionale, Isola C4, 80143, Naples, Italy
e-mail: massimo.esposito@uniparthenope.it

Luigi Gallo
Institute for High Performance Computing and Networking (ICAR) - Italian National Research Council (CNR),
Via Pietro Castellino 111, 80131, Naples, Italy
e-mail: luigi.gallo@na.icar.cnr.it
University of Naples "Parthenope", Centro Direzionale, Isola C4, 80143, Naples, Italy
e-mail: luigi.gallo@uniparthenope.it

Antonio Coronato
Institute for High Performance Computing and Networking (ICAR) - Italian National Research Council (CNR),
Via Pietro Castellino 111, 80131, Naples, Italy
e-mail: antonio.coronato@na.icar.cnr.it

Gennaro Della Vecchia
Institute for High Performance Computing and Networking (ICAR) - Italian National Research Council (CNR),
Via Pietro Castellino 111, 80131, Naples, Italy
e-mail: gennaro.dellavecchia@na.icar.cnr.it

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 431–442.
springerlink.com

1 Introduction

The ultimate goal of Ubiquitous Computing is the development of environments where highly heterogeneous hardware and software components can seamlessly and spontaneously interoperate, in order to provide a range of services to users independently of the specific characteristics of the environment and of the client devices [8]. The widespread availability of new sensor technologies makes it possible to enhance existing environments with pro-active features -the capability of a system to "foresee" beforehand resources and services requests- by adding context-aware services. This is the case of the Radio Frequency IDentification (RFID) technology [4], which relies on the use of RFID tags (or transponders) and antennas. Healthcare institutions constitute very suitable settings for deploying ubiquitous services, where wireless technologies can dramatically reduce medical costs and improve quality of service [12]. In this paper we propose an infrastructure which integrates RFID and photosensor technologies to identify, locate and track patients and doctors in a hospital department with the final aim of granting ubiquitous and transparent access to medical data gathered into standard Electronic Health Records. Using RFID alone in health facilities is not a novel idea, since in recent years several pilot research projects, investigating the feasibility of RFID for tracking patients, assets, pharmaceuticals and personnel, have been conducted in hospitals [1, 2]. Furthermore, some commercial solutions to identify and locate patients and doctors have already been presented [11]. Notwithstanding, it is a matter of fact that integrating turnkey RFID-based healthcare solutions into a real, fully operating hospital environment is a thorny issue, mainly because of the scarce, if any, compatibility with the existing software. For this reason, our research efforts have been primarily focused on the design of a framework that was open-source and cross-platform, hence capable to easily interoperate with legacy software and databases usually present in hospitals. This paper is organized as follows. Section 2 discusses some motivations and related works. Section 3 describes the infrastructure architecture. Section 4 describes a running scenario. Finally, in Section 5 some conclusions are drawn and the Reference section ends the paper.

2 Motivations and Related Works

Over the years, the scientific community has been making a remarkable effort to introduce information technologies in hospitals and clinical environments. Most work focused on the electronic version of the patient's medical file. In particular, the term *Electronic Health Record* (EHR) has been coined, meaning a computer-based record of patient health information intended primarily for use by healthcare providers. A fully-functional EHR incorporates all patient records of visits, hospitalizations, and other encounters with the healthcare system. Recently, it has been proved that new pervasive computing technologies can remarkably enhance the Quality of Service in healthcare [5]. By giving medical professionals appropriate, complete information, it is expected to deliver better care that is tuned not only to the situation but also to the patient's history. Meanwhile, the fast growing RFID technology, initially developed for applications in logistics and industry, suggested to integrate mechanisms for locating and tracking mobile objects based on RFID tagged entities in a pervasive computing environment. Many papers can be found in literature dealing with this fresh approach. Bardram et al. in [3] use RFID technology to identify patients while hospitalized, and control drugs assumption. In particular, beds in the wards are equipped with various RFID sensors which identify approaching tagged drugs so to reduce medication errors by verifying that the right drugs are given to the right patient. De et al. in [7] describe a location sensing prototype system that uses RFID technology for locating objects inside buildings is presented, where the authors demonstrate that active RFID is a viable and cost-effective candidate for indoor location sensing. Satoh in [9, 10] presents a general-purpose infrastructure for building and managing location-aware applications in ubiquitous computing settings, with the goal of providing people, places, and objects with computational functionalities to support and annotate them. Coronato et al. in [6] describe an Access&Location service for pervasive grid applications which uses RFID and WiFi technologies to grant access and locate mobile objects in a pervasive way within a computer grid infrastructure. Our contribution consists of an infrastructure that provides users with pervasive access to Electronic Health Records in a eHospital scenario. We focused on a solution capable to easily interoperate with heterogeneous software and clinical databases already present in hospitals. The system's design and implementation adopt a methodology which integrates two different sensing technologies, RFID and photoelectric sensors, in order to gather context-depended information. In particular, suitable services have been realized to implementing the following functionalities:

- Identifying mobile users - RFID tagged users (both doctors and patients) are identified via the RF system.
- Locating mobile users - accesses to specific rooms detected by photosensors trigger the RFID system which in turn reads the presence of tagged users so that they are identified and located; this service also provides a mechanism for notifying unauthorized accesses.

- Tracking mobile users - RFID tagged users are tracked while moving within the environment.
- Accessing EHRs - Users are provided with ubiquitous access to EHRs. In particular, the access is context-aware, i.e. the kind of medical information retrieved depends on the user position.

3 An Infrastructure for eHospitals

The proposed infrastructure consists of a set of services and components shown in Fig. 1 and described below

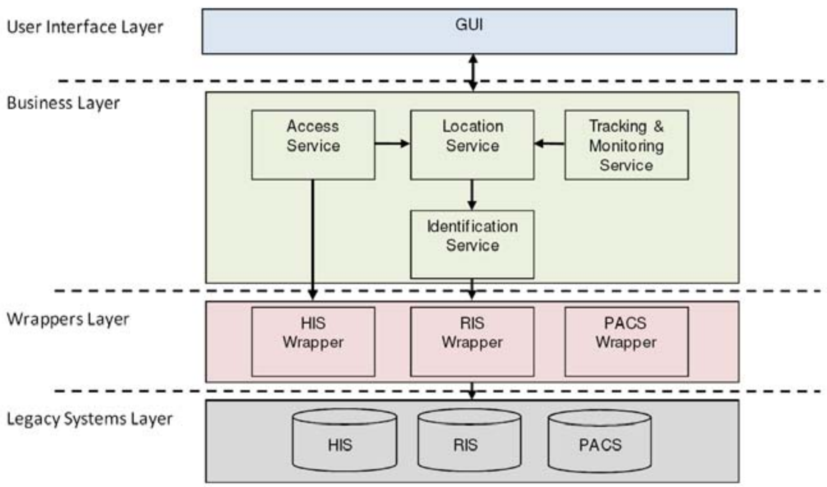


Fig. 1: System Architecture.

3.1 Identification service

This service provides features to i) register/discharge patients and ii) identify patients and medical staff.
i) Routinely, when a fresh patient is checked in at the hospital, the medical staff creates a new record in the EHR system storing the patient's personal and clinical historical data and a unique PatientID (PID) is assigned to the patient. A specific functionality of the identification service allows to initialize a wearable RFID tag -which has its own ID- with the patient's PID so that the patient is univocally

associated to that tag. The tag will be eventually given to the patient to be worn all along her/his staying in the facility. When a patient is released from the hospital, a "kill" command -a peculiar feature of the EPC protocol (see section 4), is issued to the patient's tag so that it can never be used again. It is supposed that medical staff is registered by default, each member having her/his own wearable RFID tag with unique PID.
ii) Whenever a tagged person is detected by an RFID antenna, the tag is read and that person is identified through the PID it is associated to.

3.2 Location service

This service provides features to locate RFID tagged users. It offers both locating and location functions, that is, the function Locate_person returns the position of a specific person, whereas the function Get_people returns the list of people located at a specific position. Moreover, the service uses an asynchronous communication mechanism for notifying the environment when a tagged person enters a specific room, i.e. it generates the event NEW_EVENT(person, room). The architecture of Location Service consists of two layers with the following components:

- LocatingComponent - This component is in charge of locating RFID tagged objects. An RFID location is identified by the area covered by the RFID antennas of a reader installed in a specific room which is accessible through a watched gate. A watched gate relies on two oriented photoelectric sensors to reveal accesses. The couple of photosensors is able to detect the direction of the motion in between them, so that it is possible to infer whether a person is entering or leaving the room. The powering on of RFID antennas is triggered by the photosensors, thus reducing RF exposure and power consuming. There are as many LocatingComponent as RFID locations, each component being associated to a specific reader. Whenever a person passes throughout a watched gate, that person's movement is detected by the couple of photoelectric sensors. This event is caught by the LocatingComponent which recognizes the movement direction and decides whether the person is entering or leaving the gate. In the same time, the component issues a command to power on the RFID antennas placed in the room. If a tag is sensed, it reads the tag ID and the PID stored during the initialization and uses them to possibly identify the person, who can be a registered patient or a medical staff member. If the tag ID or the PID are not recognized as valid or if no tag is detected at all, the component notifies an unauthorized access. These activities are depicted in the diagram shown in Fig. 2.
- LocationComponent - This component is in charge of handling global location states obtained by combining information coming from below components. In particular, it handles information on last detected user position. Whenever an LocatingComponent detects an RFID tag, it communicates such information to the LocationComponent, which, in turn, compares it with the last known position for that user and notifies the environment in case of change. This mechanism is also

used to detect and resolve collisions, which arise when for some circumstances the same RFID tag appears to be sensed by two readers simultaneously.

Fig. 2: Activity Diagram for locating patients and detecting unauthorized accesses.

3.3 *Tracking & Monitoring service*

This service receives location information from the Location service for keeping track of the movements of patients and medical staff and monitoring accesses to specific rooms by logging these activities. Namely, the Tracking & Monitoring service catches all the NEW_EVENTs notified by the Location service and updates the log accordingly.

3.4 Access service

This service constitutes the core business of the infrastructure. It provides the medical staff with ubiquitous access to Electronic Health Records, on the basis of the information supplied by the Location service. The Access service exhibits a reactive and context-dependent behavior as automatic actions are triggered in response to patients and medical staff movements within the environment, namely the NEW_EVENTs events above. This behavior can be described by means of a set of rules conforming to the Event-Condition-Action (ECA) architectural pattern. Generally, ECA rules have the form "when¡event¿if¡conditions¿then ¡action¿", where ¡conditions¿ specifies the circumstances that must be verified for the ¡action¿ to be carried out whenever ¡event¿ occurs. In our case ¡conditions¿ can be any logical combination of patients/doctors staying in one specific room while ¡action¿ actually consists in a series of operations that eventually will led to formatting and generating a suitable query for EHR and consequently showing the results on a display in that room. For instance, when a patient enters the surgery and a doctor is present, the service automatically queries and retrieves the patient's Electronic Health Record and shows it on the terminal in surgery. This behavior can be formalized by the following ECA rule associated to the Surgery:

```
WHEN (NEW_EVENT(patient, Surgery))
  IF( ∃ doctor ∈ Get_people(Surgery)) THEN
     <query for patient's EHR and display the results in Surgery>
```

Each room in the health facility has its own ECA rule purposely devised, reflecting the specific conditions to be met for the room. As a further example, let us consider a ward hosting some patients, say the ward no. 7. When a doctor enters the ward no. 7 for a routine visiting, queries for accessing the EHRs of those patients are proactively generated and the results stored in the PC installed in the ward. Then, the doctor -through a GUI running on the PC - can select a patient and peruse her/his own EHR. The ECA rule in this case assumes the form:

```
WHEN (NEW_EVENT(doctor, Ward#7))
  IF( ∃ patient ∈ Get_people(Ward#7)) THEN
    <retrieve the EHR of each patient in the list and
     launch a dialog box on the PC in Ward#7 >
```

3.5 Wrappers

The Wrapper components provide features to access the EHR systems, namely HIS, RIS and PACS. HIS (Hospital Information System) is the information system designed to cover all the information transactions occurring in the clinical practice, from administrative and financial issues to medical data handling. RIS (Radiological Information System) is the information system used by radiology departments

to store, manipulate and distribute radiological data and images. PACS (Picture Archiving and Communication System) is the information system in charge of storing, retrieving, distributing and presenting medical images.
HIS and RIS are typically complaint with the HL7 (Health Level 7) standard, which provides message-based protocols for the exchange, management and integration of clinical and administrative electronic health data. PACS is typically compliant with the DICOM (Digital Imaging and COmmunications in Medicine) standard, which defines a file format for medical images and provides a protocol for their exchange based on a set of commands. The wrappers implements handlers for communicating with HIS, RIS and PACS through their respective standards. Specifically, such components allow to i) formalize messages or commands in the standards HL7 ver. 3.0 and DICOM ver. 3.0, and ii) to send and receive messages, commands and data to and from HIS, RIS and PACS.

4 Experimental scenario

The infrastructure has been devised to be eventually deployed in the experimental scenario depicted in Fig. 3.

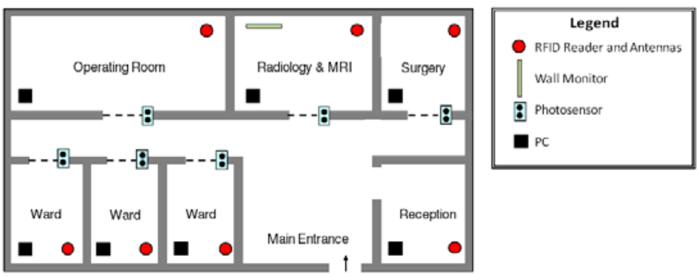


Fig. 3: Hospital layout.

It consists of a number of physical environments, as described below. Each environment is accessible through a watched gate (except the Reception) and is equipped with RFID reader and antennas; a desktop PC for accessing system functionalities is installed too.

- Reception - Here the patient is accepted for hospitalization. A desk operator (medical staff) registers the patient (i.e., a new EHR is created and all required data are input), a fresh wearable RFID tag is associated to the patient's PID and given to the patient.
- Wards - Rooms where patients reside.
- Operating room - This is the room where medical staff provides surgical services to patients.

- Radiology & MRI - This is the setting where radiological tests are carried out. This room is equipped with a wall monitor for displaying the produced medical images.
- Surgery - In this room doctors perform specific check ups.

An experimental prototype for this scenario has been realized which implements the functionalities here described:

- Initialize & Release Patient - These functionalities enable the medical staff at the Reception to assign an RFID wearable tag to a new patient and void the tag upon the patient discharge. The Initialize functionality includes accessing the HIS for retrieving the patient's EHR.
- Locate Person - This functionality allows the medical staff to locate medical personnel and patients.
- Get people - This functionality is intended to get a list of people located at a specific location.
- Visualize EHRs - This functionality provides medical staff with ubiquitous access to the patient's EHR. This functionality includes accessing HIS, RIS and PACS in order to retrieve clinical reports and images of the patient.
- Track & Monitor - This functionality enables medical staff to obtain reports on people movements and room accesses.
- Generate Location Information - This functionality yields location information gathered through the detection of people movements. It includes the identification of medical personnel and patients and is capable of notifying possible unauthorized accesses.

These functionalities are modeled in the use case diagram shown in Fig. 4.

The prototype mainly consists of a set of front-end clients running on the PCs installed around the health facility, a back-end centralized server and several sensor devices. The front-end clients expose a Graphical User Interface to let the medical staff use the above mentioned functionalities. The GUI consists of a window application program developed in Java language. The back-end centralized server hosts all the software components described in Sect. 3. Such components have been developed as standard Web Services. They are fully portable Java entities, with the exception of the Identification and Location Services, which also integrate, as native code, Microsoft Visual C++ 6.0 control drivers provided by the RFID technology and photoelectric sensor manufacturers. Specifically, RFID readers belonging to FEIG's Passive RFID Long Range Reader series, model ISC.LRU2000-A have been employed. These readers are able to interact with EPC Class 1 Gen 2 transponders for operating frequencies in the UHF range 865-870 MHz and reading range of up to 10 meters (approx. 33 feet). FEIG's UHF antennas model ID ISC.ANT.U250/250-EU have been installed, too. These antennas are circular polarized for operating frequencies in the UHF range 865-870 MHz. Omron photoelectric retro-reflective sensors, model E3F2-R4B4-E 2M have been used, too. As an example of the pervasive access to clinical data allowed by the proposed infrastructure, a screenshot of typical radiological images output by the GUI implemented in the experimental prototype for the target scenario is displayed in Fig. 5. This figure shows X-Ray images

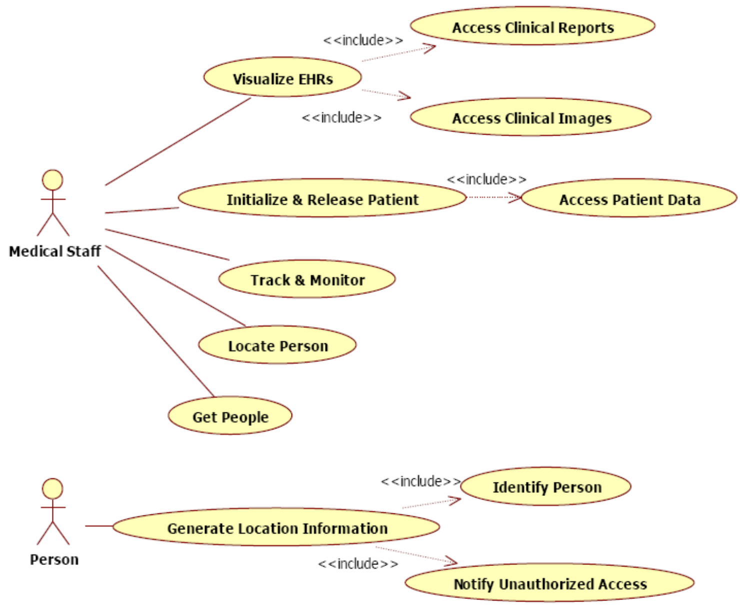


Fig. 4: System Use Cases.

pertaining to a patient detected in the Radiology & MRI room as they are displayed on the wall monitor as a consequence of the fact that the patient has been identified and the proper query to the PACS has been performed, all this being carried out automatically without any human intervention.

5 CONCLUSIONS

The emerged Ubiquitous Computing model fosters the development of environments where the smooth cooperation among distributed heterogeneous hardware and software components grants end-users the capability to access services and resources in a pervasive and totally transparent manner. This approach requires the integration in the system of suitable sensor technologies, which can also be extensively exploited to enhance ubiquitous applications with pro-active features. One such solution can be very effectively adopted in healthcare institutions, especially hospitals, for the urge of optimizing doctor and patient workflows and the need of handling a great deal of clinical data would significantly benefit from Ubiquitous Computing. In this perspective, the present paper has described an infrastructure based on RFID/photosensor technologies which implements mechanisms for

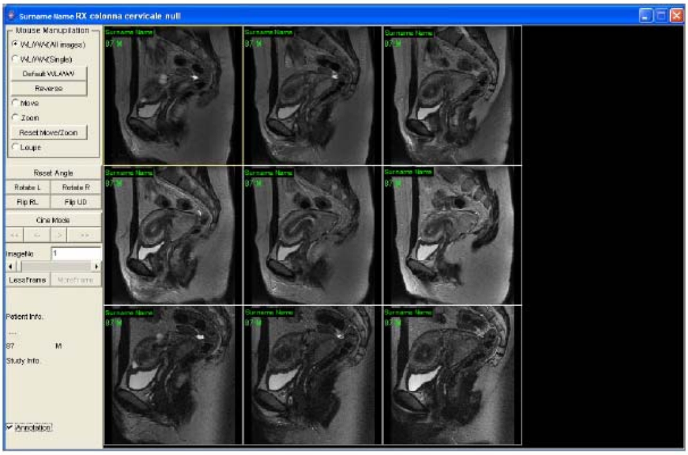


Fig. 5: A screenshot of typical radiological images output by the GUI.

identifying, locating and tracking doctors and patients equipped with wearable RFID tags with the goal of granting ubiquitous and transparent access to Electronic Health Records in a eHospital. A very simple but meaningful hospital layout has been chosen as a running scenario, for which an experimental prototype has been realized in Naples (Italy) at the Institute for High Performance Computing and Networking (ICAR) of Italian National Research Council (CNR). Using information on people movements supplied by passive RFID and photosensor devices, the prototype yields a set of basic services implemented as components which are fully portable Java entities and expose WSDL interfaces. Also, handlers for communicating with legacy HIS, RIS and PACS through their respective standards have been integrated in the system's software architecture, achieving the final aim of allowing pervasive and transparent access to clinical data. Preliminary tests performed on the experimental prototype gave a proof of its feasibility, suggesting that a scaled-up implementation can be effectively deployed in a health care facility. As a concluding remark, it is relevant to observe that, as far as this current implementation is concerned, no security & privacy issues have been considered, leaving the addressing of this yet critical matter for future work, when some mechanisms for S&P policy enforcement will be studied and possibly adopted.

References

1. Bacheldor, B.: Tags Track Surgical Patients at Birmingham Heartlands Hospital. RFID Journal (2007a), http://www.rfidjournal.com/article/articleview/3222/1/1/
2. Bacheldor, B.: HCA North Florida Expands Its RTLS to Track Patients. RFID Journal (2007b), http://www.rfidjournal.com/article/articleview/3615/1/1/

3. Bardram, J.E.: Applications of Context-Aware Computing in Hospital Work - Examples and Design Principles. In: ACM Symposium on Applied Computing (2004)
4. Borriello, G.: RFID: Tagging the World. ACM Communications (2005)
5. Borriello, G., et al.: Pervasive Computing in Healthcare. IEEE Pervasive Computing (2007)
6. Coronato, A., Della Vecchia, G., De Pietro, G.: An RFID-Based Access&Location Service for Pervasive Grids. In: Proceedings of TRUST 2006 - 1st Int. Workshop on Trustworthiness, Reliability and Services in Ubiquitous and Sensor Networks, Seoul, Korea, August 1-4 (2006)
7. De, P., Basu, K., Das, S.K.: An Ubiquitous Architectural Framework and Protocol for Object Tracking using RFID Tags. In: Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, MobiQuitous 2004 (2004) 0-7695-2208-4/04
8. Saha, D., Murkrjee, A.: Pervasive Computing: A Paradigm for the 21st Century. IEEE Computer, Los Alamitos (2003)
9. Satoh, I.: Linking Physical Worlds to LogicalWorlds with Mobile Agents. In: Proceedings of the 2004 IEEE International Conference on Mobile Data Management (MDM 2004) (2004) 0-7695-2070-7/04
10. Satoh, I.: A Location Model for Pervasive Computing Environments. In: Proceedings of the 3rd IEEE Int'l Conf. on Pervasive Computing and Communications (PerCom 2005) (2005) 0-7695-2299-8/05
11. Swedberg, C.: PCTS, Radianse Team for Hospital Tracking Solution. RFID Journal (2006), `http://www.rfidjournal.com/article/articleview/2632/1/1/`
12. Varshney, U.: Pervasive Healthcare. IEEE Pervasive Computing (2003)

Intelligent Systems for Optimized Operating Rooms

Vladimir Stantchev

Abstract Intelligent systems can support healthcare activities in various ways. As an expensive and complex resource the operating room can benefit substantially from such systems. Of particular interest are techniques and systems that allow better utilization of operating rooms. Augmented and mixed reality systems, together with context awareness, can provide such functionality. This work describes a concept for an intelligent operating room. It focuses on an aspect rarely addressed by other intelligent healthcare systems in this sector – resource planning and scheduling of operations. Such system can gather information about the locations and the preparation states of resources and patients and provide optimized operation scheduling. The work also discusses how to address two key aspects for an intelligent healthcare system – quality-of-service (QoS) assurance and integration within existing clinical infrastructures. We use a service-oriented architecture (SOA) as architectural concept and an approach for QoS assurance called architectural translucency to address these aspects of healthcare systems. Furthermore, we present and evaluate a clinical application scenario in the operating room that verifies our approach.

Keywords: public services, intelligent systems, augmented reality, healthcare, location awareness.

1 Introduction

Services in clinical environments typically take place under high workloads and have to handle many different problems in a very short time. Digital assistance through software systems and electronic devices is able to reduce administrative workload and free physicians and nurses for their core competence, taking care of

Public Services and SOA Research Group,
Berlin Institute of Technology and FOM Fachhochschule fuer Oekonomie und Management, Berlin, Germany
e-mail: vstantch@cs.tu-berlin.de

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 443–453.
springerlink.com

patients. Examples for such technologies are hospital information systems (HIS) and electronic patient records (EPR) or electronic health records (EHR) [5] which simplify the access to patient data and medical information. Furthermore, there are devices that can provide information about the location of patients, staff and medical devices in the hospital. Such localization can be done via different technologies, for example Ultra Wide Band (UWB), Bluetooth (BT) or Wireless LAN (WLAN) location applications. UWB has the advantage that it works independently of other systems, allows very precise location and sends with a very low signal strength. WLAN localization on the other hand can be done in existing WLAN networks and allows the localization of computers, such as handhelds and laptops. Hybrid approaches that use two or more such technologies typically provide higher precision and more robust position sensing [7]. Handhelds and laptops are widely used in hospital environments. They enable mobile access to hospital information systems and patient records. We can improve the access to patient data by combining the localization with the data from the HIS and EPR. A doctor can use a laptop while doing the ward round to display only information that is relevant to the patients of the room he is currently in. This would prevent the physician from searching for the patient and therefore save time.

A service-oriented architecture (SOA) is an emerging paradigm to offer functionality in such distributed intelligent environments [19]. A key requirement when using SOA in such scenarios is the assurance of acceptable service levels of Quality of Service (QoS) parameters. QoS parameters are part of the nonfunctional properties (NFPs) of a service. The approach we apply for NFP assurance in SOA is called architectural translucency [25, 26]. We also assess position sensing techniques and how we can use them in healthcare scenarios. This assessment summarizes results from previous works [7, 24]. The main contribution of this work is a design and usability evaluation concept for an intelligent operating room. Furthermore, we present and evaluate an application that demonstrates and validates the concept.

The rest of this paper is structured as follows: in Section 2 we present the technologies and our architectural approach for intelligent healthcare systems. In Section 3 we present and evaluate an application scenario – the intelligent service-based operating room. Related works are given within these sections. Section 4 contains the conclusion.

2 Technologies and System Architecture for an Intelligent Operating Room

In this section we first describe technologies for position sensing and their applicability in an intelligent service-oriented operating room. In the remaining subsections we define architecture as applied to service-based intelligent systems and describe our approach for QoS assurance in the context of intelligent healthcare systems.

2.1 Position Sensing Technologies

Several research projects and commercial products are offering positioning using WLAN or other wireless technologies. Depending on the system and used technology different levels of precision can be reached. The Horus system developed at the University of Maryland [29] and the RADAR system by Microsoft Research [4] were one of the first viable efforts to provide WLAN-based position sensing. They are using a WLAN signal strength map for positioning and reach precision of less than three meters. A downside is that they are requiring huge initial effort to set up the signal strength map. A similar system is being developed and offered by Ekahau [10]. These systems work only in areas where signals from enough WLAN access points are received (at least three) and need to be recalibrated in case the infrastructure is changed. The Place Lab system by Intel Research [14] is pursuing an opposite approach minimizing the needed initial effort. The system provides WLAN positioning in a whole city. The needed data is gathered by ”war driving” (driving around in a car and collect WLAN access point signals). The achieved precision is much less but the system is more prone to changing infrastructure.

Systems based on RFID technologies or BT for positioning are offering slightly better precision [11, 2] but cannot be built on existing infrastructure. Additional RFID tags and readers, or BT infrastructure have to be installed.

Hybrid positioning systems use two or more wireless communication technologies for position sensing. One example is MagicMap [7] where WLAN, RFID and ZigBeeare used. MagicMap delivers 33% better precision when at least two technologies are available.

2.2 Software System Architecture

The term *architecture* is intuitively associated with what humans see and experience. From such perspective people understand architecture as physical structures and their construction to form a coherent whole. Beside this external view there is the view of the building architect. This perspective allows the architect to provide the required properties in his building – shelter, light, heat, safety, accessibility, etc. [1].

A general definition of *software architecture* is as a structure composed of components, and rules for the interaction of these components [16]. Other definitions include the view of software architecture as components, connections, constraints and rationale [12], as elements, form and rationale [20], and as components, connectors and configurations [13]. All these definitions represent more or less the external view of a software system. In order to account for the inner view of the architect a *software systems architecture* should include also a collection of stakeholder-need statements and a rationale that demonstrates that the external view (components, connections and constrains) can satisfy these statements [12].

The usage of managed environments to facilitate the development of distributed applications leads to transparency of component location – the developer can write a procedure call once and the environment takes care to find the needed component (local or remote) and forward the call to it. Such transparency makes the task of writing distributed applications easier, but rarely accounts for typical NFP constrains related with remote calls, e.g., the need to limit the number of such calls or to process them asynchronously.

2.3 Service-based Environments and Architectural Translucency

Web services are emerging as a dominating technology for providing and combining functionality in distributed systems. Service-based environments are not something fundamentally new, but are rather an evolutionary step. Thus, the transparency of location and the NFP-related problems that we can observe in other distributed environments are also inherent to SOA. Furthermore, the SOAP-based interchange paradigm adds a huge performance overhead. Not only is the message size considerably larger, but every web service request using complex data types involves a serialization of object status by the requesting service before the request is submitted and its deserialization by the service that then processes this request.

SOA combines elements of software architecture and enterprise architecture. It is based on the interaction of and with autonomous and interoperable services that offer reusable business functionality via standardised interfaces. Services can exist on all layers of an application system (business process, presentation, business logic, data management). They may be composed of services from lower layers, wrap parts of legacy application systems or be implemented from scratch [17].

QoS encompasses research efforts throughout major computer science domains – networking, computer architecture, distributed systems. In the context of web services there are works in the area of QoS-aware web service discovery [18], QoS-aware platforms and middleware [28], and context-aware services [27].

The approach we apply to dynamically adapt service performance to changes in load and usage is called architectural translucency [25, 26]. It aims to assure continuous meeting of service levels by employing service reconfiguration at runtime. Thereby, it applies experimental techniques to evaluate possible reconfigurations in existing systems and enforces them to meet specified nonfunctional requirements.

3 Usability Evaluation and Application Scenario for an Intelligent Operating Room

Our application scenario focuses on the operating room – a major cost factor in healthcare services and at the same time a place, where a substantial value is created for a healthcare provider. In this section we present the scenario in more details and present our evaluation and implementation methods.

3.1 Usability Evaluation Techniques

A comprehensive overview of usability evaluation techniques is presented in [15]. It differentiates between inspection methods (heuristic evaluation, cognitive walkthrough, and action analysis) and test methods (thinking aloud, field observation, and questionnaires). These techniques are categorized according to their applicability in different phases of the system development process, to their time requirements and the number of users, evaluators and the complexity of equipment needed for the evaluation, as well as to their intrusiveness. A historic overview and recent developments in usability research of augmented and mixed reality specifically in the health care domain is presented in [6].

Our design and usability approach is described in details in [23]. We consider intrusiveness as particularly important aspect in health care, therefore we apply cognitive walkthrough and action analysis as inspection methods. This requires high expertise from the evaluators, who are either clinicians themselves, or design specialists with extensive domain knowledge. As test methods we use questionnaires and expert interviews.

3.2 Perioperative and Postoperative Processes

An important precondition for the efficient design of an intelligent operating room are the perioperative and postoperative processes.

There are many visions how to redesign and reorganize these processes for maximum operating room productivity, which also bring changes in operating room architecture. Figure 1 [21] shows a ground plan and flow diagram of patient movement through the operating room of the future. Patients are brought from the main registration area (1 and arrow) to the induction area (2). Perioperative preparation and induction of anesthesia occur in the induction area (2), concomitantly with instrument setup taking place in the operating room (3). The sequence is timed so that anesthetized patients are transferred to the operating room (2, arrow, and 3) for surgery as instrument setup is completed. At the conclusion of surgery, patients emerge from anesthesia in the operating room and are promptly transferred to the early recovery area (3, arrow, and 4), or emergence occurs in the early recovery area. After approximately 15 min of recovery, patients are transferred to the postanesthesia care unit (4 and arrow) by the perioperative nurse. The work space provides access to the hospital information system. It is used by surgeons between cases for dictation, order writing, and teleconsultation with patients' families and by the anesthesia team during surgery for perioperative planning for subsequent cases.

We presented a different room and location setting in [24]. While there our focus is on technology selection process for WLAN-based positioning systems, here, our focus lies on the provision of data in such environment. Our objective is to use the existing infrastructure (WLAN, HIS) with some minor extensions (RFID) to allow for clinicians to access EPRs from the HIS in a context-aware way.

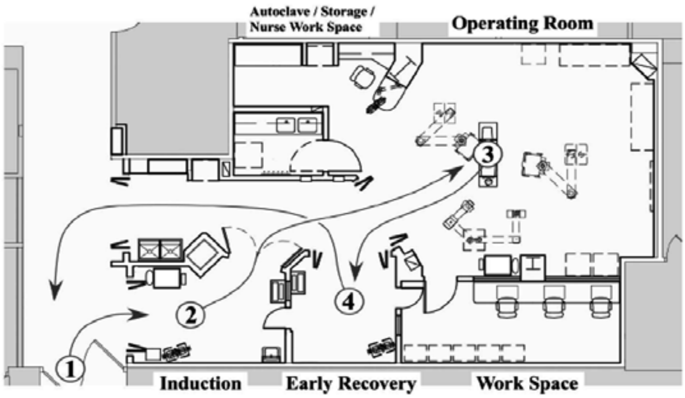


Fig. 1: Example for Optimized Process Flow in an Operating Room.

3.3 Augmented and Mixed Reality

Augmented reality (AR) and mixed reality (MR) are two approaches for advanced human-computer interaction (HCI) that emerged in the last fifteen years. Azuma [3] defines augmented reality (AR) as a variation of Virtual Reality (VR). While VR technologies present to the user a completely synthetic environment without a relation to the real world (the users perceives only the virtual world), in an AR world the user can see the real world with its artifacts, together with virtual objects superimposed upon or composited with the real world. Obviously, AR supplements and enhances reality, rather than completely replacing it. Contrary to authors who interpret AR primarily in the context of a head-mounted display (HMD) Azuma defines three characteristics of AR systems: (i) they combine real and virtual, (ii) they are interactive in real time, and (iii) they are registered in 3-D. Proposed applications of AR systems in surgery include visualization and training support, as well as the collection of 3-D patient data, using non-invasive sensors like Magnetic Resonance Imaging (MRI), Computed Tomography scans (CT), or ultrasound imaging (see [3] for more details and some early projects at the University of North Carolina and at the MIT AI Lab). Our application scenario (context-aware access to patient data) has clear similarities with these examples. Therefore, an AR system appears to be a viable solution.

The definition of an AR system is extended by the definition of an MR system as an interactive system combining physical and digital entities [9]. There is one important difference between these two definitions – while an AR system deals primarily with the output (e.g., visual, 2-D, 3-D) of a system, an MR system deals also with inputs from the real world to the system (e.g., localization and context awareness). Therefore, an MR system generally differentiates between input and output adapters. This is particularly relevant in our application scenario where we focus on localizers as input adapters that monitor proximity and exchange data with the system. This was the main reason to choose an MR system for the realization of the intelligent operating room.

3.4 Integration into an Intelligent System

A suitable WLAN positioning system can be Ekahau [10], or MagicMap. We motivated this selection thoroughly in [24]. The architecture of MagicMap (Figure 2) [7] allows us to use different position sensing techniques (or a combination of two or more of them). It also provides a cooperative positioning which can further improve precision, as well as web service interfaces for the integration in our SOA.

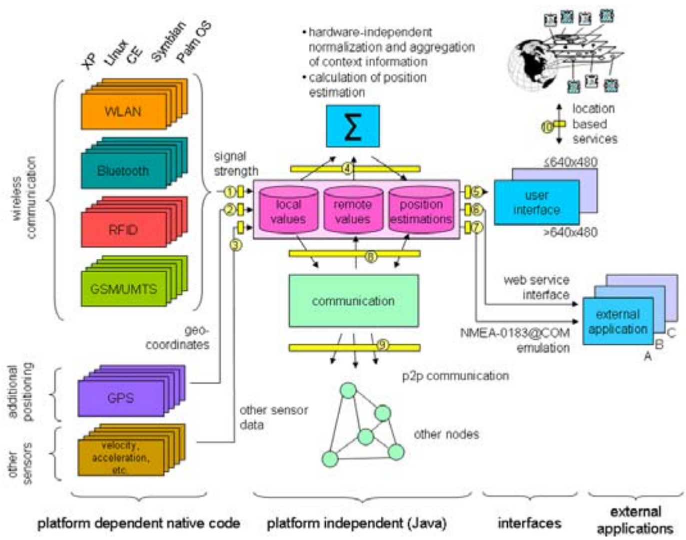


Fig. 2: Architecture of a MagicMap Client.

The health information system should also provide web service interfaces, or we need to implement web service wrappers around the proprietary interfaces. We are using Tablet PCs as mobile devices for the clinicians. Our architectural approach (see Figure 3) is to use an SOA with wrappers that provide web service interfaces to the enterprise service bus (ESB) [8]. The AT engine takes care of service QoS assurance by monitoring and management. When it notices that for example a service is experiencing higher loads or behaves faulty, it dynamically reconfigures the replication settings of the service to further provide the expected QoS. Integration of other systems (e.g., enterprise resource planning, external partner applications, health insurance systems) is facilitated by the use of WSDL-based interfaces. Representation and further information processing are depicted in the upper part of the figure. The system provides portal-based access to EPRs. Information from these ERPs is extracted from the HIS in a context-aware way (e.g., the surgeon sees only information relevant to the patient currently in the operating room and to the specific intervention the medical team is conducting. This information is visualized on different screens (including smart surfaces) within the operating room to best assist the intervention and provides thereby a unique intelligent healthcare environment.

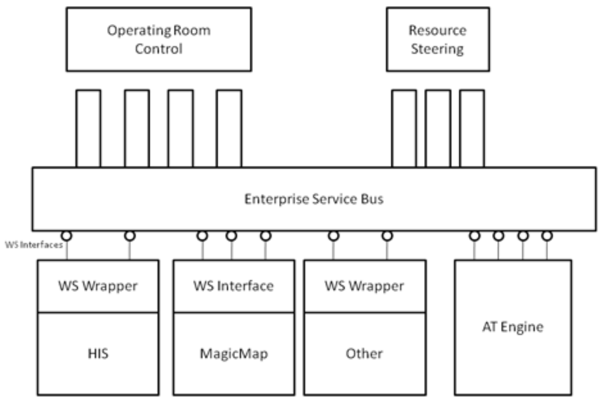


Fig. 3: Systems, Technologies and Architecture for Intelligent Systems in Clinical Environments.

3.5 *System Evaluation*

The continued assurance of performance of our system is evaluated in several works, e.g., [26, 22]. The questionnaires and expert interviews we used as usability evaluation test methods were addressed at clinicians that use our system. An overview of the surveyed group is given in Table 1, a summary of results is presented in Table 2. Overall, there is a substantial ($\Delta > 50 percent$) increase in usability.

4 Conclusion

In this paper we presented and evaluated a concept for an intelligent service-based operating room, together with an application scenario, technologies and system architecture. To address typical design requirements of a mixed reality system in

Group Characteristics	**No. of Participants**	**Percentage**
Nurses	10	33.33
Surgeons	6	20
Anesthetists	3	10
Management	1	3.33
Other	10	33.33
Total	30	100

Table 1: Group Profile

Dependent Variables	**before**	**after**
Average Patient Preparation Time (min.)	31.20	16.30
Avg. Additional Preparation Tasks Needed (Nr.)	12	7
Avg. Number of Process Errors (perioperative)	6	2
Avg. Number of Process Errors (postoperative)	5	2
Clinicians: User Satisfaction (percent)	46	83
Patients: User Satisfaction (percent)	52	89

Table 2: Summary of Usability Evaluation Results

healthcare we used SOA as architectural concept and architectural translucency as an approach to provide QoS assurance in this life-critical environment. Location awareness can be provided by any *off-the-shelf* radio-based positioning system. Specific topics that we are currently addressing in our ongoing research include the standardization of healthcare-related services within an ESB and service priorities (e.g., wireless transmitted alarms regarding vital data) within a limited-resource environment. Furthermore, we are working on a process model for the evaluation of complex process requirements in the healthcare domain and how to better address such requirements using intelligent systems.

References

1. Alexander, C.: Notes on the Synthesis of Form. Harvard University Press (1970)
2. Anastasi, G., Bandelloni, R., Conti, M., Delmastro, F., Gregori, E., Mainetto, G.: Experimenting an indoor bluetooth-based positioning service. In: Proceedings of 23rd International Conference on Distributed Computing Systems Workshops, May 19-22, pp. 480–483 (2003), doi:10.1109/ICDCSW.2003.1203598
3. Azuma, R.: A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments 6(4), 355–385 (1997)
4. Bahl, P., Padmanabhan, V.N.: Radar: An in-building rf-based user location and tracking system. In: INFOCOM, pp. 775–784 (2000)
5. Baron, R.J., Fabens, E.L., Schiffman, M., Wolf, E.: Electronic Health Records: Just around the Corner? Or over the Cliff? Ann. Intern. Med. 143(3), 222–226 (2005), `http://www.annals.org/cgi/content/abstract/143/3/222`

6. Behringer, R., Christian, J., Holzinger, A., Wilkinson, S.: Some usability issues of augmented and mixed reality for e-health applications in the medical domain. In: Holzinger, A. (ed.) USAB 2007. LNCS, vol. 4799, pp. 255–266. Springer, Heidelberg (2007)
7. Bruning, S., Zapotoczky, J., Ibach, P., Stantchev, V.: Cooperative positioning with magicmap. In: 4th Workshop on Cooperative positioning with magicmap. Positioning, Navigation and Communication. WPNC 2007, pp. 17–22 (March 2007), doi:10.1109/WPNC.2007.353606
8. Chappel, D.: Enterprise Service Bus. O'Reilly, Sebastopol (2004)
9. Dubois, E., Gray, P.D., Nigay, L.: Asur++: A design notation for mobile mixed systems. In: Mobile HCI 2002: Proceedings of the 4th International Symposium on Mobile Human-Computer Interaction, pp. 123–139. Springer, London (2002)
10. Ekahau, I.: Ekahau Positioning Engine 3.1 (2005), `http://www.ekahau.com/file.php?id=129`
11. Feldmann, S., Kyamakya, K., Zapater, A., Lue, Z.: An indoor Bluetooth-based positioning system: concept, implementation and experimental evaluation. In: International Conference on Wireless Networks, pp. 109–113 (2003)
12. Gacek, C., Abd-Allah, A., Clark, B., Boehm, B.: On the Definition of Software System Architecture. In: Proc. of ICSE 17 Software Architecture Workshop (1995), `citeseer.ist.psu.edu/gacek95definition.html`
13. Garlan, D., Shaw, M.: An Introduction to Software Architecture. In: Advances in Software Engineering and Knowledge Engineering (1993)
14. Hightower, J., LaMarca, A., Smith, I.: Practical lessons from place lab. IEEE Pervasive Computing 5(3), 32–39 (2006)
15. Holzinger, A.: Usability engineering methods for software developers. Commun. ACM 48(1), 71–74 (2005), `http://doi.acm.org/10.1145/1039539.1039541`
16. Jones, A.: The Maturing of Software Architecture. In: Software Engineering Symposium. Software Engineering Institute, Pittsburgh (1993)
17. Krallmann, H., Schröpfer, C., Stantchev, V., Offermann, P.: Enabling autonomous self-optimization in service-oriented systems. In: Proceedings of The 8th International Workshop on Autonomous Systems - Self Organisation, Management and Control, pp. 127–134. Springer, Berlin (2008)
18. Makripoulias, Y., Makris, C., Panagis, Y., Sakkopoulos, E., Adamopoulou, P., Pontikaki, M., Tsakalidis, A.: Towards Ubiquitous Computing with Quality of Web Service Support. Upgrade, The European Journal for the Informatics Professional VI(5), 29–34 (2005)
19. Papazoglou, M.P., Traverso, P., Dustdar, S., Leymann, F.: Service-oriented computing: State of the art and research challenges. Computer 40(11), 38–45 (2007)
20. Perry, D., Wolf, A.: Foundations for the Study of Software Architecture. ACM SIGSOFT Software Engineering Notes 17(4), 40 (1992)
21. Sandberg, W., Daily, B., Egan, M., Stahl, J., Goldman, J., Wiklund, R., Rattner, D.: Deliberate perioperative systems design improves operating room throughput. Anesthesiology 103(2), 406–418 (2005)
22. Stantchev, V.: Effects of Replication on Web Service Performance in WebSphere. Icsi tech report 2008-03, International Computer Science Institute, Berkeley, California 94704, USA (2008)
23. Stantchev, V.: Enhancing health care services with mixed reality systems. In: Engineering of Mixed Reality Systems. Springer, Berlin (to appear, 2009)
24. Stantchev, V., Hoang, T.D., Schulz, T., Ratchinski, I.: Optimizing clinical processes with position-sensing. IT Professional 10(2), 31–37 (2008)

25. Stantchev, V., Malek, M.: Architectural Translucency in Service-oriented Architectures. IEE Proceedings - Software 153(1), 31–37 (2006)
26. Stantchev, V., Malek, M.: Translucent replication for service level assurance. In: Dong, Paul, Zhang (eds.) High Assurance Service Computing, pp. 1–18. Springer, Heidelberg (2009) ISBN: 978-0-387-87657-3
27. Tokairin, Y., Yamanaka, K., Takahashi, H., Suganuma, T., Shiratori, N.: An effective qos control scheme for ubiquitous services based on context information management. CEC-EEE, 619–625 (2007), http://doi.ieeecomputersociety.org/10.1109/CEC-EEE.2007.20
28. Yau, S., Wang, Y., Huang, D., In, H.: Situation-aware contract specification language for middleware for ubiquitous computing. In: Yau, S., Wang, Y., Huang, D., In, H. (eds.) Proceedings of the Ninth IEEE Workshop on Future Trends of Distributed Computing Systems. FTDCS 2003, May 28-30, pp. 93–99. IEEE Computer Society Press, Los Alamitos (2003)
29. Youssef, M., Agrawala, A.K.: Handling samples correlation in the horus system. In: INFOCOM (2004)

A healthcare real-time monitoring system for multiple sensors data collection and correlation

Luigi Romano, Luigi Coppolino, Ivano Alessandro Elia, and Gianluigi Spagnuolo

Abstract Telecare and home healthcare services are an increasing healthcare research sector. In this field a novel approach is based on wearable sensor devices which provide a user-friendly acquisition of vital signs and allow the implementation of pervasive and continuous healthcare monitoring systems. The amazing amount of data continuously provided by sensors, poses challenging issues to the systems which are in charge of their collection and processing.
In this paper we present a platform allowing real-time monitoring of biomedical and environmental parameters able to collect, store and process data gathered from a wide variety of sources. The platform presents a modular architecture that easily allows its extension with additional sensors. The main feature of the proposed system is the adoption of a complex event processor to correlate and manage events extracted from the collected data. This allows a wider and more accurate knowledge of the patient's health status. A proof-of-concept implementation of the proposed platform has been realized and its implementation is detailed in this paper.[1]

Luigi Romano, Luigi Coppolino, Ivano Alessandro Elia
Dipartimento per le Tecnologie(DiT), Università degli Studi di Napoli "Parthenope", Centro Direzionale di Napoli, Isola C4, 80143 Napoli, Italy
e-mail: `luigi.romano, luigi.coppolino, ivano.elia@uniparthenope.it`

Luigi Romano, Luigi Coppolino, Gianluigi Spagnuolo
CINI - ITeM Laboratory, Napoli, Italy 80126
e-mail: `gianluigi.spagnuolo@gmail.com`

Luigi Romano
Institute for High Performance Computing and Networking (ICAR) of the Italian National Research Council (CNR), Italy

[1] The research leading to these results has received funding from the the Regione Campania within the framework of the PHDS Project. Authors wish to thank the technical staff of all partners participating in the PHDS project (namely: Synclab, Intesis , HAL Software srl, CINI, and ICAR): their contribution was key for properly steering the direction of the research activity, as well as for improving the quality of this paper.

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 455–464.
springerlink.com

1 Introduction

Telemedicine and Telecare use advanced telecommunication technologies to exchange health information and provide healthcare services across geographic, time, social and cultural barriers.[7] [8]
The healthcare industry is facing various challenges that include increasing costs, the incidence of medical errors and the lack of qualified medical workers. For these reasons there is the need to provide a better quality of service using limited financial and human resources. The use of Telemedicine technologies and in particular of ”Pervasive Healthcare” is considered a promising opportunity to address this kind of issues. Pervasive Healthcare or ”healthcare to anyone, anytime, and anywhere” is an approach that can remove locational, time and other restraints and in this way increase both the coverage and the quality of the healthcare services and so help reduce costs.
Many type of powerful medical applications can be developed in the field of pervasive healthcare: miniaturized biological sensors can be connected though increasingly widespread Wireless LANs and personal area networks, along with a wide range of different environmental sensors, and data gathered can be combined with the help of middleware for a wide range of possible medical purpose.[9]
Despite these opportunities there are still challenges to the wide acceptance and diffusion of these technologies. One of the main limit for this kind of applications is the lack of standardization resulting in the impossibility of interoperation for applications developed by different producers. Another problem is related with middleware capability of handling the possibly huge amount of data gathered from the sensors and the addition of new sensors to the systems.[7] [6]

This paper presents a proposal for an extensible medical application platform for real-time monitoring and correlation of biomedical and environmental parameters.
The remainder of the paper is organized as follows. Sect. 2 is an overview of monitoring systems for telemedicine, along with description of related works and of the tools used in our platform. In Sect. 3 the platform's architecture is presented. The section focuses on the data correlation capability and on the extensibility of the proposed platform. In Sect. 4 details are given about the actual implementation of a prototype.

2 Background and related work

Monitoring systems for telemedicine. Early signs of disease can only be detected by hospital examinations but it's difficult to perform screening tests very frequently. Continuous monitoring of patient's health status is a very powerful diagnostic tool for early detection of diseases. An approach like this can allow immediate intervention, before the appearance of subjective symptoms, which is a key feature to decrease related costs.

This kind of monitoring is very suitable for the assistance of elderly patients at home, and this is particularly interesting in advanced countries where the incidence of elderly people living alone is increasing.
Another important filed of application is for the monitoring of patients in postoperative care, in chronic condition or that undergo supervised recovery [9].
As an example of existing implementations of this kind of technologies we here briefly point at two monitoring systems for telemedicine:

- In [10] is presented a system architecture proposal named Ubimon, for an ubiquitous monitoring environment for wearable and implantable sensors based on body sensor network (BSN). It addresses issues related to the use of wearable or implantable sensors for distributed mobile monitoring. A prototype was developed for patients with ischaemic and arrhythmic heart disease.
- In [11] is presented an open platform for the remote monitoring of patients developed by IBM named Persona Care Connect (PCC). It aims at giving caregivers timely access to a patient's health status, and the possibility to provide patients with appropriate preventive interventions, helping to avoid hospitalization and to improve the patient's quality of care and quality of life. The platform contains core functions needed for the collection of biomedical sensor data and stores the data on a server for long-term persistence and analysis. The system is focused on the architecture's integration, extensibility and interoperability issues and a key goal is the use of standards and open system APIs at as many points as possible.

In both of the former systems the aim is to provide highly trained medical staff with a intuitive and context-aware visualization of the collected data, but no mean is provided for configuration and automatic detection of complex patterns of parameters. No complex detection and reaction is possible without the presence of a human operator visualizing the data.

Wearable sensors. Remote and continuous health monitoring systems require sensors easy to use and provided with comfortable sensing interfaces. Measurements have to be carried on in a "natural" environment and while the patient performs his daily activities without interfering with them. A class of sensors that fits for this purpose is that of wearable sensors that are non-invasive devices for the monitoring of biological parameters. Wearable sensors are either supported directly on the human body or a piece of clothing, and are designed to be worn in comfort enabling prolonged use [12].

In our prototype a wearable health monitoring system named WEALTHY is used. It is based on a textile wearable interface implemented by integrating sensors, electrodes, and connections in fabric form provided with signal processing features, and telecommunication systems. WEALTHY is able to acquire simultaneously several biomedical signals: respiration, electrocardiogram, activity, temperature and oxygen saturation. Sensors, electrodes and connections are realized with conductive and piezoresistive yarns. The main functions of the wearable modules are sensing, signal conditioning, signal processing, and data transmission. The sensing module consists of a patented interface connected with a device called the Portable Patient Unit (PPU), where local processing as well as communication with the network is performed. A miniaturized shortrange wireless system is embedded in and can be used to transfer signals to a PC, PDA, or mobile phone. [5]

Complex-Event-Processing. Complex-Event-Processing is a powerful computational paradigm which ideally suits for real time processing of continuous data streams to enable applications to monitor multiple streams of event data. In Complex Event Processors (CEP) sensory information is mapped as event data, or event attributes, and encapsulated in messages that travel over a digital nervous system, constituting an electronic messaging infrastructure. This technology allows the extraction of an higher level knowledge from situational information by means of correlation of events from multiple data stream sources.
CEP are based on continuous queries —i.e. queries that are evaluated continuously as data streams continue to arrive— that are performed on the data flows and trigger reactions to the matching of specific complex event patterns. At the core of a complex event processing system there is a complex event processor, that is the interpreter of an event-driven processing language which allows one to describe the complex event to be analyzed as a space-temporal relationship among elementary events.

3 System design

This section provides an overview of the proposed platform. We focused our efforts on two themes: extensibility and data correlation. Our system can continuously monitor a wide variety of sensors, perform a cross-device correlation between data collected and react in a proper way when significant changes in functional health status are observed. The considerable amount of data collected by the sensors will allow quantitative analysis of various conditions and patterns.

The presented system, as shown in Fig.1, includes seven components deployed across two main elements: Home Station and Event Server.

3.1 Home Station

In addition to collect all the sensor data, the Home Station sends the normalized data to the Event Server via Internet. It is also designed to detect some configurable events and act accordingly. Home Station is structured in three layers: data acquisition, data normalization and local data processing

Data Acquisition. The Data Acquisition layer presents a modular architecture that easily allows its extension with additional sensors. Every sensor data stream format

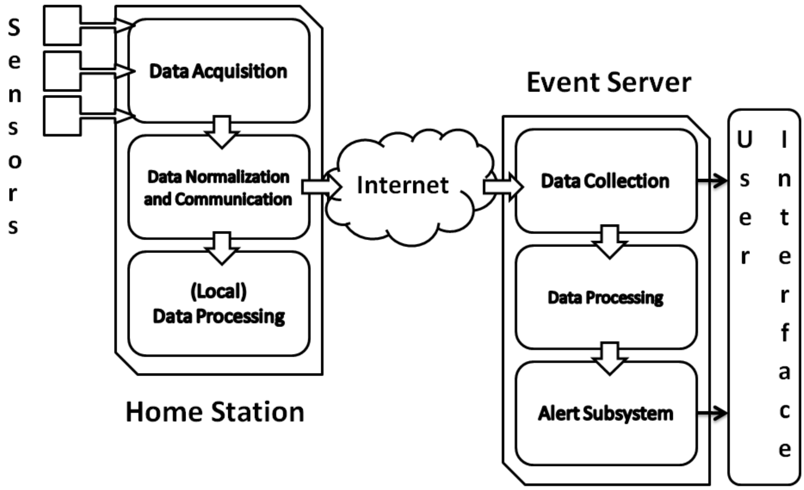


Fig. 1: System's Architecture.

is handled through an adapter that provides features for data parsing. In this way every kind of additional sensor can be easily plugged in the system only generating a new adapter with the specific data parsing capability. This sensor adapters are then connected to the Data Acquisition layer through a plug-in framework. They are specific to each type of sensor device and gather incoming data stream; each adapter implements proper methods to acquire and parse measured data.
The collected data are parsed and stored in a common, consistent format, independent from the sensor type. This layer also handle the initialization, configuration, and synchronization of the sensor devices.

Data Normalization and communication. The foremost task of the Home Station communication system is to transmit reliable medical data through a secure network from the Home Station device to the remote server and to handle replies from the medical staff to the user.
The health monitoring devices produces an amazing amount of network traffic depending on the number and frequency of vital signs monitored. The network performance can be improved by sending minimal real-time data, such as differential signals or changes since the last transmission [6]. The traffic generated can be finally compressed.
The delivery of medical information could be negatively affected by the failure of the network [4]. To address this issue a network monitor has been implemented to monitors the state of the network connection between Home Station and Event Server. The impact of network failures can be minimized by caching data for later transmission.

Local Data Processing. The core of the Home Station is the Local Data Processor that presents a minimal event handler based on a configurable monitoring system are designed. For each monitored vital sign multiple thresholds and a set of actions [6]; if the received measurement are out of range an emergency alert can be notified. For this purpose this layer provide an audio and graphical user interface that can be used for the first medical aid.

3.2 *Event Server*

Our challenge was to implement an Alarms and Event Server that correlate and manage events extracted from the collected data. In this way if there is a significant compromise in the patient's vital signs, and event is raised an alarm is sent to the patient's medical staff.
Our system is structured in four layers: data collection, data processing, alert subsystem and user interface. A detailed schema for the Event Server is shown in Fig.2.

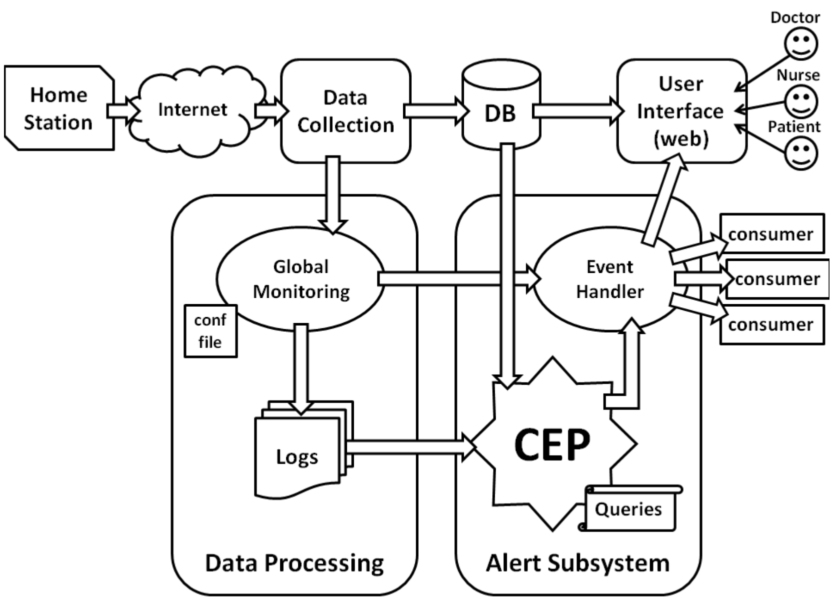


Fig. 2: Event Server Schema.

Data Collection This layer provides a service to collect data in real-time from individual patients. Apart from receiving real-time data from the Home Station via

Internet this layer also store this data in a database which is used by the User Interface for the data visualization and management, and the Complex Event Processor to perform a long-term analysis and detect undesirable patterns. Patient database is a valuable resource to predict potential risk factors through various data correlation techniques.

Data Processing In addition to the Local Data Processing this layer extends the number of detectable events and actions. It also produces a log file used by the Alert Sub-system to perform a interrelationship between multiple vital signs.

Alert Sub-system In addition to simple events monitor, a complex event processor adds the capability to manage and correlate data from the log file and patient database to perform advanced short-term and long-term analysis. In this way events from multiple sources can be correlated to identify problems that previously required operator analysis. To handle outputs received from the CEP, a configurable event handling subsystem is provided. A consumer must be able to receive event from the Event Handler service.
In order to receive events, a client application (consumer) can subscribe for one or more event notification. Users can customize their preferences for receiving event notification messages through a web based user interface.
A message priority based mechanism is used to control the transmission of emergency alert messages onto the network; this feature optimizes bandwidth utilization while minimizing message delay.

User Interface In [3] factors to consider when providing health information are identified as: consistency of health-monitoring information; convenient user interface; universal access and universal design for the disabled and elder persons; a unified health-monitoring Web server for wired Internet and mobile Internet; health-information center accessibility for the doctor nurse, and the disabled and elderly; provision for different formats of health information .
User Interface is provided by a web application that format the patient data according to user profile. Views for doctors, patients and system administrators are offered; the web application dynamically produces pages according who accesses a patient data. For example a doctor will see all the patient information and may alter the configuration files, while a patient will see a limited set of data.
All the collected data, and events, are available in an externally accessible format to allow interoperability and extensibility with other systems. For example, our platform can be seamlessly integrated with an automated prescription process and with the patient electronic medical records. In this way the system has the opportunity to immediately change prescriptions accordingly to the current health status of the patient. Then a message is generated and sent to the Home Station in order to show the updated prescriptions and remind the patients when they need to take medication.

Extensions and modification can be performed easily and with no modification to the core of the system: the addition of a new sensor only implies the generation of a new adapter and an upgrade to the event detection logic is simply performed trough the creation of a new query to be deployed on the CEP.

4 System implementation

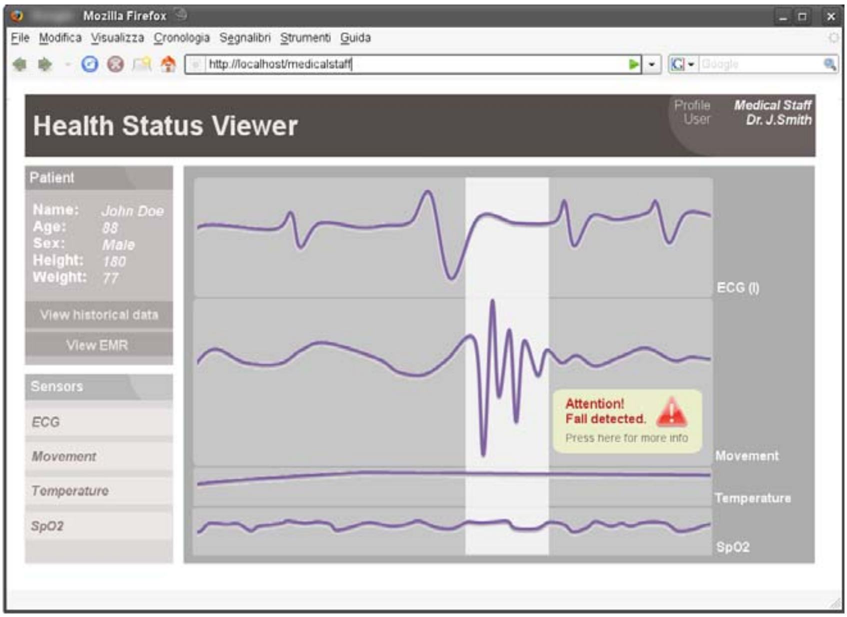


Fig. 3: A screenshot of the platform's user interface.

This section presents the implementation of the proposed system, with particular emphasis on the data correlation. The prototype of the proposed system is implemented using the Ruby language [2] on a Linux box. We chose Ruby because of its greater portability and extensibility.
Preliminary evaluation of the prototype is carried out using WEALTHY [5], a wearable system discussed in Sect. 1. The WEALTHY sensor data are gathered by the Home Station through a bluetooth connection. The specific plug-in has been implemented using BlueZ, the official Linux Bluetooth protocol stack, to acquire data via Bluetooth while the PPU packets parser has been written from scratch.
A minimal Local Data Processing use a configuration file that contains an entry for each measured parameter. Each entry is composed by a pre-configured range and a local action; this action is accomplished if a measured value is out of range. Currently, the action mechanism is limited to transmission of an alarm message delivered to the patient in three different ways: via SMS, via email or through the Home

Station GUI. As future work, we intend to add a more flexible and fine-grained framework to define events and related actions; a domain specific language is also planned.
In the Alert Sub-system, to correlate and manage events extracted from the collected data we have used Borealis [1], a distributed stream processing engine developed at Brandeis University, Brown University, and MIT.
A sample detection that could be performed using the correlation capability of the CEP is the detection of a Tachycardia, in this case the pulse rate exceeds the 110 beats per minute while the oxygen saturation happens to be lower then 89%. As shown in fig.4 the detection of this event can be performed with a query performed on the data streams on the CEP. The two filter blocks select only the anomalous values of oxygen saturation and pulse rate from the corresponding streams and then, in the join block, streams are correlated on the base of their timestamps, when for both parameters, anomalous values are measured within the same time window, alerts are generated on an output stream.

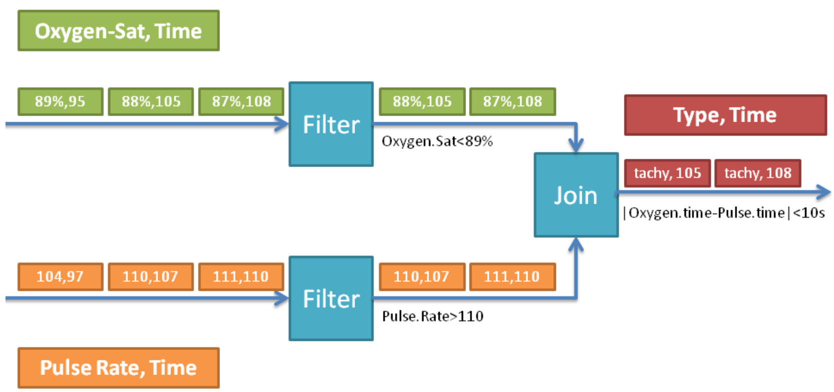


Fig. 4: An example of event detection performed trough the use of CEP.

The prototype user interface is built upon the Ruby on Rails [13] web application framework on Apache/Passenger server. Some features are developed using JavaScript and Adobe Flex technologies. In Fig.3 is shown a monitoring page during a fall detection.

Conclusion

In this paper, we have described a platform to allow real-time monitoring of biomedical and environmental parameters to perform a cross-device correlation and have an accurate knowledge of the patient's health status. We focused our efforts on extensibility and data correlation in order to provide an highly configurable platform.

A prototype of the platform has been implemented using Borealis to correlate data collected from a wearable system.

References

1. Abadi, D.J., Ahmad, Y., Balazinska, M., Cetintemel, U., Cherniack, M., Hwang, J.-H., Lindner, W., Maskey, A.S., Rasin, A., Ryvkina, E., Tatbul, N., Xing, Y., Zdonik, S.: The Design of the Borealis Stream Processing Engine. In: 2nd Biennial Conference on Innovative Data Systems Research (CIDR 2005), Asilomar, CA (2005)
2. Fulton, H.: The ruby way: solutions and techniques in ruby programming. Addison-Wesley Professional, Reading (2006)
3. Kim, Y.B., Yoo, S.K., Kim, D.: Ubiquitous Healthcare Technology and Service. In: Intelligent Paradigms for Assistive and Preventive Healthcare. Springer, Heidelberg (2006)
4. Lee, H., Park, K., Lee, B., Choi, J., Elmasri, R.: Issues in data fusion for healthcare monitoring. In: Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments, Athens, Greece, July 16-18 (2008)
5. Paradiso, R., Loriga, G., Taccini, N.: A wearable health care system based on knitted integrated sensors. IEEE Transactions on Information Technology in Biomedicine 9(3), 337–344 (2005)
6. Varshney, U.: Pervasive healthcare and wireless health monitoring. Mobile Networks and Applications 12(2-3), 113–127 (2007)
7. Alexander, Q., Xiao, Y., Hu, F.: Telemedicine for pervasive healthcare. In: Mobile Telemedicine: A Computing and Networking Perspective. Auerbach Publications, Taylor & Francis, New York (2008)
8. Choi, Y.B., Krause, J.S., Seo, H., Capitan, K.E., Chung, K.: Telemedicine in the USA: standardization through information management and technical applications. IEEE Communications Magazine 44, 41–48 (2006)
9. Ogawa, M., Togawa, T.: The concept of the home health monitoring In: Proceedings. 5th International Workshop on Enterprise Networking and Computing in Healthcare Industry. Healthcom 2003, June 6-7, pp. 71–73 (2003)
10. Ng, J.W.P., Lo, B.P.L., Wells, O., et al.: Ubiquitous monitoring environment for wearable and implantable sensors (UbiMon). In: Proceedings of the 6th International Conference on Ubiquitous Computing (UBICOMP 2004), Nottingham, UK (2004)
11. Blount, M., Batra, V.M., Capella, A.N., Ebling, M., Jerome, W.F., Martin, S.M., Nidd, M., Niemi, M.R., Wright, S.P.: Remote health-care monitoring using Personal Care Connect, pp. 95–114 (2007)
12. Glaros C, Fotiadis: Wearable Devices in Healthcare, pp. 237–264. Springer, Heidelberg (2005)
13. Thomas, D., Hansson, D., Breedt, L., Clark, M., Davidson, J.D., Gehtland, J., Schwarz, A.: Agile Web Development with Rails. Pragmatic Bookshelf (2006)

Personal Health Monitor

Almudena Vicente Tocino, Juan José Andrés Gutiérrez, Iván Álvarez Navia, Francisco J. García Peñalvo, Esteban Pérez Castrejón, and José Rafael García-Bermejo Giner

Abstract This paper introduces the *Monitor de Salud Personal* created by Telefónica I+D in cooperation with the Departamento de Informática y Automática of the University of Salamanca. This project fits within the activities related to the management of a "Digital Home", and it is one of the projects that are currently under development in this field in Telefónica I+D. It is the aim of the project to explore various enabling technologies based on environmental intelligence, by means of which the user interacts with his or her home in various scenarios: home assistance and wellbeing, entertainment, identity management in a home environment, and location management in a home. The *Monitor de Salud Personal* we describe is intended to satisfy the objective of home assistance and wellbeing, and it fits well within *Ambient Assisted Living* Joint Programme of the European community [1].

1 Introduction

Current homes must face an ever-growing number of consumer electronic devices. These are intended for family leisure, and also to help in daily chores. However, most of them work in isolation, and require some knowledge in order to make them work, all the more so if one wants to make full use of their capacities. Nowadays, users demand simplicity, that is to say, they want technology to be a servant of users, and not the opposite. A new concept rises, that of Environmental Intelligence, which is based on the existence of a natural, non-intrusive relationship between people and the technological environment that surrounds them.

Juan José Andrés Gutiérrez · Esteban Pérez Castrejón
Telefónica I+D. e-mail: `jjangu,esteban@tid.es`

Almudena Vicente Tocino · Iván Álvarez Navia · Francisco J. García Peñalvo · José Rafael García-Bermejo Giner
Dpto. Informática y Automática. Universidad de Salamanca
e-mail: `almudenavt@gmail.com`, e-mail: `inavia,fgarcia,coti@usal.es`

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 465–475.
springerlink.com

Environmental Intelligence is about a type of technology that remains transparent for the user, embedded in his or her environment, present whenever it is needed and accessible by means of a simple and natural interaction with the user. This interaction must be addressed to the whole sensorium, must adapt to user, and must be context-sensitive. This means accessing a high quality of information, with customized contents, and it must be available for everyone, in any place and at all times. [2]. *Environmental intelligence* is a concept that was introduced by [3] as a vision of the future, within a near-future horizon, in which people will leave in an environment that surrounds them with intuitive and intelligent systems, embedded in everyday objects that will create an ambience capable of recognizing people, and with capabilities to respond to people's presence and also to assist them.
Ideally, devices would have an autonomous behaviour; the user should not have to worry about them. Everyday objects must incorporate environmental sensors with wireless connections; interfaces should be natural and non-intrusive, and applications should be predictive and endowed with enough intelligence to take correct decisions based on the users' needs. Homes are one of the environments in which Environmental Intelligence makes most sense. The proliferation of domestic devices with the ability to connect, geared towards wellbeing, entertainment, communication, security and assistance, make our homes sites of high technological penetration. However, the relationship between people and technology turns out to be not as friendly as one might expect.
These are the goals of the project *Hogar Asistido - Ambientes Inteligentes en el Hogar*. Its main aim is to explore various technologies that enable the use of environmental intelligence in a home setting, by offering technical components in a range of scenarios: health and wellbeing, home assistance, entertainment, identity management in a home environment, and location within homes.

Fig. 1 shows an overview of this project. A Home consists of heterogeneous devices with communications capabilities (WiFi, Bluetooth, etc.), that is to say, a wide ecosystem of devices with sensors that enable them to identify and locate users (RFID readers), to measure vital constants (pulsioxymeter, scales) and to obtain environmental data (meteorological station). In order to incorporate these devices and of course to put their functionalities at the disposal of a Home, a new concept must be introduced: that of **service discovery**. This means that the functionality of each device must be perfectly defined and published, so that other systems may access the services they offer.
Once devices are able to put their services at the disposal of the Home, one needs another component that will organize and manage the intelligence provided by the system. That is the main task of the global platform (Fig. 1). This system is in charge of scanning the home network and of discovering the devices found in that network. Additionally, it controls the devices by invoking the services they provide. This control capability, along with the intelligence of the platform, can analyze events and context information provided by the sensor in the home environment; thus it can carry out actions based on available information. This platform is intended to act in order to assist the user, and can be an important help in certain daily activities.
This is the context in which the *Monitor de Salud Personal* project is being

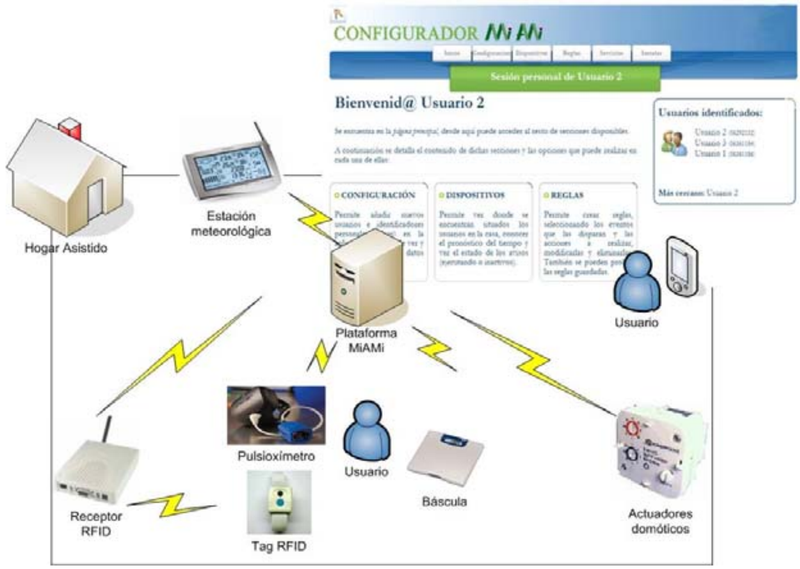


Fig. 1: Overview of the “Hogar Asistido - Ambientes Inteligentes en el Hogar” project.

developed, in the framework of a cooperation agreement between the Departamento de Informática y Automática de la Universidad de Salamanca and Telefónica I+D, which is placed in the Parque Tecnológico de Boecillo. According to previsions, by 2026 more than 21% of Spain’s population will be more than 65 years old. Besides, 32% of them will suffer some kind of disability [4]. This project is aimed to fulfil the need for intelligent assistance in home environments. With this objective in mind, the system makes use of a mobile sanitary device that measures several vital constants of the users; it is called pulsioxymeter. The scope of this project is not restricted to a home environment; rather, it is considered as a personal *gateway* that extends the concept of remote assistance in the home to a wider-ranging mobile remote assistance. Since these devices monitor the vital constants of patients, it is essential for sanitary personnel to be able to access their values.
Telemedicine or *e-health* is an emerging field, that arises from the intersection of medical computing with public and private healthcare, and which is related to the services and healthcare information that are being offered through Internet and related technologies. In a broader sense, the term includes not just the technical development and new business models, but also a change in our way of thinking, and the compromise of enhancing health both locally and in the widest possible scope by means of information and communications technologies. [5]. The following sections introduce the most relevant aspects of the *Monitor de salud personal*project. Section 2 offers an overview of its scope. Afterwards Section 3 shows the architecture of the system that has been developed. The interactions between the various

actors that comprise the scenario of this project and the diverse components used to fulfil each goal are detailed in Section 4. Finally, Section 5 offers the project's conclusions and the future developments.

2 Personal Health Monitor

The proliferation of *wearable* devices [5,7,8] with sensors that measure certain vital parameters, as well as the enhancement of communications technologies constitute a starting point in the research for telemedicine systems [9,10], which allow sanitary personnel to perform a remote follow-up of patients. In this project, a Bluetooth pulsioxymeter by Nonin [11] makes it possible to measure cardiac frequency, oxygen saturation and pletismography in the patient. On the other hand, when the user gets home with his or her pulsioxymeter, it is to be desired that the device will integrate transparently in the intelligent environment. Hence, one can anticipate the following challenges in a home environment: (1) standardizing device access and (2) facilitating the introduction of new device with a minimal intervention on the user's part. This makes it necessary for all devices to put their services at the disposal of the home in a standard way. In order to achieve this in the *Monitor de salud personal*, one must build an overlay that will allow the pulsioxymeter to announce its services to the home network. This is done by means of a PDA, which exports the pulsioxymeter's services. Thus, when the user acquires this device (Pulsioxymeter + PDA) and takes it home, the platform will discover it automatically, with no intervention on the user's part. The platform will be able to access all of the functionality published by the pulsioxymeter.

At this time, the patient can do a personal follow-up of all vital constants measured by the pulsioxymeter; this information is restricted to the home environment. However, by making use of wireless communication technologies (WiFi, GPRS or UMTS) one can extend this scenario in such a way that the measurements of vital constants of said patient would be exported to the sanitary environment. In this way, the user would no have to travel to the health centre. This avoids waiting times, and can obviate a large number of consultations whose only purpose is to monitor certain vital constants.

Briefly, the most relevant functional aspects of this research are as follows:

- To monitor vital constants of users in their homes.
- To remotely consult vital constants from outside their homes.
- An alert mechanism that can act in case of anomalous situations; should one arise, alerts will be sent as short messages to mobile phones, and also as e-mail messages.

Fig. 2 shows the layout of the various physical elements, the placement of software components and the individuals involved in a remote-assistance scenario. One can appreciate several facts:

- There exist a remarkable variety of devices of diverse nature (pulsioxymeter, PDA, gateway) with well-defined functions within the project. This diversity of

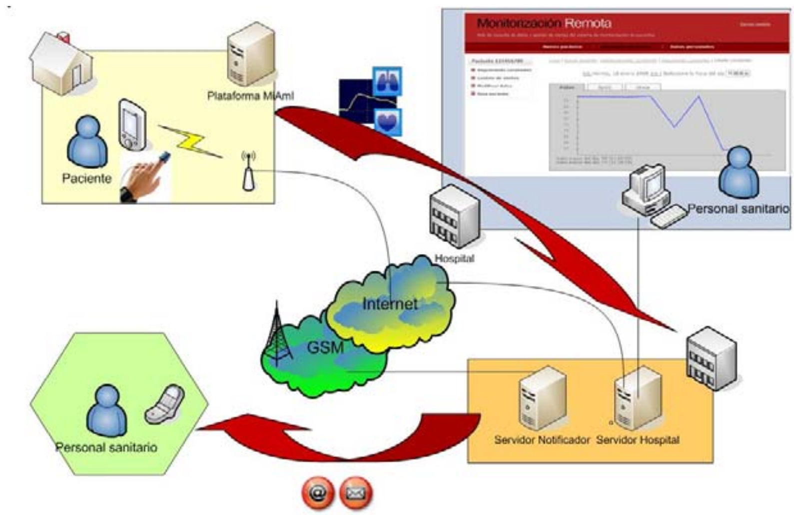


Fig. 2: Scenario for the project “Monitor de Salud Personal”.

devices, each with its own software characteristics, influences the selection of the technologies used to implement their functionality, which are equally quite variable in nature.

- The need arises to integrate the various types of communication technologies involved: **UPnP** [11], **Bluetooth**, **WiFi**, **Cable**, **ADSL** or **GSM**. This wide spectrum plays a very important role in the scenario, since it makes it possible to exchange data and to invoke services.

One of the most significant features of the system’s architecture is that it consists of various independent services and components. Basically, services expose an interface to which all incoming messages are sent. Being a distributed system, the physical architecture of the system consists of various computational nodes that host each of the components that make up the system. Sanitary devices are connected to mobile devices (mobile phones/PDAs) by means of **Bluetooth**. These, in turn, send data through **WiFi** towards a Web access platform, in which medical information is stored. Thus it is possible to communicate the patient with sanitary personnel; this makes it easier for the sanitary personnel to evaluate the patient, to take decisions and to generate an adequate response. Besides, data and communications are guaranteed to be secure. It is possible to create graphics with historic data and health states. One can also send alarms, which let the medical personnel know immediately about health situations that require immediate action. Fig. 2 shows the following computational nodes:

- Pulsioxymeter. This is a mobile device that sends vital constant measurements through **Bluetooth** to the PDA.

- PDA. A mobile device with various functionalities. The PDA sets up a direct communication with the Pulsioxymeter in order to recover measurements of vital constants of the patients. Besides, it acts as a bridge between the entre Pulsioxymeter, in the home environment, and the sanitary personnel. This is done by exporting the values of vital constants through **UPnP** to the home network, and **Servicios Web** [13] to the health centre.
- Home node. This system is in charge of discovering any services exported by the Pulsioxymeter and the PDA through **UPnP**, thus allowing the integration of new devices into the home network.
- Data Server. This node stores a historic collection that contains all of the measurements of patients.
- Notification Server and **SMS** gateway. This node manages the broadcasting of short text messages, by means of the **SMS** gateway and e-mail.
- Web application server. This node processes any petitions submitted by sanitary personnel to the web application intended to follow-up vital constants. The node codifies the proper responses.
- Client host of sanitary personnel. This node submits requests to the Web application server, thus allowing sanitary personnel to carry out a follow-up of patients that belong to the Remote Assistance Program.

3 System architecture

Once the computational nodes have been identified, on can realize a deployment model for the physical architecture. Fig. 3 shows the connections between the computational nodes mentioned above. It is based on the TCP/IP protocol; there are different *software* components in each node.

Description of the various component used in this project:

- DllUPnP: this library produces a **UPnP** overlay of the Pulsioxymeter. The **UPnP** overlay exports three services that are measured by the Pulsioxymeter: pulse, oxygen saturation and pletismographic wave.
- MonitorConstantesPDA: this application is in charge of communicating with the Pulsioxymeter and of graphing obtained data. This component is deployed in the patient's PDA and makes use of the DllUPnP methods in order to export measurement towards the home **UPnP** networks; it also invokes the **Servicio Web** called ServicioGuarda_RecuperaConstantes, which is in charge of keeping a historic file of the vital constants of patients, in order to export said measurement towards the health centre.
- ServicioGuarda_RecuperaConstantes: As explained above, this component offers an interface for each of the operations it implements. Client processes need only know the interfaces and purposes of each operation. The client is thus totally isolated from any implementation details, which could be modified at any time without affecting the system.

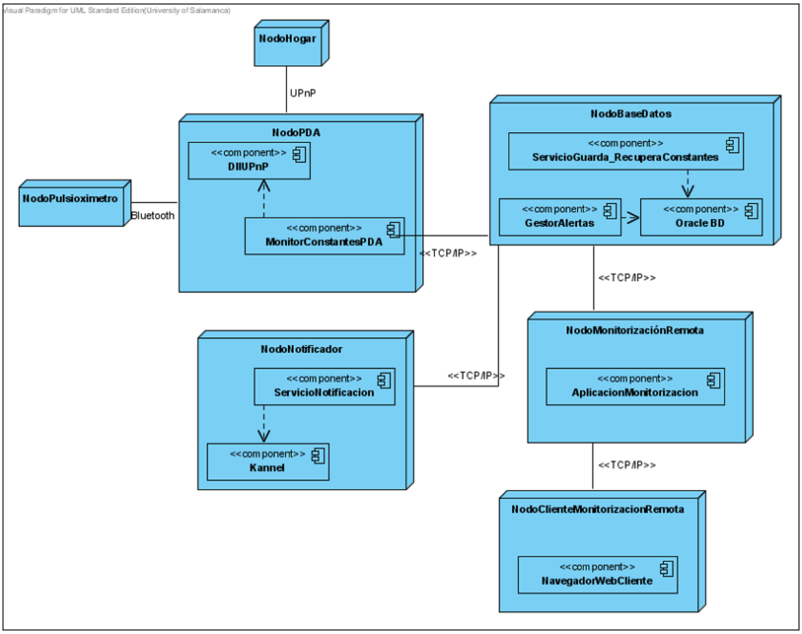


Fig. 3: Physical architecture. Deployment model.

- ServicioNotificacion: **Servicio Web** in charge of processing alerts as short text messages sent to mobile phones by means of the **Kannel** gateway; it also sends e-mail. The purpose of this service is analogous to that of the previous one. It is a fully independent and reusable component, which offers an interface that lets potential clients know all available functionality.
- AplicacionMonitorizacion: Web application accessible for sanitary personnel in order to realize a follow-up of the patients' constants. In order to access the patient's constants, it becomes a client of the **Servicio Web** ServicioGuarda_RecuperaConstantes. External systems description:
- NavegadorWebCliente: This represents the browser from which sanitary personnel accesses the Web application AplicacionMonitorizacion, which makes it possible to realize a follow-up of the patients.
- Kannel: SMS gateway used to send short text messages to mobile phones.
- Oracle BD: Database management system used for persistent storage of information.

4 Case study

Paula is a young 25-year-old woman with a chronic obstructive pulmonary disease (EPOC) that forces her to continuously monitor oxygen saturation (SpO2) in her blood. Ever since she was diagnosed, she has been continuously monitored by means of a device that explores her vital constants, thus trying to anticipate any crisis. Today Paula is returning home. She is a participant in a new “Programa de Telemonitorizaci0́n de constantes vitales” (Remote monitoring of vital constants program). Her doctor, Lucas, has given a health monitor to her. He has explained the way it works: the Pulsioxymeter will monitor her constants and send her data to the PDA through **Bluetooth**; the PDA, in turn, will send her data to hospital through internet, by means of the **WiFi** router that Paula has at home. From hospital Lucas will be able to check the evolution of Paula’s vital constants. He can also define alert threshold values, and in that way he can react proactively against a possible respiratory crisis.
Lucas inscribes Paula in the “Programa de Telemonitorización de constantes vitales”, through a Web application in which he can register new patients. Then he boots the PDA application in order to configure the parameters:

- User name and password, so that Paula’s measurements can be sent to the external server. This can be done from Preferences in the PDA’s application.
- URL of the Web Service for data storage in the hospital’s server, and server certificate.

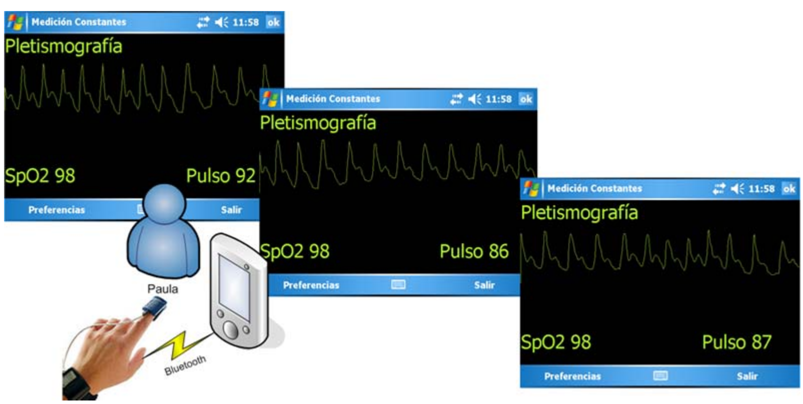


Fig. 4: Screen capture of vital constants evolution in the PDA.

Lucas defines an alert that is associated to the SpO2 level in the web application. The alert will run when SpO2 is less than 90%, since this parameter indicates that Paula could eventually experience a respiratory crisis. This value, which is relatively high, can be considered as a safety margin, since the alarm will run before

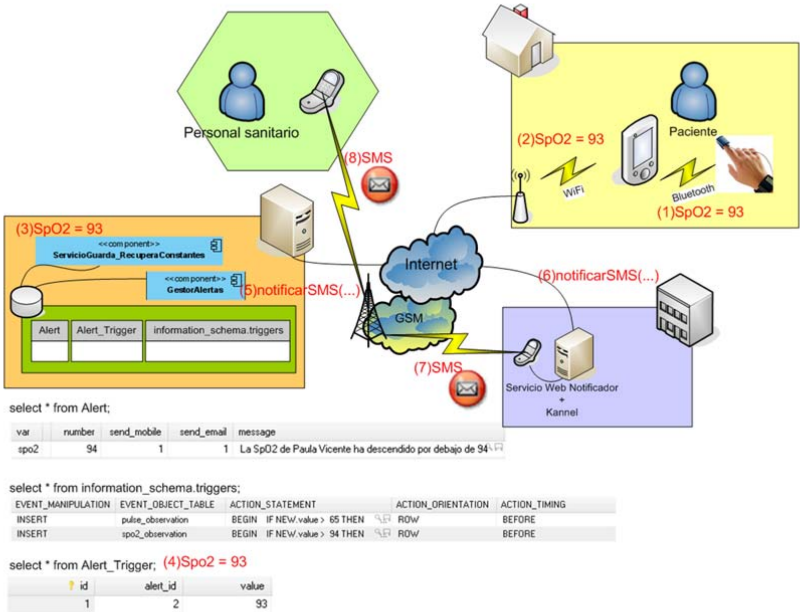


Fig. 5: Execution of an alert when SpO2 is less than 90.

a greater decrease in saturation is observed. On the other hand, one can also define a combined value of the absolute value of saturation and saturation descent as an indication of the onset of a potentially dangerous process.
SpO2 measurement suffers from errors and noise, due to the relative movement between patient and sensor. This necessitates the use of algorithms to reduce false alarms. These algorithms can be simply based on hysteresis, and one can also use more complex classifiers to evaluate risk situations. Paula can be constantly monitored both at home, through WiFi, and when she is out, by means of the 3G networks, if the PDA that is used admits these communications technologies. The alert will be sent both through e-mail and by means of an SMS to Lucas' mobile phone. The interaction between Paula and the Monitoring System is as simple and automatic as possible; technology is abstracted as much as possible. When the user starts up the application, it sets up communications in order to receive the measurements of the patients' vital constants, and it updates the **UPnP** variables returned by the Pulsioxymeter services in the home environment. Besides, it establishes a connection with the proper **Web Service** in order to export the patients' vital constants to the sanitary environment. At a given time, the measurement of Paula's oxygen saturation starts decreasing and reaches a minimum value of 89%. This triggers the alert mechanism shown in Fig. 5, which produces the sending of an SMS and an e-mail to the proper sanitary personnel. When Lucas receives the messages in his own mobile phone and through e-mail, he can act accordingly. Anyway, Lucas can always check the evolution of Paula's vital constants from the web application. By

means of the program, he can select the date and time to be consulted, and a graphic is produced that shows the evolution of a given vital constant during that period of time. Consequently, the system interacts with sanitary personnel in three possible ways: by sending an SMS, by sending e-mail, and by means of a remote follow-up web application.

5 Conclusions

Terms like remote assistance or remote medicine are becoming commonplace in society and in technology, and communications experience enhancements continuously. Some of these efforts intend to enhance medical attention given to patients, by offering them interfaces that are accessible, transparent and, most importantly, intelligent, since they anticipate the user's decisions. A number of projects related to remote assistance or remote medicine are being carried out with the patient's wellbeing in mind. This is a sample of the evolution of technologies, and of the fact that they can and should be at the disposal of users in all possible aspects of their daily lives, with an special interest in their health.

A personal health system has been designed an implemented. For this we have explored and used a wide range of technologies. The result is a system that shows that technologies can be adapted to people, being the base for the development and deployment of future projects intended to enhance wellbeing.

The philosophy behind this project is that of allowing a friendly interaction when patients relatively mature in years, which can possibly have no technical expertise, use a technologically complex system. This is done by means of wireless technologies, which give the user a greater mobility without losing the capability of continuous monitoring. The project is specially intended for chronic patients who need constant monitoring of their illness, since it allows sanitary personnel to be informed at all times of any anomalous situation that may arise. The system trite to reduced to a minimum the amount of intervention on the patient's part; in fact the patient must just place the sensors, and the system will carry out the transmission of data by any means available to the home environment, and to the hospital server. Data transmission to the home environment is a remarkable aspect of the project, as well as the use of service-discovery technologies, that are used to integrate the diverse devices.

Several lines of research remain open for exploration: (1) To incorporate biomedical devices based on the Zigbee technology, (2) to participate in standardization organisms in order to produce templates for UPnP medical devices, (3) to cooperate with health professional in order to achieve a deeper understanding of medical applications, (4) to insure privacy and to guarantee the fulfilling of international regulations regarding health safety and privacy, (5) to enhance the platform's learning system.

References

1. Ambient Assisted Living Joint Programme, http://www.aal-europe.eu
2. Weber, W., Rabaey, J.M., Aarts, E.: Ambient Intelligence (2005)
3. Information Society Technologies Advisory Group, http://cordis.europa.eu/ist/istag.htm.
4. Atención a las personas en situación de dependencia en España, http://www.imsersomayores.csic.es/documentacion/documentos/libro-blanco-dependencia.html
5. Eysenbach, G.: What is e-health? J. Med. Internet. Res. 3(2), e20 (2001), http://www.jmir.org/2001/2/e20/
6. Park, S., Jayaraman, S.: Enhancing the Quality of Life Through Wearable Technology. IEEE Engineering in medicine and biology magazine (2003)
7. Paradiso, R.: Wearable Health Care System for Vital Signs Monitoring. In: Proc. of the 4th Annual IEEE Conf. on Information Technology Applications in Biomedicine, UK (2003)
8. Anliker, U., Ward, J.A., et al.: AMON: A Wearable Multiparameter Medical Monitoring and Alert System. IEEE Transactions on Information Technology in Biomedicine 8(4) (2004)
9. Oliver, N., Flores-Mangas, F.: HealthGear: A Real-time Wearable System for Monitoring and Analyzing Physiological Signals. In: Proceedings of Int. Conf. on Body Sensor Networks (BSN 2006), April 2006. MIT, Boston (2006)
10. Konstantas, D., Halteren, A.V., Bults, R., Wac, K., Widya, I., Dokovsky, N., Koprinkov, G., Jones, V., Herzog, R.: Mobile patient monitoring: the mobihealth system. In: Proc. Int. Conf. on Medical and Care Compunetics, NCC 2004 (2004)
11. Nonin 4100 Bluetooth Oximeter, http://www.nonin.com
12. Web Services Architecture, http://www.w3.org/TR/ws-arch/
13. UPnP Forum, http://www.upnp.org/

Basic Actions in Art Study: Toward an Intelligent Environment for Human Creative Interaction

Fuminori Akiba

Abstract The aim of this paper is to make clear basic actions of art researchers on which a new intelligent environment can be built. To achieve this aim, firstly, with the help of philosophical aesthetics, I make clear two ways of thinking which underlie art study –genealogical thinking and classificatory thinking (Ch.2). Then, from these two ways of thinking I make explicit basic actions of professional art researchers. From genealogical thinking I draw out basic actions of collecting and relating (Ch.3), behind the classificatory thinking I find basic actions of extracting, assigning, and reading (Ch.4). Furthermore, in order to encourage beginners to do these actions, I try to suggest two directions of possible intelligent environment. For genealogical actions I suggest an intelligent environment that makes it possible to transform a pictorial image and to collect images that are historically relevant to the image. For classificatory actions I suggest an intelligent environment that helps beginners to read the hidden meaning of reviewer's comments.

1 Introduction

How can we facilitate human interaction? To give an answer to this question I pay attention to human interaction mediated via artworks. Artworks themselves are products of creativity; in addition, artworks have the power to stimulate viewers' latent creativity and to turn viewers to their own creative action. Even conversation between viewers about artworks is the first step to their own creative interaction. So if there is an intelligent environment which encourages viewers' active participation in artworks, this environment can also facilitate human creative interaction to the future.

Department of Systems and Social Informatics,
Graduate School of Information Science,
Nagoya University, Furo-cho, Chikusa-ku, 464-8601, Japan
e-mail: akibaf@is.nagoya-u.ac.jp

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 477–485.
springerlink.com

We already have many tools for "art study." For example, iconclass, web gallery of art, etc. However, they are completely useless because they are tools for specialists. Certainly they contain the results of previous art study, but they are only waiting for the time when people input keywords suitable for this previous study. In front of these tools, we mistakenly come to believe that understanding art is equivalent to cramming innumerable knowledge about art history into our heads. However, just as scientific research begins with basic actions, for example, observing the behavior of objects, so art study must begin with some basic actions, not with knowledge. So if there is an intelligent environment through which beginner art viewers can imitate basic actions of professional art researchers, it might encourage beginners to participate in creative interaction mediated via artworks. But what are basic actions of art researchers? The aim of this paper is to make clear the basic actions of art researchers on which a new intelligent environment can be built.

To achieve this aim, in the next chapter, with the help of philosophical aesthetics, I will make clear two ways of thinking which underlie art study. One is genealogical thinking, which I will explore in Ch.3, and the other is classificatory thinking, which I will discuss in Ch.4. In these two chapters I would like to point out the basic actions which underlie the two ways of thinking in art study.

2 Two Ways of Thinking Underlie Art Study

Philosophers have revealed our prejudices about art and proposed alternative ideas. In this chapter, with the help of philosophers, I try to make clear two ways of thinking in art study.

2.1 Genealogical Thinking

In his article "Art as Action", Gregory Currie announced his idea about artworks and the mission of art study (Currie 1989, see also Levinson 1980). At first he pointed out our prejudice that the essence of an artwork is in its structure. Contrary to our prejudice, he alternatively proposed the idea that the essence of an artwork is in the process in which the artist found the structure of the artwork. In order to explain this idea, Currie took the following example; if two artworks, which have the same structure, are made through two different processes, we must think they are two different artworks. Currie called this process "the heuristic path" and formalized an artwork as an event; *the discovering of S via heuristic path H by x at t*. This means that an artist (x) discovered the structure of his artwork (S) via the heuristic path (H) at the time t.

From this formula, he drew out the mission of art study. The mission of art study is to reveal "the way in which the artist arrived at the final product. He [for example, art critics] must show us in what ways the artists drew on existing works for his inspiration, and how far that product was the result of an original conception. He must show us what problems the artist had to resolve in order to achieve his end result, and how he resolved them. His job, in other words, is to trace, as closely as he can, the artist's heuristic path to the final product" (Currie 1989, 114). Here we can see a genealogical thinking.

2.2 *Classificatory Thinking*

In addition to the study of the art-making process, there is another research field in art study. It is the study of the art-propagating process. In order to explain this propagating process, Charles Nussbaum borrowed Ruth Garrett Millikan's notion of proper function(Nussbaum 2003, see also Dutton 2006, Cohen 2004) . According to Millikan and her interpreter (Millikan 1984, Maeda 2004), the proper function of a thing comes from a reciprocal relationship between the function of a thing and a system as a whole. The function of a thing which contributes to the survival of a system is, at the same time, selected by the system and survives. Millikan applied the notion of proper function to natural language, especially to declarative sentences.

Nussbaum used her notion and said, "Like declarative sentences, if musical symbols ceased doing whatever it is they do for producers and users, if they ceased to fulfill their derived proper functions, they would not continue to be produced and would stop proliferating" (Nussbaum 2003, 277). This is a remark about musical symbols, that is, about various forms of records from musical scores to digitally recorded data,, but we can apply it to artworks and their records in general, for example, various type of sentences such as comments about paintings, reviews on exhibitions, and various type of recorded images of paintings such as portfolios, catalogues, and moreover, various type of archival images produced by information technologies such as X-ray photograph, cross-sections, etc. If artworks and their records ceased to fulfill their proper functions, they would not continue to be produced or reproduced and would stop proliferating. Contrary to our prejudice, production and reproduction of artworks and their records have nothing to do with the intention of the author.

If so, there would be another mission of art study. The mission of art study would be to make clear proper functions of artworks and their records. This raises the questions: What kind of function should each artwork and its record fulfill in order to make both an art system and the artwork itself survivable? What kind of function of an artwork do viewers expect? Here we can see a classificatory thinking.

3 Basic Actions in Genealogical Research

Here I try to make explicit basic actions of art researchers when they begin to make genealogy. To achieve this aim, I refer to an article which resolved the genealogical problem of Rubens's paintings (Olivier et al. 2005).

3.1 How to Solve the Genealogical Question

The question in the article is: it is said that the painting "The Judgment of Paris" by Rubens (London, National Gallery of Art) was imitated by his studio. This smaller studio work is now in the Dresden, Gemaelde Gallery. However, the painting imitated by his studio differs in a number of ways from the National Gallery's paintings by Rubens, for example, the number of Cupids, the postures of Mercury and Paris, the existence of three satyrs. If we assume that a studio work made by assistants should faithfully imitate their master's model, these differences are strange. If so, is the genealogical assumption that the studio work in Dresden was imitated directly after the Rubens's work in London true? To give this question an answer, the researchers took the following procedures. 1) They compared the painting by Rubens with the painting by his studio, and 2) found an unnatural spot which suggests something repainted in Rubens's painting, 3) took X-ray photographs of Rubens's painting, and 4) beneath the surface of Rubens's painting they found another painted layer which contained the initial composition quite similar to the studio painting. 5) This changed their previous assumption; after the completion of the initial composition, Rubens considered the painting good and ordered his studio assistants to copy it. However, Rubens then had further ideas about the composition and repainted his painting (Olivier et al. 2005, 10-11).

3.2 Basic Actions in Genealogical Research -Collect and Relate

Here we can easily find basic actions in genealogical research. To solve the genealogical problem the researchers tried to collect as many similar images as possible. Above I mentioned only three images, Rubens's painting in London, the studio painting in Dresden, and the newly found layer beneath the Rubens's painting, but they collected more than 10 similar images in the article and related each to each. So the basic actions of the researchers is collecting and relating. The basic actions are the same, even if the research object is different. We can see the same actions in studies on Eduard Manet (Flam 1996), Korin Ogata (The MOA museum of art et al. 2005), and the IBM logo painted with 35 atoms (Hennig 2004).

3.3 Deficiency of Search Engines

However, for beginners, it is hard to collect or recollect proper similar images in their heads. In order to encourage them to act like professional researchers, what kind of intelligent environment should we offer? One might think of search engines such as Google and TinEye[1]. But do they already offer a sufficiently intelligent

[1] TinEye is a reverse image search engine. `http://tineye.com/`

environment for beginners? -I do not think so. Certainly TinEye can search more than 1 billion images in under 1 second. However, it only gathers images which are completely the same as the query image and images which contain parts of the query image. So, if we input Rubens's painting in London as the query image in TinEye (beta version), we can not find the image of the studio work in Dresden in the search results. On the contrary, if a beginner who wants to study art input the keywords "The Judgement of Paris" in Google, the result is too huge for the beginner to relate the images with each other. Such insufficiency comes from the concept level, not from the technological level. The aim of current search engines is to identify illegal images which use the whole or the parts of the query image without permission. So these engines can find parodies and appropriations of the query image. But they do not care about historical and gradual transformations of the query image. In addition, these engines lack the sense of excitement we can feel when we see the gradual transformation of images.

3.4 Intelligent Environment by Which We Can Encourage Beginners (1)

Then, how can we encourage beginners to act like professional researchers? Here I can introduce a tiny but important attempt by Kida (Kida 2007). She proposed a unique method of producing genealogical images. In her article she assumed that we can compare the succession of artistic style from teachers to the pupil with genetic inheritance of features from parents to the child. So, if we know the style of a teacher A and the style of the pupil C, then we can estimate the requested style of another teacher B. In the estimation Kida used morphing and feature points matching. On the basis of this assumption, she produced the style of the painter Giovanni Santi, who influenced young Raffaello as well as his teacher Perugino. Of course, the legitimacy of such a direct comparison of succession and transformation of artistic style with genetic inheritance is open to discussion. And her method must be improved because it was applied only to facial features of a painting, not to a painting as a whole. Nevertheless, from the point I mentioned above, this research direction is indispensable. If we could have an intelligent environment which makes it possible to transform a historical image and to collect/recollect images which are art historically relevant to the image, then we could enjoyably relate images to images and imitate actions of art researchers.

4 Basic Actions in Classificatory Thinking

In the final chapter I consider basic actions in classificatory thinking. Contrary to genealogy which is relatively objective, ways of classifying an artwork tend to vary from person to person. They depend on expectations which viewers have about the artwork. If the function of an artwork could satisfy viewer's expectations, then the artwork would have more chances to be reproduced and survive (See ch.2.2).

4.1 Five Functions of Painting

In relation to this point, Kishi's pictorial act theory is highly suggestive (Kishi 2008). In the book he quotes a report by Kyoshin Iijima. It concerns about an Ukiyo-e print by an artist Kunichika Toyohara. According to the report, Iijima went to see the play which Kunichika depicted in his Ukiyo-e in order to check whether the depicted play is the same as the real play. From this report Kishi supposes that a painting which depicted a play had not only a function to announce the opening of the play but also another function to invite people to go and see the play. In addition, Kishi also supposes that the person called Iijima had strongly expected that a painting must have a function to depict things truly. The painting by Kunichika could satisfy Iijima's expectations so it was written in his dairy report and survived.

According to Kishi, a painting functions to convey to viewers more than one function out of five: 1) referential function; if a painting can convey to viewers referential information of depicted objects, then it satisfies its referential function, 2) emotional function; if it can convey emotional information of the author, then it satisfies its emotional function, 3) directive function; if it can convey directive information to viewers, for example, go and see the show, then it satisfies its directive function, 4) aesthetic function; if it can convey aesthetic information of the painting itself, then it satisfies its aesthetic function, 5) meta-imagery function; if it can convey to viewers information about rules and techniques with which the author created the painting, then it satisfies its meta-imagery function. Kishi also says that every comment on a painting is inevitably related to more than one function of the painting. The report in the diary of Iijima is related to referential and directive functions of the painting. Through analysis of the diary, Kishi reveals expectations of the viewer to the painting.

4.2 Basic Actions in Classification (1) -Extract and Assign

Instead of diaries, now we have innumerable comments on the Internet. Here I take comments on an exhibition of a painter Andrew Wyeth (the English title of this exhibition was "Andrew Wyeth: Emotion and Creation") as an example[2] and try to make explicit some basic actions of classificatory procedure. What are the basic actions? In order to classify both functions of a painting and viewers' expectations to the painting, we have at least two approaches. The first is to extract keywords from viewers' comments, and according to the keywords, we assign functions to

[2] I chose the first ten out of innumerable comments collected with Google (query words, "Wyeth, exhibition")

the painting. For example, if we find keywords about how to create a painting such as “watercolor,” “drybrush,” “tempera,” “process,” etc. we can assign meta-imagery function to the painting. And if we find keywords about the impressions of the painting such as “loneliness,” “desolation,” “tranquility,” we assign emotional function to the painting. The basic actions are to extract and assign. I think this is easy to do because there are innumerable web sites which have glossaries of “art words.” We can consult with them about technical terms like “tempera.” And everyone can easily understand the meaning of emotive words such as “loneliness.”
In this way I made a table (Table 4.1). Of course, the number of collected sample is insufficient, so it might only roughly suggest the tendency both of viewers’ expectations and of functions of the paintings at the exhibition. It also might indicate that the viewers’ expectations straightforwardly reflect expectations of the exhibition organizer who named the subtitle of the exhibition as “Emotion and Creation.”

On referential function	8
On emotive function	53
On directive function	1
On aesthetic function	4
On meta-imagery function	65

Table 1: Functional classification of the total numbers of keywords in the first ten comments.

4.3 *Basic Action in Classification (2) -Read*

The second approach for classification is to read in comments, which have no apparent keywords, hidden expectations of viewers and functions of paintings. I take an example from a weblog collected by the search engine *Biglobe Minnano Hyoban* (beta version): “the exhibition was worth seeing. It was too good, so I could not buy a catalogue.” (13 February 2009 `http://yaplog.jp/meltwater/archive/186` The comment is originally written in Japanese. Akiba literally translated it into English.) “It was too good, so I could not buy a catalogue” seems strange. It means the viewer could not be satisfied with the reproductions in the catalogue because they could not reproduce the high quality of the genuine paintings. So the comment shows that the viewer had great expectations for the technical aspects of the paintings and their meta-imagery function. If we compare this comment with another comment which was also collected by the same engine Biglobe Minnano Hyoban (beta-version), we can easily find what kind of intelligent environment we should offer to beginners. The comment is as follows: “Wyeth often used techniques such as watercolor, drybrush, tempera. His paintings are realistic and warmhearted. I was deeply impressed by the subtle expressions produced by the painted layers.” (13 February 2009 `http://ameblo.jp/`

petit-citron/entry-10198404798.html The comment is also originally written in Japanese. Akiba also literally translated it into English.) In spite of apparent difference, these two comments have the same content. The second one also shows the viewer's expectation for technical aspects of the paintings and their meta-imagery function. So we need an intelligent environment which can improve beginners' ability to read the hidden meaning of other's comments.

4.4 Intelligent Environment by Which We Can Encourage Beginners (2)

Again I return to the first comment in order to make clear what kind of intelligent environment we should offer. If an intelligent environment could analyze the comment in the following procedures, it could encourage beginners. 1)The environment judges the first half of the second sentence, "it was too good," as a sentence which evaluates the exhibition positively, 2) from this judgment it generates various sentences which state possible reasons why someone evaluates an exhibition positively, for example, "in the exhibition the paintings are beautiful," "the paintings are real," "the paintings are technical," etc.(results A), and, at the same time, 3) the environment also judges the second half of the second sentence, "so I could not buy a catalogue," as the sentence which evaluates a catalogue negatively, 4) from this judgment it generates various sentences which state possible reasons why someone might evaluate a catalogue negatively, for example, "the reproductions in the catalogue are not true to the originals," "the articles in the catalogue are difficult to understand," "the catalogue is too expensive," etc.(results B), and 5) the environment compares two groups of results and proposes some suitable combinations between A and B.

5 Conclusion

In this paper I made the basic actions of art researchers explicit in the following way: At first, on the basis of philosophical aesthetics, I made clear the missions of art study and two ways of thinking which underlie it –genealogical thinking and classificatory thinking. I treated genealogical thinking in Ch.3, and found its basic actions, that is, collecting and relating. I discussed classificatory thinking in Ch.4 and drew out its basic actions, that is, extracting, assigning, and reading. So this paper has achieved its expected aim. In addition, I suggested two kinds of intelligent environments by which we could encourage beginners to act like professional researchers. On the one hand, in order to encourage genealogical actions, I suggested an intelligent environment which makes it possible to transform a historical image and to collect/recollect other images which are art historically relevant to the image. On the other hand, in order to encourage classificatory actions, I suggested an intelligent environment which helps beginners to read the hidden meaning of reviewer's comments. Of course I only proposed directions for future intelligent environments, not the detailed design of them. However, if these environments could be realized, they would be helpful not only for beginner art viewers but also for professional art researchers because they could make their basic actions more efficient and make it possible for them to spend more time on their own creative interaction.

References

1. Cohen, J.: Information and Content. In: Floridi, L. (ed.) The Blackwell Guide to the Philosophy of Computing and Information, pp. 215–227. Blackwell, Malden (2004)
2. Currie, G.: Art Works as Action Types. In: An Ontology of Art, pp. 46–86. Macmillan, London (1989); Rpt. in: Lamarque, P., Olsen, S.H. (eds.): Aesthetics and the Philosophy of Art: The Analytic Tradition, pp. 103–122. Blackwell, Malden (2004)
3. Dutton, D.: A Naturalist Definition of Art. Journal of Aesthetics and Art Criticism 64, 367–377 (2006)
4. Flam, J.: Looking into the Abyss: The Poetics of Manet's A Bar at the Folies-Bergére. In: Collins, B.R. (ed.) 12 Views of Manet's Bar, pp. 164–188. Princeton University Press, Princeton (1996)
5. Hennig, J.: Vom Experiment zur Utopie: Bilder in der Nanotechnologie. Bildwelten des Wissens 2(2), 9–18 (2004)
6. Kida, Y.: Gazo Shori Gijutsu o Mochiita Rafello no Kaiga Gijutsu no Bunseki: Bijutsugaku ni okeru Gazo Shori Gijutsu no Donyu [Analysis of Techniques in Rafaello's Paintings by Image Processing Technology: Introduction of Image Processing Technology into Science of Art] (in Japanese), Graduate Thesis. Nagoya University (2007)
7. Kishi, F.: Kaiga-Koi Ron: Ukiyo-e no Pragmatism [Pictorial Act Theory: Pragmatism of Ukiyo-e] (in Japanese). Daigo-Shobo, Kyoto (2008)
8. Levinson, J.: What a Musical Work is. Journal of Philosophy 77(1), 5–28 (1980)
9. Maeda, T.: Sikosei to Mokutekironteki Kino Shugi [Intentionality and Teleological Functionalism] (in Japanese). In: Nobuhara, Y. (ed.) Kokoro no Tetsugaku [Philisphy of Mind: Causation, Intentionality and Consciousness], pp. 85–129. Keiso-Shobo, Tokyo (2004)
10. Millikan, R.G.: Language, Thought, and Other Biological Categories: New Foundations for Realism. The MIT Press, Cambridge (1984)
11. The MOA museum of Art. National Institute for Cultural Properties Tokyo. Ogata Korin Hitsu, Kokuho Kohakubaizu Byobu [The National Treasure: Korin Ogata's Red and White Plum Blossoms Screens] 05390 (in Japanese). Chuo Koron Bijutsu Shuppan, Tokyo (2005)
12. Oliver, L., Healy, F., Roy, A., Billinge, R.: The Evolution of Rubens's Judgement of Paris (NG194). National Gallery Technical Bulletin 26, 4–22 (2005)

A Simple method for 3-dimensional Photorealistic Facial Modeling

Ippei Torii, Yousuke Okada, Manabu Onogi, and Naohiro Ishii

Abstract The process of creating photorealistic 3-dimensional computer graphic (3DCG) images is divided into two stages, i.e., modeling and rendering. Automatic rendering has gained popularity, and photorealistic rendering is generally used to render different types of images. However, professional artists still model characters manually. Moreover, not much progress has been achieved with regard to 3-dimensional shape data acquisition techniques that can be applied to facial modeling; this is an important problem hampering the progress of 3DCG. Generally, a laser and a highly accurate camera are used to acquire 3-dimensional shape data. However, this technique is time-consuming and expensive. Further, the eyes may be damaged during measurements by this method. In order to solve these problems, we have proposed a simple method for 3-dimensional shape data acquisition using a projector and a web camera. This method is economical, simple, and less time-consuming than conventional techniques. In this paper, we describe the setup of the projector and web camera, shape data acquisition process, image processing, and generation of a photorealistic image. We also verify the accuracy of this method by comparing the photograph of a face with its rendered image.

Ippei Torii
Aichi Institute of Technology Management and Information Science,
Yachikusa 1247, Yakusa-cho, Toyota-shi, Aichi, Japan
e-mail: mac@aitech.ac.jp

Yousuke Okada
Aichi Institute of Technology Management and Information Science,
Yachikusa 1247, Yakusa-cho, Toyota-shi, Aichi, Japan
e-mail: x07232xx@aitech.ac.jp

Manabu Onogi
Aichi Institute of Technology Management and Information Science,
Yachikusa 1247, Yakusa-cho, Toyota-shi, Aichi, Japan
e-mail: x08016xx@aitech.ac.jp

Naohiro Ishii
Aichi Institute of Technology Management and Information Science,
Yachikusa 1247, Yakusa-cho, Toyota-shi, Aichi, Japan
e-mail: ishii@aitech.ac.jp

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 487–497.
springerlink.com

1 Introduction

Recent developments in 3-dimensional computer graphics (3DCG) have made it possible to generate photorealistic images. Research is being conducted on various 3DCG applications and generation of photorealistic images is one of the main research topics. The generation of photorealistic images is divided into two stages, namely, modeling and rendering. Modeling is a process which defines and creates the data of facial shape and rendering is a process whereby the final image is generated from the modeling data. The rendering is comparatively easy for automatic creation and it can generate photorealistic images from complicated form easily. However, modeling must be carried out manually by an artist, and its efficient improvement is late. Achieving highly precise geometric modeling is an important challenge in 3DCG[4]. Conventional methods such as stereo imaging and 3-dimensional scanning are used for 3-dimensional modeling. Stereo imaging makes use of a stereo camera that can simulate binocular vision, and 3-dimensional scanning involves the use of a laser scanner and a CCD camera. In 3-dimensional scanning, the vertical planes of an object are measured using the laser scanner by the 2-dimensional scanning method, in which a method of measuring the object shape of carrying the cross section is developed. However, conventional methods have some disadvantages. Stereo imaging is difficult to use under varying light conditions because the shadow of the object is misinterpreted by the turbulence light. In 3-dimensional scanning, we can acquire large high-density data by only one scan in short time; however, a large amount of memory is required because it is necessary to scan the object in all directions. In 3-dimensional scanning, the object size in the passage direction is misinterpreted when the speed and direction of objective change in the measurement domain[2, 3].

In this paper, we propose a new method that reproduces the shape data of a person's face on the computer in a short period of time. In comparison with the laser scan method, the proposed method is economical, simple, and less time-consuming. The following steps are carried out in order to acquire the shape data of the face:

(1) We show the projection method that can measure delicate ups and downs precisely.

(2) Swift calculation of shape data of the face is performed in the image processing stage.

An inexpensive web camera and a projector device have been developed for the step (1), and line extraction, noise processing, and line thinning tools have been developed for the step (2)[8, 17].

2 Facial Modeling

It can be confirmed that the line is horizontal even if it is observed from anywhere when the horizontal line projected to the plane. However, the line has been often curved when this line is projected to complex ups and downs, and observed from the upper part. The shape of a face can be acquired by comparing and analyzing this difference with a web camera. In addition, we can generate a model of the facial shape by converting the acquired shape data to a file format that can be read with the help of a general software application. The method that we have used for facial modeling is explained below. The measured values are transformed to coordinates in a Data Exchange File (DXF)[10]. DXF adopts polygon technology, and it has an advantage that modeling data is converted polygon at the same time. However, polygons are incapable of accurately representing curved surfaces because they are represented by a set of vertices. Therefore, we use a subdivision surface[14] to represent smooth surfaces. Further, the polygon model is UV mapped[15, 16] to obtain a texture for the surface of the object. We can render the texture image by photorealistic rendering methods such as ray tracing and obtain a photorealistic image.

3 Stages in Facial Modeling

3.1 Environment

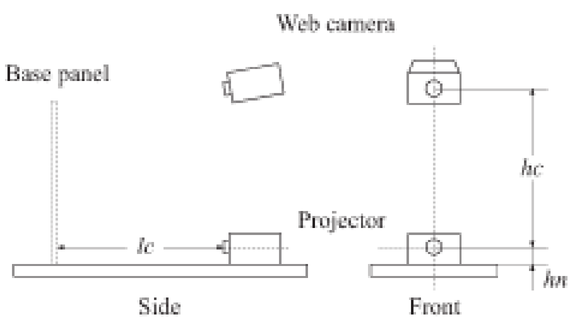


Fig. 1: Projector and web camera set up

The basic devices required for facial scanning in this method are a projector and a web camera. The projector and the web camera must be set up in the position as shown in Fig. 1. The projector and the web camera should be set up based on the lens.

3.2 Initialization of base line and base panel

Before facial scanning, we generated a line of 1 pixel (base line) that is standard on the computer. The line is projected in three colors on the base panel (white drawing paper), i.e., red (R: 255 G: 0 B: 0), green (R: 0 G: 255 B: 0), and blue (R: 0 G: 0 B: 255), and compared them on the basis of extraction accuracy. We find that red is not suitable for extraction because it interferes with the color of the lips and skin. Green is also inappropriate because it would mask the color of a blood vessel if it is prominent. Finally, blue is used as the background for chroma key as it has the highest extraction accuracy. The base line generated by the computer is projected on the base panel. We adjust the zoom of the projector and the distance between the projector and the base panel in order to maintain a distance of 7 mm between all four sides of the base panel and the base line. The adjustment is completed by measuring the distances with a ruler. We also measure the distances between the projector's lens and the base panel (lc), the center of the projector lens and web camera (hc), and the installation side and center of the projector lens (ln). Then, we obtain a static image with the web camera and map the base line on the global coordinate system to the camera coordinate system. Since we have defined the 7-mm interval in the global coordinate system, it automatically becomes a constant in the camera coordinate system. The distortion of the camera lens must be corrected.

3.3 Generation and projection of scan line

We generate a scan line projected to the face with the projector based on information of the base line set in the preceding section. The color of the scan line is the same as that of the base line, i.e., blue. The reason for this is explained in the previous section. The direction of movement of the scan line should be the same as that of the base line. Here, sequential scanning is used in order to reduce processing complexity and mutual interference by multiple base lines. It is necessary for the web camera to take a picture after every line is scanned. This operation is performed at very high speeds, which reduces eye strain. The danger of the scan line passing over the eyes is not as great as that posed by a laser beam; however, the examinee must close his/her eyes when a picture is being taken.

3.4 Extraction and analysis of scan line

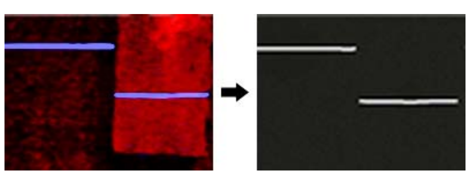


Fig. 2: Analysis of scan line

Only the scan line is extracted from the image data and all additional information is discarded. It has been confirmed that the scan line is identified more clearly when the intensity of the blue background is greater than 187. Furthermore, the noise in the direction of the x-axis can be removed that deviated from the scan line greatly using continuing on an image. Line thinning is performed in order to increase the clarity of the extracted line. The thickness of the line is set to 1 pixel by deducing the average of the point of the topmost part and the lowermost part of the line. In this way, the scan line is extracted completely (Fig. 2)[12, 13].

3.5 Computation of coordinate values

The scanned data must be sampled along the x-axis, and the amount of data must be determined. Minute facial contours are considered as noise. According to the sampling theorem[11], the x-axis should be divided into at least 5 intervals. It calculates the coordinates on the basis of the analyzed image after sampling. The values of lc, ln, and hc have already been measured. The scan line which is projected from the projector, becomes a real image at point y_i on the base panel in Fig. 3. Point y_i is also measured in section 3.2. However, when the object is placed in front of the base panel, it becomes a real image at point P of the object. It holds the same as the point observed by point y' on the base panel by the web camera. And, the straight line of **a**, **b**, and **c** is linear function, therefore it is easy to estimate an intersection coordinates (x_P, y_P). We show a schematic diagram of the coordinates calculation in Fig. 3.

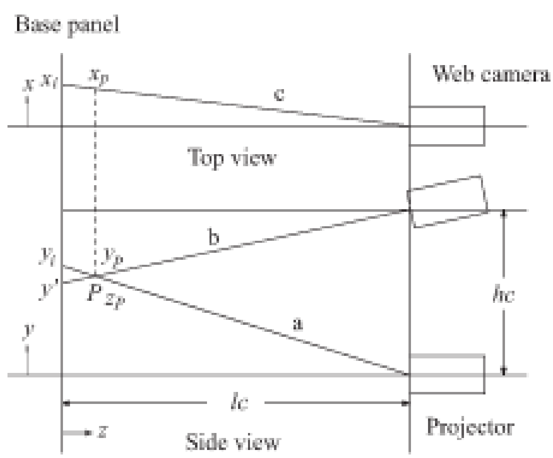


Fig. 3: Computation of coordinate values

$$y = -\frac{y_i}{lc} z + y_i \tag{1}$$

$$y = \frac{hc - y'}{lc} z + y' \tag{2}$$

$$x = -\frac{x_i}{lc} z + x_i \tag{3}$$

$$z_P = \frac{lc(y_i - y')}{hc + y_i - y'} \tag{4}$$

$$y_P = -\frac{y_i}{lc} z_P + y_i \tag{5}$$

$$y_P = -\frac{y_i(y_i - y')}{hc + y_i - y'} + y_i \tag{6}$$

$$x_P = -\frac{x_i}{lc} z_P + x_i \tag{7}$$

$$x_P = -\frac{x_i(y_i - y')}{hc + y_i - y'} + x_i \tag{8}$$

3.6 Conversion of modeling data to polygonal data

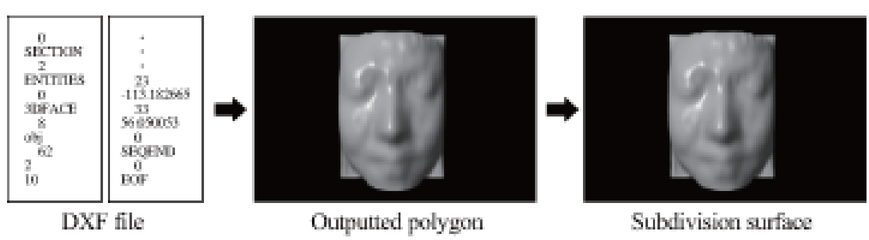


Fig. 4: Output modeling data of face

The acquired facial data are converted to polygonal data by a general software application (Fig. 4). Representing objects in the form of polygons is a standard modeling technique used in 3DCG as the object can be easily edited. Therefore, we adopt the DXF-3DFACE[10] file format. The simple file structure of DXF simplifies the process of mapping facial data to polygon coordinates. Furthermore, since many software applications support DXF, it is possible to import DXF data from one application to another.

3.7 UV mapping and rendering

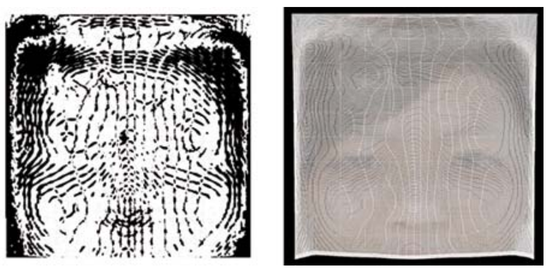


Fig. 5: Mesh parameterization image (left) and Mesh parameterization (right)

We adopt UV mapping[15, 16] to obtain the texture image of the modeling object. Since it is a method that makes use of a 2-dimensional image to represent a 3-dimensional object (Fig. 5), we think it makes the best use of the advantage of method of facial modeling.
Firstly, mesh parameterization of the 3-dimensional modeling object is performed, which results in a 2-dimensional mesh image. We resample the textured mesh image obtained from the photograph of the face. We map the sampled photograph of the modeling object and render it with the help of ray tracing a photorealistic rendering technique.

4 Verification, comparison, and evaluation of the acquired data

4.1 Error estimation

We defined equations for error estimation from section 3.5.

$$C = \frac{y_i - y'}{hc + y_i - y'} \tag{9}$$

$$z_P = lcC \tag{10}$$

$$y_P = -y_i C + y_i \tag{11}$$

$$x_P = -x_i C + x_i \tag{12}$$

We calculate the error values from these equations. From equation (10), we can see that the accuracy of z_P depends on lc. Therefore, in order to reduce the error in z_P, the value of lc needs to be controlled. We can also see that x_P depends on x_i, which means sampling internal (number of partitions in vertical direction) of x-axis. Further, y_P (accuracy of projector and web camera) depends on y_i (capture angle). We calculate C from equation (9). In our method, equation (9) has to satisfy $y_i - y' > 0$ and $hc > 0$ so that $C > 0$. However, if $y_i - y' > 0$ and $hc > 0$ are not large enough, then it will result in a large error margin.

4.2 Verification of acquired data by geometric form

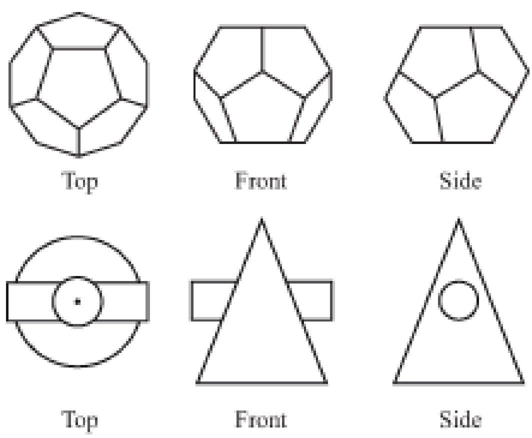


Fig. 6: Drafting of geometric form

It is necessary to measure the dimensions of a known object and verify them with the acquired data to measure the accuracy of the proposed method. In this paper, we verify the accuracy of our proposed method by using a geometric form. (Fig. 6) A geometric form is a solid model used for sketches etc., whose sizes are known. The accuracy of acquisition data is evaluated by using two geometric forms. The actual dimensional values and the measurement data acquired by this method are shown in Fig. 7 and Table 1.

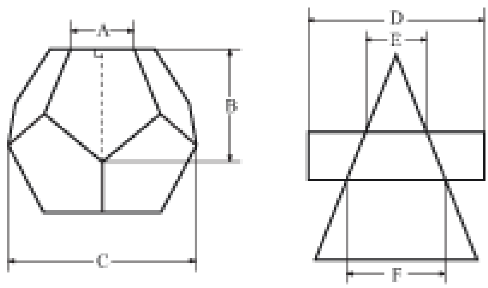


Fig. 7: Dimensions of the solid model

Table 1: Measurement data

	A	B	C	D	E	F
Actual size	70 mm	105 mm	175 mm	185 mm	80 mm	100 mm
Measured value	68 mm	103 mm	171 mm	182 mm	77 mm	98 mm
Difference	- 2 mm	- 2 mm	- 4 mm	- 3 mm	- 3 mm	- 2 mm

According to Table 1, line **c** (Fig. 3) can be confirmed that the margin of error is the largest when full scale is compared with measurement scale. Therefore, x_P is related to equations (7), (8), (9), and (12) which means length of x-axis. From equations (9) and (12), the possibility that the problem occurs is low because x_P does not depend on lc. Moreover, y_i and y' has the influence to all variables. As a result, it is guessed that there is a factor in x_i. The x_i means sampling internal of x-axis (number of partitions in vertical direction). This error margin in accuracy is due to less sampling rate. Therefore, the sampling rate needs to be increased in order to reduce the error margin.

Table 2[1] shows the anthropometric results from the mannequin by methods of the stereo camera and the solo camera. The Table show the difference between distances in the 3-dimensional reconstructions and the real object. Table 1 that shows this method has higher accuracy than Table 2. The accuracy of this method is verified by comparing the actual dimensional values of the solid model and the measurement data acquired by this method.

Table 2: Anthropometric results

	Stereo camera			Solo camera		
Measure(mm)	Shift	Gray	Stripe	Shift	Gray	Stripe
Mouse Width	-0.22	0.72	0.39	-0.78	-1.54	-2.27
Iris Diameter	0.37	0.08	0.55	0.26	-0.53	-0.90
Eye Separation	-0.24	0.43	0.92	-0.45	-0.49	2.36
Eye Width	-0.60	-0.86	-2.15	1.14	1.78	-7.19
Nose Height	1.53	-0.33	-0.92	-0.30	-0.32	0.00
Nose Width	0.45	0.53	0.23	-1.25	-0.98	-2.40
Nose to Mouth	0.50	0.58	0.05	-0.05	2.21	-0.81
Average	3.31	2.67	3.06	3.09	6.06	8.75
Average	3.01			5.96		

4.3 Comparison and evaluation of rendered image and face

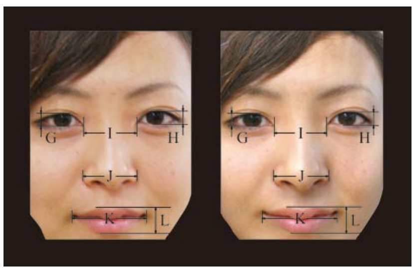


Fig. 8: Photographed face (left) and rendered image (right)

Table 3: Comparing photography with rendering image

	G	H	I
Photographed image	8.005 mm	8.344 mm	33 mm
Rendered image	8.360 mm	8.615 mm	33 mm
difference	+ 0.355 mm	+ 0.271 mm	0 mm (standard)
	J	K	L
Photographed image	33.68 mm	46.32 mm	16.21 mm
Rendered image	34.53 mm	47.32 mm	16.47 mm
difference	+ 0.850 mm	+ 1.000 mm	+ 0.260 mm

We compare the rendered image with the photograph of the face. So, we set up standard length 33 mm with the length of the horizontal line that connected heads of eyes of right and left. We compare each part of the face with the actual size. The results in Fig. 8 and Table 3 show that the UV mapping technique used in this study has sufficient accuracy since the error margin is less than 1 mm.

5 Conclusions

In this paper, an economical and simple method for generation of photorealistic images is proposed. The image processing method and calculations for this study are described. Furthermore, muscle animation[5, 6, 7] can also be applied if muscular motions of the face are compiled into a database. The simplicity of our method of facial modeling will be applied to the entrance checking information, the interface of the robot and the medical support apparatus and so on[9].

References

1. An, D., Woodward, A., Delmas, P., Gimelfarb, G., Morris, J.: Comparison of active structure lighting mono and stereo camera systems: Application to 3d face acquisition. In: ENC 2006: Proceedings of the Seventh Mexican International Conference on Computer Science, Washington, DC, USA, pp. 135–141. IEEE Computer Society, Los Alamitos (2006)
2. Lee, Y., Terzopoulos, D., Keith: Realistic Modeling for Facial Animation. In: SIGGRAPH, ACM Computer Graphics (1995)
3. Bickel, B., Lang, M., Botsch, M., Otaduy, M.A., Gross, M.: Pose-Space Animation and Transfer of Facial Details. In: ACM SIGGRAPH Symposium on Computer Animation (2008)
4. Akimoto, T., Suenaga, Y., Wallace, R.: Automatic creation of 3D facial models. IEEE Computer Graphics and Applications 13(5), 16–22 (1993)
5. Doyle, J., Philips, J.: Manual on Experimental Stress Analysis, 5th edn. Society for Experimental Mechanics (1989)
6. Essa, I.A.: Visual Interpretation of Facial Expressions using Dynamic Modeling. Ph.D thesis, MIT (1994)
7. Frick, Hans: Human Anatomy, vol. 1. Thieme Medical Publishers, Stuttgart (1991)
8. Gray, H.: Anatomy of the Human Body, 29th edn. Lea Febiber, Philadelphia (1985)
9. Guenter, B.: A system for simulating human facial expression. In: State of the Art in Computer Animation, pp. 191–202. Springer, Heidelberg (1992)
10. Ochiai, S.: DXF Handbook. Ohmsha, Ltd. (2003)
11. Shannon, C.E.: Communication in the presence of noise. Proc. Institute of Radio Engineers 37(1), 10–21 (1949); Reprint as classic paper in: Proc. IEEE 86(2) (1998-2)
12. MSDN Library, Microsoft, `http://msdn.microsoft.com/ja-jp/library/default.aspx`
13. MSDN Microsoft, `http://msdn.microsoft.com/ja-jp/visualc/default.aspx`
14. Zorin, D.: Modeling with Multiresolution Subdivision Surfaces. In: ACM SIGGRAPH 2006 Course, pp. 30–50. ACM SIGGRAPH (2006)
15. Sander, P.V., Snyder, J., Gortler, S.J., Hoppe, H.: Texture Mapping Progressive Meshes. In: Computer Graphics (SIGGRAPH 2001 Conf. Proc.), pp. 409–416 (2001)
16. Tarini, M., Yamauchi, H., Haber, J., Seidel, H.-P.: Texturing Faces. In: Parco. Graphics Interface 2002 (2002)
17. Bodyline Scanner, Hamamatsu Inc., `http://www.hpk.co.jp/`

Automatic Linefeed Insertion for Improving Readability of Lecture Transcript

Masaki Murata, Tomohiro Ohno, and Shigeki Matsubara

Abstract The development of a captioning system that supports the real-time understanding of monologue speech such as lectures and commentaries is required. In monologues, since a sentence tends to be long, each sentence is often displayed in multi lines on the screen and becomes unreadable. In the case, it is necessary to insert linefeeds into a text so that the text becomes easy to read. This paper proposes a technique for inserting linefeeds into a Japanese spoken monologue sentence as an elemental technique to generate the readable captions. Our method inserts linefeeds into a sentence by applying the rules based on morphemes, dependencies and clause boundaries. We established the rules by circumstantially investigating the corpus annotated with linefeeds. An experiment using Japanese monologue corpus has shown the effectiveness of our rules.

1 Introduction

Real-time captioning, which displays transcribed texts of monologue speech such as lectures, is a technique for supporting the speech understanding of deaf persons, elderly persons, or foreigners. In monologues, since a sentence tends to be long, each sentences is often displayed in multi lines on the screen. In the case, it is necessary to insert linefeeds into a text so that the text becomes easy to read.

Masaki Murata
Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, 464-8601, Japan, e-mail: murata@el.itc.nagoya-u.ac.jp

Tomohiro Ohno
Graduate School of International Development, Nagoya University, Furo-cho, Chikusa-ku, 464-8601, Japan, e-mail: ohno@nagoya-u.jp

Shigeki Matsubara
Information Technology Center, Nagoya University, Furo-cho, Chikusa-ku, 464-8601, Japan, e-mail: matubara@nagoya-u.jp

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 499–509.
springerlink.com

This paper proposes a technique for inserting linefeeds into a Japanese spoken monologue sentence as an elemental technique to generate readable captions. We assume that a screen which displays only multiline caption is placed to provide the caption information to the audience on the site of lectures and commentaries. In our method, the linefeeds are assumed to be inserted into only the boundaries between *bunsetsus*[1]. Our method applies the rules for inserting linefeeds to a sentence. The rules are created in consideration of the boundary into which linefeeds are not inserted, the boundary into which linefeeds should be inevitably inserted, and the boundary into which linefeeds can be inserted.

We established the rules based on the emerging pattern of morphemes, dependencies and clause boundaries by circumstantially investigating the corpus annotated with linefeeds. We conducted an experiment on inserting linefeeds by using Japanese spoken monologue corpus. As the results, the precision and recall of our method was 82.7% and 79.0%, respectively. Our method improved the performance dramatically compared with the baseline method, which is implemented based on bunsetsu boundaries and the maximum number of characters per line, and has been confirmed to be effective.

2 Caption display of spoken monologue

2.1 Linefeeds insertion in monologue sentences

In our research, as an environment in which captions are displayed on the site of lectures, we assume that a screen for displaying only captions is used. Figure 1 shows our assumed environment in which captions are displayed.

One line of the displayed text switches to other line and multiline text is always displayed, being scrolled.

As shown in Fig. 2, if the transcribed text of monologue speech is simply displayed in accordance with only the width of a screen without considering the proper point of linefeeds, the caption becomes not easy to read. Especially, since the audience are forced to read the caption in accordance with the speaker's utterance speed, it is important that linefeeds are inserted into the displayed text in consideration of the good readability as shown in Fig. 3.

In our research, we set the following concepts as the proper points into which linefeeds are inserted on captioning.

- Linefeeds have to be inserted so that each line constitutes a semantically meaningful unit.

[1] *Bunsetsu* is a linguistic unit in Japanese that roughly corresponds to a basic phrase in English. A bunsetsu consists of one independent word and zero or more ancillary words. A *dependency* is a modification relation in which a *modifier bunsetsu* depends on a *modified bunsetsu*. That is, the modifier bunsetsu and the modified bunsetsu work as modifier and modifyee, respectively.

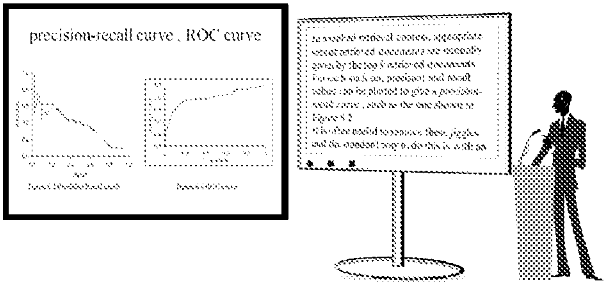


Fig. 1: Caption display of spoken monologue

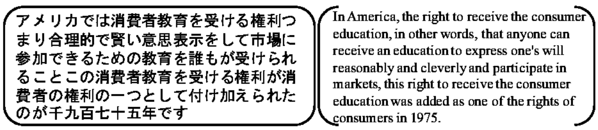


Fig. 2: Caption of monologue speech

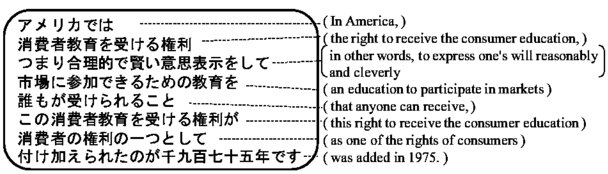


Fig. 3: Caption into which linefeeds are properly inserted

- The number of characters in each line have to be less than or equal to the maximum number of characters per line, which is established based on the width of a screen.

Here, since a bunsetsu is the smallest semantically meaningful language unit in Japanese, our method adopts the bunsetsu boundaries as the candidates of points into which linefeeds are inserted.

2.2 Related works

There exist a lot of researches about captioning [1, 8, 2]. However, there are few conventional researches about inserting linefeeds on captioning except the following researches. Monma et al.[5] proposed the method for inserting linefeeds based on patterns of a sequence of morphemes. They analyzed the point into which linefeeds were inserted on the closed-captions of Japanese TV shows, and then made the rules for inserting linefeeds. However, in this research, the linefeeds are inserted on the constraint that the text displayed in a screen all switches other text at a time, that is, the readability in case of our assumed caption display system is not considered.

Saikou et al.[7] proposed the method for captioning based on the gradually chunking. This method chunks morphemes into a "*constituent*," which corresponds to the nominative, predicates, case elements and so on in a sentence, and then chunks "*constituents*" into a "*phrase*." This method can insert linefeeds so that each line becomes a linguistic unit, by concatenating character string of the *constituents*, which do not constitutes a different *phrase* mutually, until the length of each line reaches 15 characters. However, this research did not verified the relation between the proper point into which linefeeds should be inserted, the *constituent* and the *phrase*.

3 Linefeeds in monologue sentences

We investigated the actual spoken monologue data to make the rules for inserting linefeeds based on the concepts described in Section 2.1. In our investigation, we used the spoken monologue corpus "Asu-Wo-Yomu[2]," annotated with information on morphological analysis, clause boundary detection, bunsetsu segmentation, dependency analysis[6], and linefeeds inserted by hands. In what follows, we organize bunsetsu boundaries by classifying them into the following three categories: the boundary into which linefeeds are not inserted, the boundary into which linefeeds should be inevitably inserted, and the boundary into which linefeeds can be inserted.

3.1 Boundaries into which linefeeds are not inserted

As the result of the investigation, we observed that linefeeds were not inserted into the following bunsetsu boundaries.

- The end boundary of the bunsetsu of which the part-of-speech of the rightmost morpheme is "adnominal particle," "case particle-*toyu*," "case particle-*no*," "auxiliary verb," or "adnominal" and depends on the right-hand neighbor bunsetsu.
- The start boundary of the bunsetsu which consists of one of 36 different verbs such as "," " ," which play an auxiliary role and have no corresponding word in English.

[2] Asu-Wo-Yomu is a collection of transcriptions of a TV commentary program of the Japan Broadcasting Corporation (NHK). The commentator speaks on current social issues for 10 minutes.

Table 1: Strong clause boundary

compound clause	*-toka*, *-ga*, *-shi*, *-de*, *-keredomo*, *-tari*
condition clause	*-kagiri*, *-ba*, *-tara*, *-kekka*, *-tokoro*
time clause	*-tokini*, *-atoni*, *-ima*, *-ato*, *-tokino*, *-tokiniwa*, *-sonota*
reason clause	*-node*, *-kara*
adverbial clause	*-tameniha*, *-tame*, *-nagara*, *-nado*, *-tewa*, *-yo*, *-sonota*
others	Indeclinable words stopping, Interjection

3.2 Boundaries into which linefeeds are inevitably inserted

Since a clause constitutes a semantically meaningful language unit, a clause boundary can be widely-accepted as the candidate of the proper point into which linefeeds should be inserted. However, the role of each clause on a sentence is different by the types. This means that the likelihood that a linefeed is inserted into a clause boundary is different by the type of the clause boundary. As the result of the above-mentioned analysis, there existed 29 types of clause boundaries into which linefeeds should be inevitably inserted. Table 1 shows the clause boundary types into which linefeeds should be inevitably inserted. We call these clause boundaries the **strong clause boundary** as a whole hereafter. The strong clause boundary accounted for 49.1% of clause boundaries which appear in the analysis data.

3.3 Boundaries into which linefeeds can be inserted

3.3.1 Insertion of linefeeds based on clause boundaries

There exist clause boundaries into which linefeeds are not necessarily inserted but are inserted with high probability in a context. As such a clause boundary type, there are “adverbial clause,” “adverbial adjective clause,” “supplement clause,” “concessive clause-*temo*,” “indirect interrogative.” In this paper, we call these 5 types of a clause boundary the **weak clause boundary**. The weak clause boundary becomes the point into which linefeeds are inserted, if there does not exist the strong clause boundary around it. Furthermore, the clause boundary “condition clause-*to*” and “compound clause-*te*” tend to become the point into which linefeeds are inserted although the tendency is not as great as as that of the strong clause boundary and weak clause boundary.

3.3.2 Insertion of linefeeds based on dependency relations

A dependency relation is a modification relation in which a modifier bunsetsu depends on a modified bunsetsu. A sequence of bunsetsus from the modifier bunsetsu to the modified bunsetsu constitutes a semantically meaningful unit. Therefore, linefeeds tend to be inserted into the end boundaries of modified bunsetsus although the tendency is not greater than that of clause boundaries.

The bunsetsu boundaries, into which linefeeds are most easily to be inserted among the linefeed locations based on dependency relations, the bunsetsu boundaries into which linefeeds are easiest to be inserted are the end boundaries of modified bunsetsus of adnominal clauses. Here, in Japanese, the rightmost morpheme of an adnominal clause is congruent with that of a sentence end. Since, if a linefeed is inserted into the end boundary of adnominal clause, the end of the line is misunderstood as a sentence end, a linefeed is inserted not there but into the end boundary of the modified bunsetsu of an adnominal clause.

Since the clause boundary "topicalized element *wa*" does not strictly represent clause boundaries but can be regarded as syntactically independent elements, it is the dominant candidate of the linefeed points. However, in case that the number of characters in the clause boundary "topicalized element *wa*" which appears at the start of a sentence is few like "(this)," it is not appropriate to insert a linefeed into the boundary. If the length of the character string between the start of a line and the clause boundary "topicalized element *wa*" is long to some extent, a linefeed tends to be inserted into the clause boundary "topicalized element *wa*."

In addition, the dependency structure of a line displayed as a caption tends to be closed. That is to say, all bunsetsus, except the final bunsetsu, in a line tend to depend on one of bunsetsus in the line. Conversely, a linefeed tends to be inserted into the end boundary of the modified bunsetsu of which the dependency distance is long.

4 Automatic insertion of linefeeds

In our method, a sentence, on which morphological analysis, bunsetsu segmentation, clause boundary analysis and dependency analysis are performed, is considered the input. Our method outputs the sentence into which linefeeds are inserted. The insertion of linefeeds is executed as follows by using the rules for deciding the point into which linefeeds should be inserted.

We made the rules for inserting linefeeds based on the analysis described in the previous section. Table 2 shows the rules. The each rule number indicates the priority order in which each rule is applied, and the application of rules is performed in accordance with the priority order until the length of all lines becomes less than or equal to the maximum number of characters per line. The first rule is for the boundaries into which linefeeds are not inserted, and the second is for inevitable insertion. Furthermore, the rules 2-4 are for the insertion based on clause boundaries, 5-9 are on dependency relations, and 10 is on the number of characters of line.

Figure 4 shows the processing flow of linefeed insertion. The candidates of points into which linefeeds are inserted are denoted by a slash "/." First, the end boundaries of the bunsetsus which are the boundaries into which linefeeds are not inserted are excluded from the candidates of linefeed points. Next, a linefeed is inserted into the end boundary of the bunsetsu "(because they have already been subjected)," which

Table 2: Ten rules for inserting linefeeds

	rule
1	Exclude bunsetsus boundaries into which linefeeds are not inserted from the candidates.
2	Insert into the strong clause boundary.
3	Insert into the weak clause boundary.
4	Insert into the clause boundary "condition clause-*to*."
5	Insert into the clause boundary "compound clause-te."
6	Insert into the end boundary of a modifier bunsetsu of "adnominal clause."
7	Insert into the clause boundary "topicalized element-*wa*," if the number of characters between the start of a line and"*wa*" is over than 70% of the maximum number of characters per line.
8	Insert into the end boundary of the **long dependency distance bunsetsu** in case that there exits only one **long dependency distance bunsetsu**. Within the maximum number of characters from the start of the line.
9	Insert into the end boundary of the leftmost bunsetsu among **long dependency distance bunsetsus** of which the modified bunsetsu is different from that of the next one.
10	Insert into the rightmost bunsetsu boundary within the maximum number of characters.

*A bunsetsu which is located within the maximum number of characters from the start of the line and which depends on a bunsetsu located outside the maximum number of characters from the start of the line is called **long dependency distance bunsetsu**.

is the strong clause boundary "reason clause-*node*." As mentioned above, the rules for linefeed insertion are applied in accordance with the priority order. The texts of the caption are finally generated so that the length of each line is less than or equal to the maximum number of characters per line.

5 Experiment

To evaluate the effectiveness of our method, we conducted an experiment on inserting linefeeds by using Japanese spoken monologue data.

5.1 Outline of experiment

We used 3 programs (219 sentences, 2,121 bunsetsus) in the syntactically annotated spoken monologue corpus "Asu-Wo-Yomu" as the test data, annotated with information on morphological analysis, clause boundary detection, and dependency analysis by hands.

We applied our method to the test data. In addition, we compared our method with the baseline one, which inserts linefeeds into the rightmost bunsetsu boundary among the bunsetsu boundaries into which linefeeds can be inserted so that the length of the line does not exceed the maximum number of characters. We obtained the precision (the ratio that the points of the inserted linefeeds correspond with the

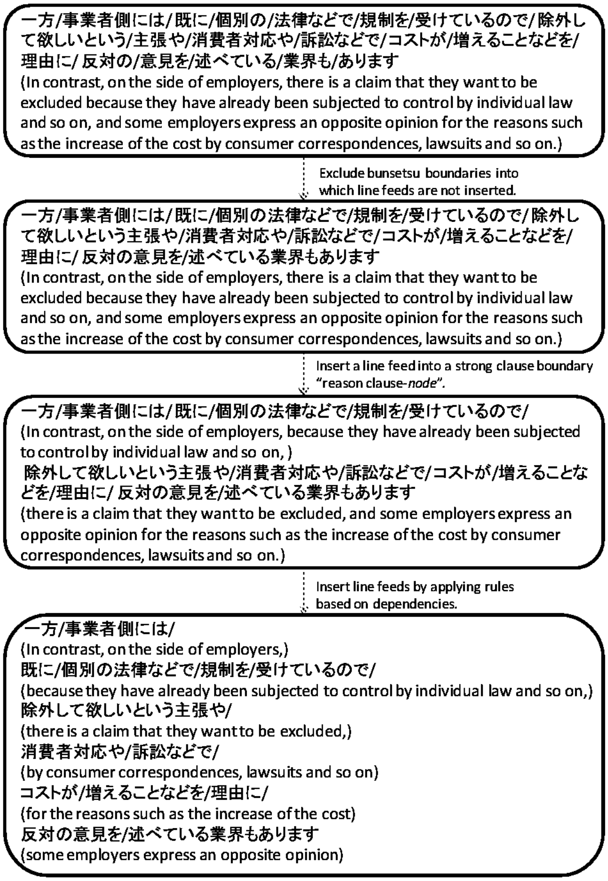


Fig. 4: Processing flow of linefeeds insertion

correct ones) and the recall (the ratio that linefeeds were inserted into the correct ones).

Two experts collectively created the correct1 data. They created the data independently and then they decided the correct data through the consultation based on each data. We considered that there are other acceptable linefeed points other than those of the correct data. Therefore, we also obtained the precision in case that the union of the linefeed points in the two data which two experts created independently is considered to be the correct points (hereinafter called **acceptable precision**).

Table 3: Experimental results

	our method	baseline
recall	82.7% (426/515)	30.1% (155/515)
precision	79.0% (426/539)	37.9% (155/409)

Table 4: Causes of incorrect linefeed insertion

causes	#
exclusion of bunsetsu boundary from the candidates	10
insertion into the strong clause boundary	6
insertion into the weak clause boundary	9
insertion based on “adnominal clause”	1
insertion based on “topicalized element *wa*”	10
insertion based on dependency distance	24
others	11
total	71

5.2 *Experimental results*

Table 3 shows the experimental results. The recall and precision were 82.7% and 79.0% respectively, and we confirmed that our method had higher performance than the baseline method. Furthermore, the acceptable precision was 86.8% (468/539). As mentioned above, we confirmed the effectiveness of our method.

Figure 4 shows the causes of incorrectly inserting linefeeds. The largest cause is the line feed insertion based on the dependency distance. There were a lot of cases that the length of a line becomes short by inserting a linefeed into the end boundary of a bunsetsu which depends on a distant bunsetsu. We need to establish detailed rules based on not only the dependency distance but also the number of characters in a line and the morphological information such as the part-of-speech of a particle.

On the other hand, one of the reasons for not inserting linefeeds into the correct linefeed points is the existence of the clause boundaries which did not appear in the analysis data. Figure 5 shows an example. In this example, the clause boundary “condition clause-*nara*h becomes correct linefeed point. However, there did not exist the clause boundary “condition clause-*nara*.h Therefore, the linefeed was incorrectly inserted into the clause boundary “compound clause-*te*.” Since the current rules are not covered enough, we need to increase the size of the learning data.

6 Conclusions

This paper proposed a method for inserting linefeeds into Japanese monologue sentences to support the understanding of monologue speech by the deaf persons, elderly persons or foreigners. Our method can insert linefeeds so that captions become

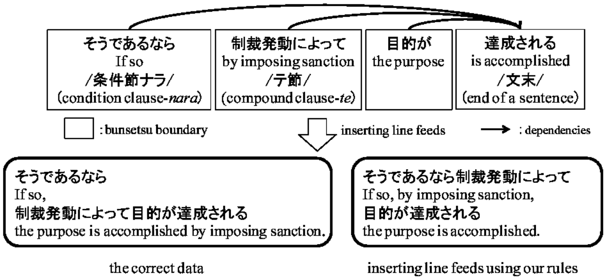


Fig. 5: Example of the clause boundary which did not appear in the analysis data

easy to be read by applying the rules which are established based on the emerging pattern of morphemes, dependencies, clause boundaries, pauses, fillers and so on. An experiment on inserting linefeeds by using monologue corpus showed the recall and precision was 82.7% and 79.0%, respectively, and we confirmed the effectiveness of our method.

To make the rules linefeeds by hands has limitations in enlarging, refining and organizing them. Future research will include considering the automatic acquisition of the rules. In addition, we will plan to reveal the linguistic relation between the points into which linefeeds should be inserted on captioning and the points into which pauses should be inserted on speech synthesis [4].

Acknowledgements. This research was partially supported by the Grant-in-Aid for Scientific Research (B) of JSPS and by The Telecommunications Advancement Foundation.

References

1. Boulianne, G., et al.: Computer-Assisted Closed-Captioning of Live TV Broadcasts in French. In: Proceedings of ICSLP 2006, pp. 273–276 (2006)
2. Daelemans, W., Hothker, A., Sang, E.T.K.: Automatic sentence simplification for subtitling in Dutch and English. In: Proceedings of LREC 2004, pp. 1045–1048 (2004)
3. Holter, T., Harborg, E., Johnsen, M.H., Svendsen, T.: Asr-based subtitling of live TV-programs for the hearing impaired. In: Proceedings of ICSLP 2000, pp. 570–573 (2000)
4. Iwata, K., Mitome, Y., Watanabe, T.: Pause Rule for Japanese Text-To-Speech Conversion Using Pause Insertion Probability. In: Proceedings of ICSLP 1990, pp. 837–840 (1990)
5. Monma, T., Sawamura, E., Fukushima, T., Maruyama, I., Ehara, T., Shirai, K.: Automatic Closed-Caption Production System on TV Programs for hearing-Impaired People. Journal of Systems and Computers in Japan, 71–82 (2003)
6. Ohno, T., Matsubara, S., Kashioka, H., Kato, N., Inagaki, Y.: A Syntactically Annotated Corpus of Japanese Spoken Monologue. In: Proceedings of LREC 2006, pp. 1590–1595 (2006)

7. Saiko, M., Takanashi, K., Kawahara, T.: Cascaded Chunking of Spontaneous Japanese using Bunsetsu Dependency and Pause Information. IPSJ SIG Technical Reports, pp. 19–24 (2006) (in Japanese)
8. Xue, J., Hu, R., Zhao, Y.: New Improvements in Decoding Speed and Latency for Automatic Captioning. In: Proceedings of ICSLP 2006, pp. 1630–1633 (2006)

An Ad-hoc Unicursal Protocol for Human Communication in Disaster Situations

Koichi Asakura, Junya Oishi, and Toyohide Watanabe

Abstract In this paper, we propose an ad-hoc unicursal protocol for communication systems in disaster situations. This protocol is designed for gathering information of rescuees effectively in a big earthquake. Since the longevity of networks is one of the most important aspects for such situations, terminals are connected to each other in linear form in our protocol, which makes the communication loads among terminals relatively equal in comparison with other structured networks. Experimental results show that our protocol needs less packets for generating an information sharing network and enlarging the shared information.

1 Introduction

In disaster situations such as an earthquake, a big fire and so on, information sharing is important for both rescuers and refugees. Especially, it is essential that the newest information has to be gathered and provided continuously. For example, for rescuers, correct and newly updated information about disaster areas and casualties: namely where and how much casualties exist, is very helpful for making rescue plans appropriately.

In order to gather and share information in disaster situations, it is necessary to develop a robust communication system. However, it might happen frequently that a communication system cannot be used in disaster situations although it is fully functional at ordinary times. For example, in wireless communication systems such

Koichi Asakura
School of Informatics, Daido University, 10-3 Takiharu-cho, Minami-ku, Nagoya, 457-8530 Japan, e-mail: asakura@daido-it.ac.jp

Junya Oishi and Toyohide Watanabe
Department of Systems and Social Informatics, Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8603 Japan
e-mail: oishi@watanabe.ss.is.nagoya-u.ac.jp,
watanabe@is.nagoya-u.ac.jp

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 511–521.
springerlink.com

as cellular phones and WiFi networks, communication infrastructures such as base stations and access points are broken by earthquakes, which causes the malfunctions of the whole systems.

Under the above circumstance, we aim to develop a communication system which enables to gather information about rescuees rapidly in disaster situations. In this paper, we propose a communication protocol based on ad-hoc network technology for disaster situations. In our protocol, terminals form information sharing network in ad-hoc network manner without any communication infrastructures. Information about rescuees is managed, shared in the network, and sent to refugees appropriately when refugees are moving near rescuees. This protocol aims to make power consumption equipped with each terminal lower than other existing protocols. In our protocol, a unicursal path is adopted for information network of rescuees. The unicursal path makes it possible to balance the communication loads among rescuees. Also, intentional delay of packet sending is adopted in order to decrease the number of packets. We can achieve long-lived information sharing networks and effective information delivering by these methods.

The rest of this paper is organized as follows. Section 2 expresses requirements for a communication system. Section 3 describes our proposed protocol called an ad-hoc unicursal protocol. Section 4 describes the experimental results. Section 5 concludes this paper and gives our future work.

2 Preliminaries

In this section, requirements of a communication system for disaster situations are described.

2.1 Requirements for Communication Systems

In disaster situations such as a big earthquake, we have two types of objects from the viewpoint of communication systems. One is a static object like a rescuee who cannot move in a collapsed building. Another is a moving object like a refugee who moves to the nearest shelter. Information about rescuees is important for making initial rescue plan correctly and quickly[1, 2]. Thus, it is essential to gather information about rescuees rapidly for communication systems in disaster situations.

Under the above disaster situation, it is clear that technology of ad-hoc network is mandatory for communication systems[3, 4]. This is because the ordinary communication systems based on infrastructures cannot be used fully in disaster situations. In the assumed communication systems, information shared among rescuees and refugees is not updated so dynamically. Since rescuees cannot move in collapsed buildings, this type of information is rarely updated. Thus, in this network, frequency of packet sending can be reduced without disadvantage.

For rescuees, the most important point is that information is delivered to refugees successfully. In order to provide information to refugees effectively, it is clear that each rescuee make an effort to share the information among neighboring rescuees rather than to deliver the information individually. This is because rescuees never deliver their information if they cannot contact with moving refugees directly. Thus, rescuees have to generate information sharing network among neighboring rescuees and share their information periodically.

For constructing networks, it is also important to take longevityness of network into consideration[5, 6]. A packet flooding method may be adequate from the viewpoint of robustness of information delivery. However, the flooding method increases communication cost, which causes the short liveness of information network and terminals themselves. The objective of information network for rescuees is to deliver information to refugees unfailingly as well as to share their information among rescuees. Thus, it is important to make the age of networks long in order to increase the chance to contact with refugees.

In order to achieve long-lived networks, load balancing among terminals has to be also taken into consideration. In networks with complicated topologies, communication load of terminals may be distributed unequally. For example, in tree structured networks, a root terminal takes more important role for maintaining network in comparison with leaf terminals. This causes imbalance of communication load. If battery exhaustion occurs in the root terminal, the whole network becomes out of work and additional network construction effort has to be performed, which leads excess power consumption of terminals in the network. At worst, information sharing is never achieved. Thus, for constructing sharing network for rescuees, equality of communication loads of rescuees has to be taken into account as well as constructing an effective network topology.

2.2 Unicursal Networks

In order to satisfy requirements described above, we propose a unicursal network. In unicursal networks, terminals are connected to each other in linear form as many as possible[7, 8]. Namely, in order to prevent branches in the network, terminals do not connect with more than two terminals. As mentioned above, in tree-structured networks, the quantity of communication loads depends on terminal positions in the network. On the other hand, in unicursal networks, communication loads for maintaining the networks can be balanced except for edge terminals. This load balanceness leads to the longevity of the network. Figure 1 shows examples of the unicursal network and the tree-structured network.

In unicursal networks, the maximum number of hops tends to be large in comparison with tree networks. If the number of terminals in the network is denoted as n, the maximum number of hops is proportional to n in unicursal networks, while it is proportional to $\log(n)$ in tree networks. The number of hops affects elapsed time of information delivery. However, as mentioned in Section 2.1, information

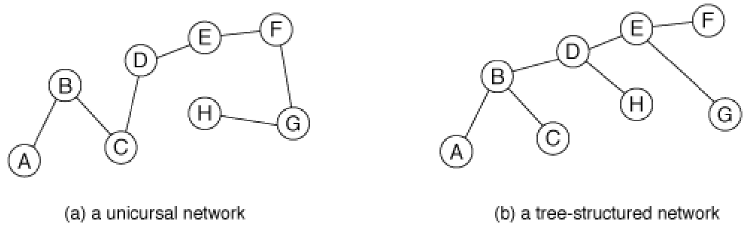


Fig. 1: Network structures

shared in the rescuee network is not changed so frequently due to immobility of rescuees. Thus, the influence of elapsed time poses little problem for information sharing among rescuees.

Constructing a unicursal network is reduced to constructing a Hamiltonian path in graph theory[9]. In the next section, we describe a protocol in which terminals generate a Hamiltonian path cooperatively in ad-hoc network manner.

3 Ad-hoc Unicursal Protocol

An ad-hoc unicursal protocol consists of two phases: a route constructing phase and an information sharing phase. In the route constructing phase, a Hamiltonian path is generated by autonomous communication of neighboring terminals. In the information sharing phase, information of rescuees in the same network is delivered to the terminals.

3.1 Data Structures

In order to generate a unicursal route, each terminal has the following four kinds of terminal lists in our protocol.

Neighboring terminal list This list consists of the terminals which can be communicated directly.

Connecting terminal list This list consists of the neighboring terminals next to this terminal in a unicursal network. In unicursal networks, up to two terminals are managed in this list basically. Obviously, the terminals in this list also exist in the neighboring terminal list. However, the opposite is not always true.

Group terminal list This list records the member terminals in the network.

Unknown terminal list This list manages the terminals which are in the communication range but are not members of the network. Namely, neighboring terminals which are not in the neighboring terminal list are added to this list.

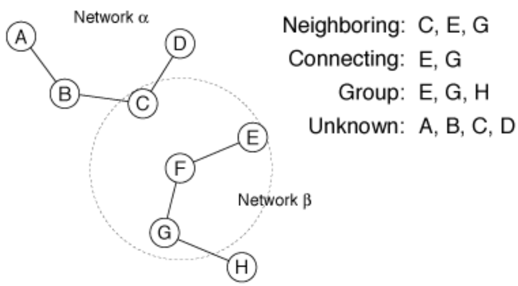


Fig. 2: Data structures

Figure 2 shows an example of these lists in the terminal F.

3.2 Unicursal Route Construction

For route construction, the following three functions are required: a group join, a group merge and a route restructure.

Group Join

When a terminal begins to join into the network of rescuees, the terminal sends a request packet (REQ) to neighboring terminals. Terminals, which receive the REQ packets, send back reply packets (REP) to the destination terminal. A REP packet includes the number of connecting terminals. In the destination terminal, the terminal which has the lowest number of connecting terminals is selected as the connection terminal. The selected terminal is added to the connecting terminal list and an acknowledge packet (ACK) is sent to the terminal.

Figure 3 shows an example of a group join. In this figure, the terminal E begins to join the network. Since three terminals are in the communication range, the REQ packet is delivered to these terminals. In this situation, the terminal F is selected for the connection terminal. The terminal E sends an ACK packet to the terminal F, and both terminals update the connecting terminal lists.

If no terminals exist within the communication range, the terminal generates a group itself. This isolated group may be merged to another network when other networks are generated near the terminal.

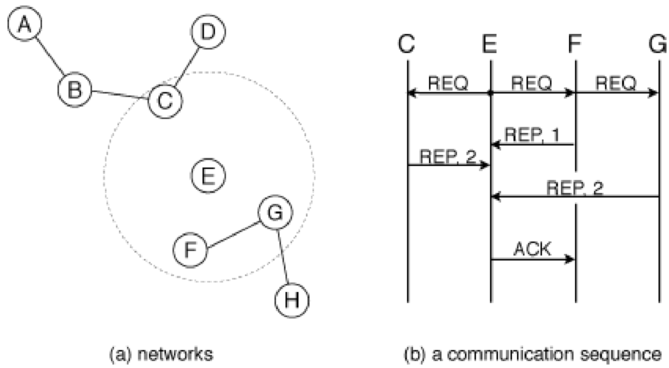


Fig. 3: A group join

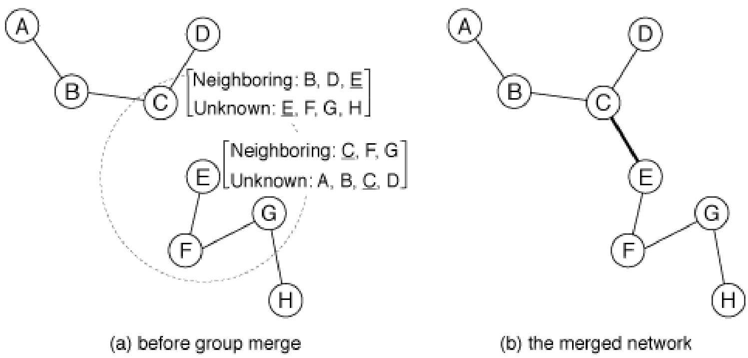


Fig. 4: A group merge

Group Merge

In this function, two network groups are merged into one network. If the following two conditions are satisfied in a terminal, two networks can be merged:

- a terminal exists in the unknown terminal list for a certain period and
- the above terminal also exists in the neighboring terminal list.

If such terminal exists in the lists, a merge request packet (MREQ) is sent to the candidate terminal. The terminal which receives the MREQ packet also verifies the lists whether the above conditions are satisfied or not. If satisfied, a merge reply packet (MREP) is sent back to the terminal. Both terminals update their connecting terminal lists and unknown terminal lists.

Figure 4 shows a workaround of this function. In this situation, both the terminals C and E satisfy the conditions and the two networks are merged.

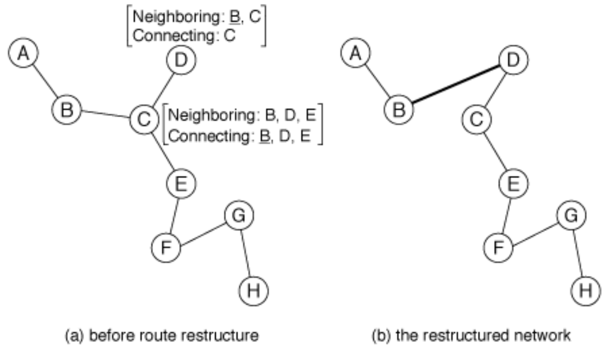


Fig. 5: A route restructure

Route Restructure

In the unicursal protocol, a route is constructed so that the number of connecting terminals does not exceed two. However, a terminal which has more than two connecting terminals may occur temporarily during the group merge function. A route restructuring function is invoked in such terminal. The terminal sends a list request packet (LREQ) to the surrounding terminals in order to collect the neighboring terminal lists and the connecting terminal lists. If the following conditions are satisfied, a route restructure is performed:

- difference of the number of connecting terminals for two terminals is two or more and
- the same terminal exists in both the neighboring terminal list of the invoked terminal and the connecting terminal list of the surrounding terminal.

Figure 5 shows an example of a route restructure. In this network, the terminal C has three connecting terminals and thus the route restructuring function is activated in this terminal. In this network, the terminal D satisfies the above conditions. This is because the terminal D has only one connecting terminal and can communicate with the terminal B. Thus, the connection between the terminals B and C is canceled and new connection between the terminals B and D is generated. Figure 5(b) shows the new network with no branches.

3.3 *Information Sharing with Intentional Delay*

Information about rescuees is shared thorough the network described in Section 3.2. As described in Section 2.1, the frequency of information update is relatively low. Thus, in order to suppress communication cost, a method for intentional delay of packet sending is introduced. If a data update packet (DATA) is received from

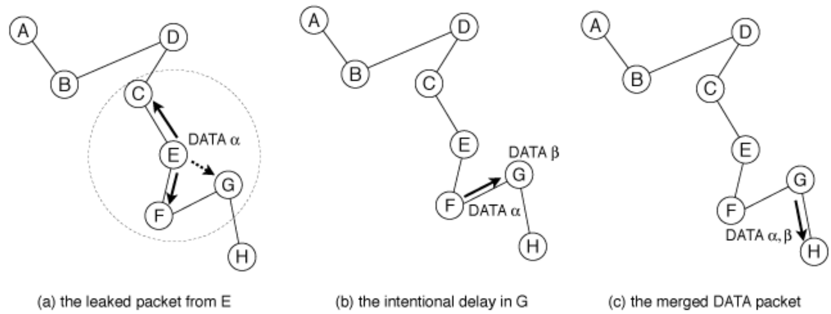


Fig. 6: Intentional delay for data transfer

terminals except for connecting terminals, this terminal stores this packet and suspends sending its DATA packet. This is because there is a possibility that the packet merging is achieved. When the same packet as the leaked packet is received from the connecting terminals, DATA packets are merged into one packet, which decreases the frequency of DATA packet transfer.

Figure 6 shows this mechanism of intentional delay. When the terminal E sends the DATA packet α, the terminal G also receives the packet as the leaked packet. Then, the intentional delay is activated in the terminal G. The terminal G postpones sending its DATA packet β until the packet α is received from the connecting terminals. When the packet α is received, the terminal G merges two packets and sends it to the connecting terminals.

4 Experiments

In order to evaluate the amount of communication and equality of communication load among terminals, we conduct simulation experiments. In the simulation, we assume that rescuees cannot move and share their location information. We use a 600m-squared field with no obstacles. The communication range of each terminal is defined as 100m. In the route constructing phase, the number of packets is measured when terminals appears in the field one by one. In the information sharing phase, the number of packets and elapsed time for spreading information of all rescuees are measured. In the experiment for information sharing phase, each terminal sends its information at one time and at random timing. We use ACP (Ad-hoc Community Protocol)[10] and AODV (Ad-hoc On-demand Distance Vector) multicast protocols[11, 12] as competitors. Our proposed protocol is denoted as AUP (Ad-hoc Unicursal Protocol).

Table 1: The number of packets in the route constructing phase

Terminals	25	50	75	100
AUP	135	584	2,566	4,461
ACP	547	3,213	13,614	26,114
AODV	454	2,005	5,538	10,696

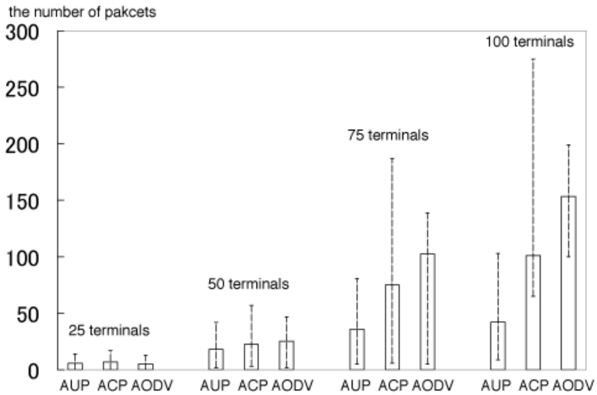


Fig. 7: The number of packets in the information sharing phase

Table 2: The results in the information sharing phase

(a) Standard deviation of packets

Terminals	25	50	75	100
AUP	4.6	11.6	19.0	18.7
ACP	5.5	16.0	36.3	43.1
AODV	4.1	17.1	41.8	49.3

(b) Average delay time

Terminals	25	50	75	100
AUP	2.5	26.9	118.2	329.5
ACP	2.8	6.3	11.6	13.5
AODV	2.1	7.6	16.2	18.1

4.1 Experimental Results

Table 1 compares the number of packets in the route constructing phase. This result declares that our proposed protocol works with less packets in all situations. This is because a route construction is achieved by local communication of terminals in our protocol.

Figure 7 shows the average number of packets for information sharing. This graph also shows the maximum and the minimum number of packets by dashed lines. In addition, Table 2(a) presents the standard deviation of the number of packets in order to show the equality of communication loads, and Table 2(b) shows the average elapsed time of information sharing. From the result of the average number of packets, we can clarify that our protocol makes it possible to share information of rescuees in the low communication cost. As the number of terminals increases, this trend also increases. This is because intentional delay is effective if the number of terminals becomes large. However, due to the intentional delay, the elapsed time for information delivery in the whole network is very large in comparison with other protocols. As shown in Section 2.1, information shared in the unicursal network is not changed so dynamically. Thus, this delay does not take effect to the communication systems in disaster situations. Instead, the durability of networks has to be paid attention to. Thus, we can conclude that our protocol decreases the number of packets and equalizes the communication load among terminals, which leads the longevity of networks.

5 Conclusion

In this paper, we propose an ad-hoc unicursal protocol for communication systems in disaster situations. This protocol takes the low and balanced communication cost among terminals into consideration. Experimental results show that the number of packets in our protocol is small in comparison with other protocol in both the route constructing phase and the information sharing phase. This protocol is suitable to human interaction in disaster situations.

For our future work, we have to pay much attention to the power consumption in terminals. We must provide a precise power consumption model and develop a power-aware communication protocol.

Acknowledgement

This research has been supported by the Kayamori Foundation of Informational Science Advancement.

References

1. Sugiyama, H., Tsujioka, T., Murata, M.: Victim Detection System Consists of Networked Mobile Robots. Trans. on Information Processing Society of Japan 46(7), 1777–1788 (2005) (in Japenese)
2. Urabe, H., Tsukamoto, J., Sato, K., Umedu, T., Higashino, T.: A MANET Based System for Gathering Location and Personal Information from Victims in Disasters. In: IEICE Tech. Report on ITS, vol. 2005, pp. 47–52 (2005) (in Japanese)
3. Toh, C.-K.: Ad Hoc Mobile Wireless Networks: Protocols and Systems. Prentice Hall, Englewood Cliffs (2001)
4. Mase, K.: Communications Supported by Ad Hoc Networks in Disasters. Journal of the Institute of Electronics, Information, and Communication Engineers 89(9), 796–800 (2006) (in Japenese)
5. Morita, K., Sato, Y., Okazaki, N., Park, M.-R.: Hop-vector Based Greedy Forwarding for Sensor Networks. In: IPSJ SIG Notes on Groupware and Networks, vol. 2007, pp. 171–178 (2007) (in Japanese)

6. Ono, M., Higaki, H., Furuta, K.: Client Participated Model of Sensor Networks for Sensor Databases. In: IEICE Tech. Report on Information Networks, vol. 2007, pp. 339–344 (2007) (in Japanese)
7. Higa, Y., Furukawa, H.: A Highly Efficient Packet Forwarding Scheme for Wireless Multihop Networks with String Topology. Trans. on IEICE japan J90-B(12), 1225–1238 (2007) (in Japanese)
8. Momcilovic, P., Squillante, M.: On Throughput in Linear Wireless Networks. In: Proc. of the 9th ACM Int'l Symp. on Mobile Ad Hoc Networking and Computing, pp. 199–208 (2008)
9. Diestel, R.: Graph Theory. Springer, Heidelberg (2005)
10. Ota, K., Machida, M., Otsuji, K., Sugimura, T.: Group Communication Protocol for Community on Ad-hoc Networks. Trans. of Information Processing Society of Japan 42(7), 1801–1810 (2001) (in Japanese)
11. Royer, E., Perkins, C.: Multicast Operation of The Ad-hoc On-demand Distance Vector Routing Protocol. In: Proc. of the 5th Annual ACM/IEEE Int'l Conf. on Mobile Computing and Networking, pp. 207–218 (1999)
12. Bae, S., Lee, S.-J., Su, W., Gerla, M.: The Design, Implementation, and Performance Evaluation of The On-demand Multicast Routing Protocol in Multihop Wireless Networks. IEEE Network 14(1) (2000)

Geographic Information Management, Based on Version Control System *Subversion*

Kazuya Hirobe and Toyohide Watanabe

Abstract In order to use geographic data flexibly in the temporal GIS, two conditions are required. One is a revision independency to avoid link complexity. The other is an index structure to be accessible to geographic data easily. In this paper, a data management structure of version control system *Subversion* is addressed. We propose a method of describing difference information without concern to the data structure and a method to manage revision information in GIS by a HR-tree structure. We manage the geographic information in the university campus. The geographic information in the map is updated by putting the lusterware image rectangle.

1 Introduction

Until today, geographic information systems (GISs) have been used in various fields such as economics, city administration, etc., and take important roles as the social information infrastructure systems. In these fields, many researches have focused on the successive manipulate of spatial change among geographic objects. However, users have been overloaded with maintaining geographic objects on a traditional temporal GIS. In a temporal GIS, needs to maintain the connectivity with surrounding objects in near future and near past. Therefore the amount of information on connectivity becomes large. Our research objective is to establish a method of managing the geographic objects independent of the data structure in the temporal GIS. In this paper, we propose the method of describing a revision which is independent of the data structure. Additionally, we propose an index structure to

Department of Systems and Social Informatics,
Graduate School of Information Science, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan
e-mail: `hirobe@watanabe.ss.is.nagoya-u.ac.jp`,
`watanabe@is.nagoya-u.ac.jp`

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 523–533.
springerlink.com

retrieve any version of map rapidly. For this purpose, we refer to the data management structure of "Subversion" which is a version control procedure for a file system. For geographic objects, we use the structure of "Subversion" and HR-tree as a spatial indexing structure. A difference without depending on the data management structure makes possible to extract revision in a temporal GIS. In addition, it makes easy to exchange revision between two or more systems. We assume that geographic objects in the map are updated by putting a newer geographic object with the newer date. The geographic object is presented by a rectangular lusterware image with the area. A university campus map is a candidate of experimentation. We are not concerned here with route information like the road. We address the approach to manage the version history of geographic objects in Section 2. In Section 3, the management structure and the difference description method in the version control system *Subversion* are introduced. In Section 4, we propose external delta which is the abstract difference notation and internal delta which represent a concrete change of the internal data structure. And the technique for converting external delta into internal delta in is described in Section 5. In Section 6, an experimental result is shown. The summary and future issues are explained in Section 7.

2 Approach

In order to use the geographic object independently in the temporal GIS, a method of describing a revision independency is required. Obviously, geographic objects have to be accessible easily. For that reason, we refer to *Subversion*. *Subversion* is a version control procedure for the file system. To manage revisions of directory structures simply, *Subversion* achieves revisions of directory tree with data structure like an overlapping method [4]. Moreover, Subversion does not describe identifier of the node in revision information, but the revision is described with the revision number and the string of the directory path. In this paper, an internal data is managed with HR-tree which applies the idea of overlapping tree to R-tree. And the difference of revision is described using the tree delta which is the difference notation of directory structure in Subversion. Subversion uses the directory path string for the node specification. We use the MBR(Minimum Bounding Rectangle) for the node specification in HR-tree. This notation is defined as an internal delta. The internal delta makes the node specified without node identifier. An internal delta concretely describes the change HR-tree. This means user has to be aware of internal nodes in HR-tree. This is not appropriate to the difference description that does not depend on the data management structure. Then, we define an external delta as a notation to describe an abstract difference of geographic objects and we translate the external delta into the internal delta. The external delta only describes the lusterware image that should be put and the location where image should be put. Fig. 1. shows the overall architecture. An external delta is a difference description that does not depend on the structure of a database. An internal delta is generated based on the node structure in the database and external delta description. An internal delta is a concrete difference description including the operation of the data structure within a database. The temporal GIS database is renewed by an internal delta. Geographic objects can be referred to concisely by HR-tree structure. As a result, the temporal GIS can describe differences which do not depend on the data management structure, and can be referred to at high speed.

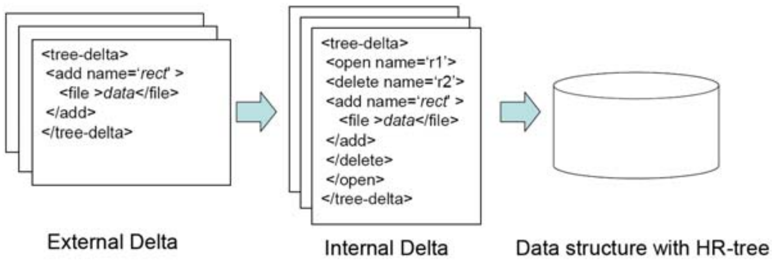


Fig. 1: Overall process of translation.

3 *Subversion* Data Structure

Subversion uses three types of difference definitions. In this research, we pay attention to tree-delta.

- Tree-delta: describe the difference between two arbitrary directory trees.
- File-delta: describe the difference between two contents of files.
- Property-delta: describe the difference between two lists of attributes beyond their contents.

3.1 Data Management Structure in Server

Subversion has the management structure based on overlapping tree. As a result, the file tree of specified revision can be easily acquired by searching the node from the root node of specified revision. *Subversion* structure has the feature that the parent node maintains the file name of the node. This means that the operation without changing the content of the file affects only to the parent node. For example the rename operation and the copy operation affect to parent node only. Fig. 2. shows the node when the content of file "D.txt" is changed in new revision. First, the leaf node of "D.txt" is generated. Next, the parent node is generated as a copy of revision until reaching the root node (folders A and C). Then, the link is generated to the node in previous revision which is not changed (folder B). Finally, the root node is linked with the latest revision number (revision 2).

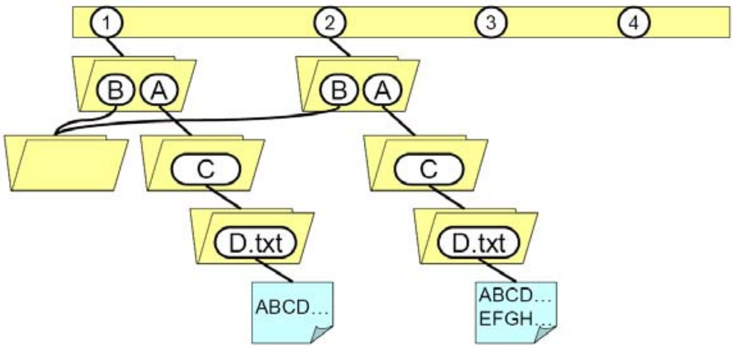


Fig. 2: Change contents of file “D.txt”.

3.2 Tree Delta Description in Subversion

We explain the notation of the tree delta in *Subversion*. The notation of the tree delta in *Subversion* is a hierarchical expression of change functions which should be interpreted one by one. Table 1 shows the component of the tree delta. The root element of the tree delta is *tree-delta*. *Tree-delta* has two or more *changes*. *Change* is change operation which is either of *open, add* or *delete*. *Open* shows that content is changed to described content, and add shows that content is generated. *Delete* shows that content is not exist in new revision. *Name* is an object file name or a directory name in new revision in each change operation. *Content* shows the content of the change, and is shown by either of *file* or *directory*. File shows the file, and has *ancestor, prop-delta,* and *text-delta* as an argument. *Ancestor* shows a directory path of content in previous revision. *Prop-delta* shows the property delta after it changed, and *text-delta* shows the file delta after it changed. *Directory* shows the directory, and has *ancestor, prop-delta,* and *tree-delta* as an argument. In *Subversion*, the example of describing the tree delta with XML is shown in Fig. 3. In this example, content text-delta of “file1” included in directory “dir1” has been changed. Text-delta presents the change of the file contents, and is described by the difference description form of the text delta. Then, the directory name of directory “dir2” included in directory “dir1” is changed to “dir3”.

```
<tree-delta><open name='dir1'>
  <directory><tree-delta>
    <open name='file1.txt'><file>text-delta</file></open>
    <delete name='dir2' />
    <add name='dir3' ><directory ancestor='/dir1/dir2' /></add>
  </tree-delta></directory>
</open></tree-delta>
```

Fig. 3: XML description of tree delta.

tree-delta(*change*$_1$,*change*$_2$, …*change*$_n$)	The root element of change operations.
change	An operation
open([*name*],[*content*])	An element by *name* has changed to *content*.
add([*name*],[*content*])	An element by *name* has created with *content*.
delete([*name*])	An element by *name* has deleted.
name	An file name or a directory name or path string
content	Content of element
file([*ancestor*],*prop-delta*, *text-delta*)	A file
file(*ancestor*)	
directory([*ancestor*], *prop-delta*, *tree-delta*)	A directory
directory(*ancestor*)	
prop-delta	Property delta
text-delta	Text delta
ancestor	Path string at pre-revision
(*path*, *new*)	*path* : Path string *new* : Boolean value whether file is new or not

Table 1: Elements of tree delta.

4 Data Management Structure in Temporal GIS

4.1 Data Management Structure in Server

The data management structure in the server is HR-tree. The non-leaf node maintains only MBR and node ID in HR-tree though the non-leaf node which shows the folder has a variety of attribute values in *Subversion*.

4.2 Internal Delta

We define the notation of an internal delta. The notation conforms to the tree delta in Subversion. However, the part which corresponds to the file paths, the file name, and the directory name of the changed node is changed to MBR. Moreover, the file node indicates the leaf node in HR-tree, and the directory node indicates the node that is not the leaf. The internal delta describes change functions which should be interpreted with a hierarchical expression.

- Revision:
- Source-mbr:
- Target-mbr:

In the root of the internal delta, source-mbr and target-mbr maintain all MBRs of the root node in the same way. Target-mbr maintains MBR of the generated node while searching for the tree delta, and source-mbr maintains MBR of the node that corresponds in a R-tree of previous revision. The internal delta is expressed as follows.

$$tree-delta([change-op_1, change-op_2, ..change-op_n]) \tag{1}$$

$Change-op_i$ shows a method of editing contents of source-mbr to target-mbr. Three kinds of changes exist. Non-specification of nodes in previous revision shows that the change does not occur in a new revision.

- *Open*
- *Delete*
- *Add*

Open is expressed as follows.

$$open([mbr],[content]) \tag{2}$$

Open changes both source-mbr and target-mbr to the node with *mbr*. Content shows file or directory to be changed.

Delete is expressed as follows.

$$delete([mbr]) \tag{3}$$

Delete shows that the node with mbr in source-mbr does not exist in target-mbr.

Add is expressed as follows.

$$Add([mbr],[content]) \tag{4}$$

Add shows that the node with mbr which has not existed yet in target-mbr exists in source-mbr now. *Content* is a file or a directory that a new directory entry will show. *Content* shows the content of change in the file or the directory. Two kinds of contents exist.

- File
- Directory

File is expressed as follows.

$$\begin{array}{c} file([ancestor], text - delta) \\ or \\ file(ancestor) \end{array} \tag{5}$$

File shows that current node of target-mbr is change of image. *Ancestor* is MBR of the node in previous revision. Text-delta is a lusterware image rectangle which changes based on ancestor. Abbreviation tree-delta shows that there is no change in the content of the lusterware image.

Directory is expressed as follows.

$$\begin{array}{c} directory([ancestor], tree - delta) \\ or \\ directory(ancestor) \end{array} \tag{6}$$

Directory shows that current node of target-mbr is change of non-leaf node. *Ancestor* is MBR of the node in previous revision. Tree-delta is a tree delta which shows how to construct directory which changes based on ancestor. Abbreviation tree-delta shows that there is no change in the content of child nodes. *Ancestor* is expressed in the file element and the directory elements as follows.

$$Ancestor = ([mbr], [new]) \tag{7}$$

Mbr shows that MBR of node which is new or changed in revision is *mbr* in previous revision. When this appears as an attribute of the file element, the text delta of the element is applied to *mbr*. When this appears as an attribute of the directory element, *mbr* is treated as new source-mbr in interpreting the tree delta of the element. *New* is a Boolean value which shows whether the node have new context or not. When *new* is true and has the text delta, the delta is applied to an empty node as a new node. When new is true and has the tree delta, the delta is evaluated as if source-mbr was an empty node. Abbreviation both *mbr* and *new* is the same as ancestor=('mbr',) with the same revision number.

5 *External Delta*

A structural change of R-tree is described in an internal delta notation, Although a structural change of R-tree is described in an internal delta notation, user and an external system cannot specify the node in R-tree at updating the map data. An external delta notation is defined for the update of a user and an external system.

Only *tree-delta, add,* and *file* notations are used in an external delta. Therefore, *tree-delta* is always composed in the following form.

Directory is expressed as follows.

$$\begin{gathered} Tree-delta(change-op_1, change-op_2, \ldots change-op_n) \\ change-op_i = add(mbr, file([ancestor], text-delta)) \end{gathered} \tag{8}$$

In an external delta notation, the operation to nodes in R-tree is not described. Moreover, the leaf node to which MBR is corresponding needs not exist in HR-tree.

6 Translation from External Delta to Internal Delta

Translation from an external delta to an internal delta is done as follows.

1. A trial revision tree is created along insertion algorithm of HR-tree.
2. New nodes are searched hierarchically from a trial revision root and change operation is written by depending on difference between a trial revision tree and previous revision tree.

Fig. 4. shows the example of generating an internal delta by addition of the object. In this example, the rectangle division of node C by addition of object J makes nodes H and I. In this case, an internal delta describes as follows.

1. Selection of node A'
2. Deletion of node C which is a child node of A
3. Addition of nodes H, I into node A'
4. Addition of nodes E, F, G, J into nodes H, I

7 Experiment

We recorded the geographic objects of Higashiyama campus in Nagoya University, Japan of 11 revisions from 1980 to 2001. The map on a brochure of the university was taken with the scanner. A revision 1, a whole map is delimited to 11 images as the base data. In R-tree, the child node upper bound concerning each node is 4, and the lower bound is 2. Fig. 5. shows the screenshot of experiment system. Fig. 6. shows the number of *nodes* in HR-Tree at each revision. In the tree used by the experiment, one image is counted as one node. Nodes are the cumulative sums of the number of total nodes in each revision. *Images* show the number of accumulations of images in each revision. *SearchNodes(Full)* shows the number of search nodes to display the entire map in each revision. *SearchNodes(1/4)* shows the number of search nodes to display the area of 25 percent in the entire map in each revision. *ExternalDelta* shows the number of nodes described with *ExternalDelta* in each

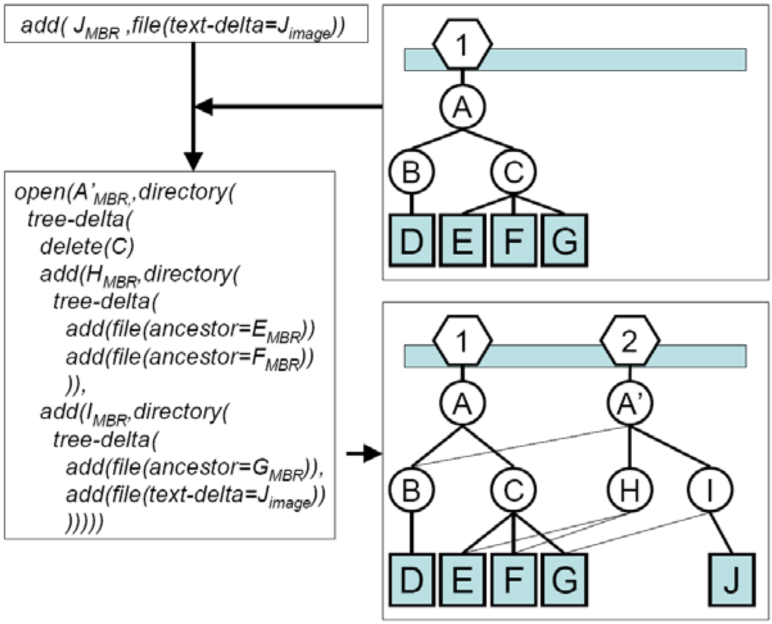


Fig. 4: Generating internal delta.

revision. The reason why the number of search nodes is increased as the revision increased is that the number of images to display increases. It is not influenced from past revision on the character of HR-tree. The amount of the description does not increase even if revision increased because the amount of the description in *ExternalDelta* is proportional to the number of images put in each revision.

References

1. Negishi, Y., Osawa, Y.: A spatio-temporal geographic information system based on implicit topology description: Stims. In: DMGIS 2001 (2001)
2. Collab .Net "Subversion Design", `http://subversion.tigris.org/design.html`
3. Nascimento, M.A., Silva, J.R.O.: Towards Historical R-trees. In: Proceedings of the 1998 ACM symposium on Applied Computing, pp. 235–240 (1998)
4. Burton, F.W., Kollias, J.G., Matsakis, D.G., Kollias, V.G.: Implementation of Overlapping B-Trees for Time and Space Efficient Representation of Collections of Similar Files. The Computer Journal 33(3), 279–280 (1990)

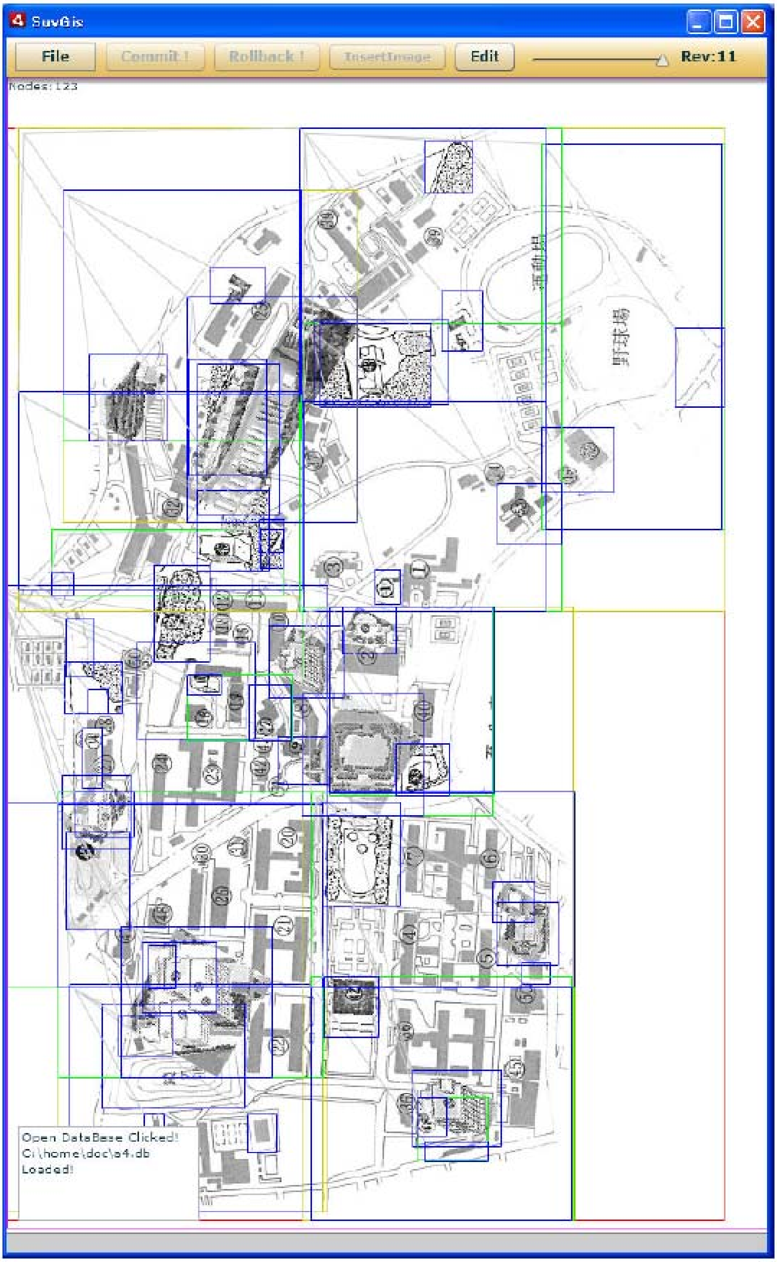


Fig. 5: Screenshot.

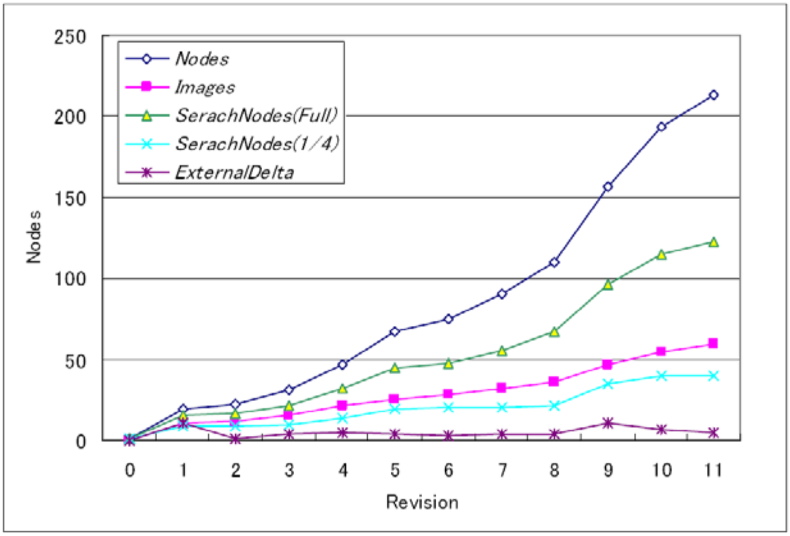


Fig. 6: Nodes per revision.

Multiuser Network Administration Training in LiNeS: Connection Function between Virtual Networks

Yuichiro TATEIWA and Takami YASUDA

Abstract Fostering network administrators for networks including Linux servers is essential. We developed a system to provide a Linux network administration training environment by exploiting User-mode Linux virtual machine software and called it the Linux Network Simulator (LiNeS). LiNeS works on a Linux PC and provides virtual networks consisting of virtual Linux servers, routers, switching hubs, and clients. LiNeS are designed to provide training environments where students administer virtual networks alone. However, the virtual networks in each student's PC are isolated: meaning they cannot telecommunicate with each other. This study developed a function that students practice by performing Linux network administration tasks with considering networks administrated by other students. In this paper, we describe how to interconnect each isolated network and discuss performance by preliminary evaluation experiences.

1 Introduction

Since computer networks are spreading in the world, fostering network administrators is essential. In Japan, the lack of Linux server administrators is emerging as a social issue because they cannot meet the demand of increasing facilities that introduce Linux servers; it is becoming crucial to foster administrators of networks including Linux servers which follow Linux networks. Actually, universities have

Yuichiro TATEIWA
Graduate School of Engineering, Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya-shi, 466-8555, Japan
e-mail: `tateiwa@nitech.ac.jp`

Takami YASUDA
Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601, Japan
e-mail: `yasuda@is.nagoya-u.ac.jp`

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 535–544.
springerlink.com

started to provide classes for learning the administration of Linux networks. However, students cannot practice sufficiently with enough network components because they are expensive. The number of such universities is not large, and some cannot provide classes in which students learn efficiently and effectively. Therefore, we developed a system for providing a Linux network administration training environment by exploiting Usermode Linux [1] (following UML) and call it the Linux Network Simulator (LiNeS) [2, 3, 4]. LiNeS runs on Linux PC and provides virtual networks whose maximum scale is about fifteen components that include Linux servers, routers, switching hubs, and clients, which are realized by UML. Hence, one student can practice with virtual Linux networks on one Linux PC in a practice room. In other words, LiNeS can easily provide classes for Linux network administration in most educational facilities. Our study developed a new function of LiNeS to interconnect virtual networks isolated in each PC. Earlier versions of LiNeS were designed to provide training environments where students construct virtual networks which are isolated on each PC. The virtual networks in each student's PC are isolated and cannot telecommunicate each other. The new function, however, makes it possible for students to administer their own networks in view of networks administered by other students, such as setting up a network by considering other networks and checking its behavior by cooperation with the others. The following training is enabled:

- setting TCP/IP parameters by considering other networks
- setting access control from other networks
- providing services in one's own networks with other networks
- using services of other networks

Since this function enables LiNeS to provide a network training environment that more closely resembles network forms in the real world, more practical education becomes possible.

2 RELATED WORKS

This section describes the originality of LiNeS by comparing with other systems that support exercises for setting up servers and constructing networks based on virtual machine technology. OVL [5] provides virtual machines to users by exploiting virtual environment software Xen [6] on a server, like a virtual hosting service. One user is given a virtual machine with administration privileges and performs programming training and server configuration training in the virtual machine with ssh remote operations. OVL provides environments for computer training in distance learning and is useful for TCP/IP configuration and server software setup in network administrator education. However, OVL is inadequate to train network administration consisting of multiple machines. Furthermore, because OVL requires highperformance servers for employing many Xen instances, OVL cannot satisfy our purpose that is easy to use in the existing facilities of PC practice rooms. A system developed by Nakagawa [7] provides environments for network construction training with virtual machines by VMware Workstation [8]. Students can construct networks consisting of virtual machines (Linux and FreeBSD) and Cisco routers. The system consists of PCs (each one employs VMware instances) and VLAN networks (connecting each PC). Nakagawa's system also has facility cost problems, such as Cisco routers (for VLAN) and VMware licenses. By contrast, LiNeS only requires general Linux PCs in PC practical rooms, and most universities have such facilities. Netkit [9] and VNUML [10] are designed to support management of virtual networks of UML. They construct virtual networks based on definition files written by users. The virtual networks are mainly useful for testing application software and server services. VIOLIN [11] constructs virtual networks consisting of UML over real networks. It has functions to comprise virtual networks of various topologies by assigning IP addresses to UML. In the networks, it is possible to perform network services, such as IP multicast and IP anycast used in video conferences and peer selection in P2P systems. These systems have the potential to help to practice network construction. However, they have no features to practice other fields covered by LiNeS such as network troubleshooting and network security.

3 NETWORK ADMINISTRATION TRAINING WITH ISOLATED NETWORKS IN LINES

In this section, we describe the training fields in LiNeS and the basic system structure for isolated virtual networks on one PC with LiNeS. We defined the core skills in university network administrator education (Fig. 1). Skills (1)-(4) are the aims of LiNeS. Skill (1) is used to construct LANs including Linux servers. Skill (2) is used for understanding relations between LAN construction skills and TCP/IP. LiNeS has a function for visualizing the behavior of virtual networks for learning such skills. Skill (3) identifies causes and resolves problems in abnormal networks. LiNeS provides virtual networks with problems for learning such skills. Skill (4) copes with network incidents in network security. LiNeS provides network incidents such as DOS attacks. A core function of LiNeS is to provide virtual networks and GUIs for operating them on one Linux PC by exploiting UML (Fig. 2). (1) is a GUI for a network topology, and (2) is the virtual components implemented by specifying UML. LiNeS provides a network environment that reduces costs, eases teacher burdens, and increases the learning effects of students for the above skills. Fig. 3 shows an execution example where a student constructed a virtual network in LiNeS. Window (1) is a GUI for constructing virtual networks. A student selects network components by pressing the desired buttons (2) and then sets the selected components in the blank area below by mouse. Window (3) is a console window for controlling a Debian/GNU Linux server (4). Window (5) controls client (6). Button (7) runs a web browser, and button (8) is used for running a mailer. Window (9) is a browser client (6) and is displaying the web page of server (4).

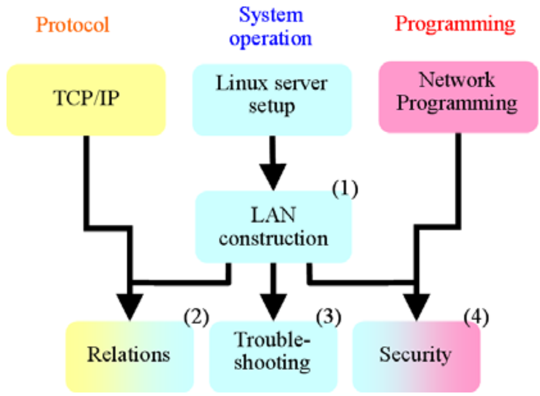


Fig. 1: Network administration skills.

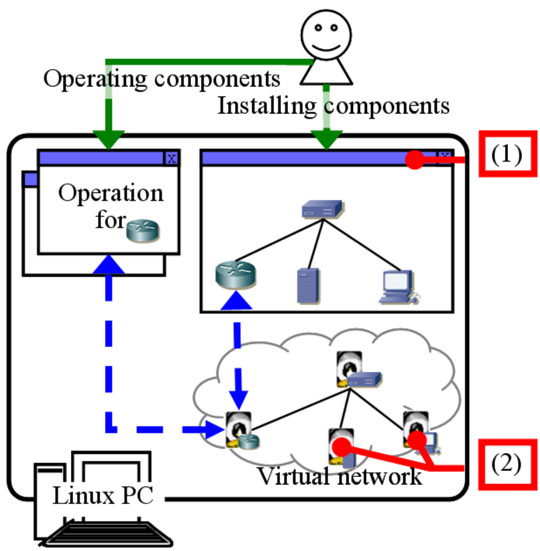


Fig. 2: Basic structure in LiNeS.

4 IMPLEMENTING NEW FUNCTIONS

In earlier versions of LiNeS, since students could not telecommunicate with each other over the virtual networks constructed on their own PCs, we developed a function that interconnects such isolated networks by exploiting VPN technology. We also developed a support function to reduce the burdens of teachers and students for such complex interconnection operations.

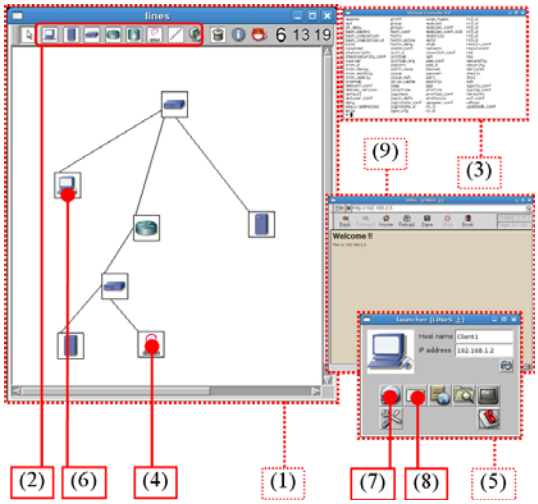


Fig. 3: Constructing a virtual network.

4.1 Function for interconnecting virtual networks

Fig. 4 shows our implementation method for interconnecting virtual networks of students. “Physical network” shows the structure of the actual networks. Virtual networks constructed by students on their own PCs are connected to “LiNeS-Net” by actual networks, such as LANs and the Internet. LiNeS-Net is a virtual network, like a pseudo Internet, and has a DNS root name server and a package server as network resources. “Implementation” shows an image of telecommunication by the implementation method. We adopted VPN technology to connect the LiNeS-Net and the student virtual networks. “Virtual VPN Gateway” is a furnished Open VPN [12] VPN software and three Ethernet interfaces: one is connected to networks constructed by students and a teacher, another is connected to real networks, and the other is used for forming VPN. By establishing a VPN connection between the Virtual Gateway of LiNeS-Net and students, networks of students can telecommunicate with each other by the LiNeS-Net’s Virtual Gateway. In such conditions, the students’ virtual networks and the LiNeS-Net form an independent network shown in the Logical network on actual networks. Such a network form, which is useful for preventing both adverse effects on actual networks and student misunderstandings, promotes effective and efficient learning.

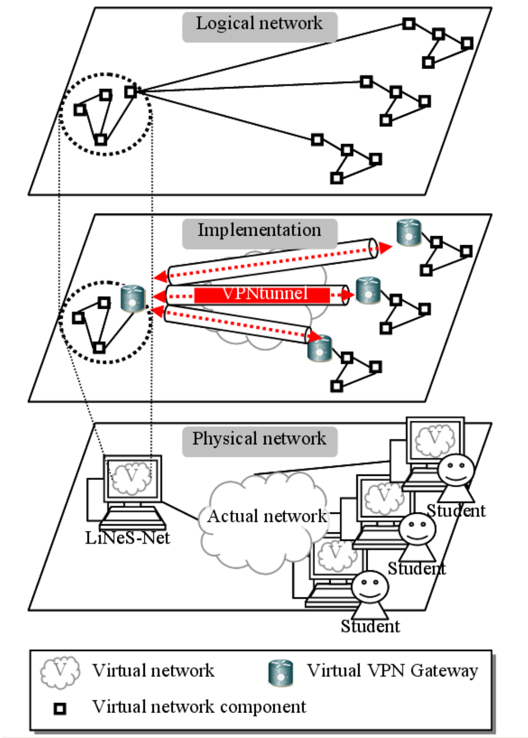


Fig. 4: Connection between virtual networks by VPN.

5 EXECUTION EXAMPLES

In this section, we describe the usage examples of the new function for network construction training. First, students connect their VPN Gateway to the VPN Gateway in the LiNeS-Net. Next, they get information of the assigned networks from the VPN server (Fig. 5) and start to construct and administer their own networks under the VPN Gateway. Fig. 6 shows the LiNeS-Net and networks constructed by Students A and B, whose virtual networks established the VPN connection to the LiNeS-Net by each VPN Gateway. Both networks can telecommunicate by the LiNeS-Net with each other. Student B finished uploading a test page in HTML to a web server in her own virtual network. Then Student A confirmed whether he can browse the test page on the server of Student B from a client in his network. If he can watch the test page (like this case), Students A and B have constructed their networks correctly. If not, either A and/or B constructed their networks mistakenly. So they must search for mistakes and modify them by communication

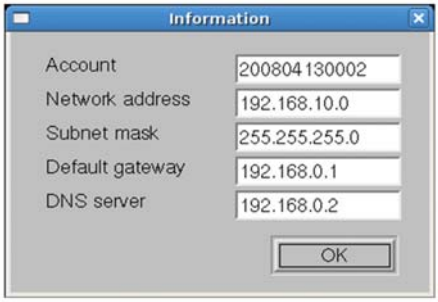


Fig. 5: Information about Student A, assigned to network.

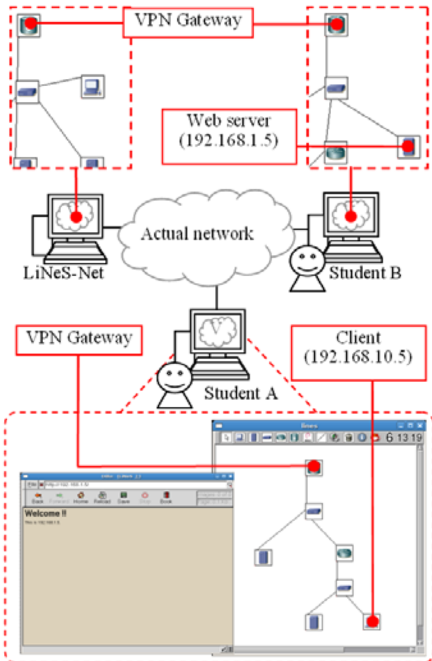


Fig. 6: Telecommunication between virtual networks by VPN technology.

6 PRELIMINARY EVALUATION EXPERIMENT

The VPN Gateway of LiNeS-Net can become a bottleneck in this system. All student telecommunications for the external networks are routed through the VPN Gateway. Therefore, by measuring the performances of both the VPN Gateway and the networks connected to it when a huge amount of telecommunication data is simultaneously routed through the VPN Gateway, we illustrate the worst performance of LiNeS. Then, based on this result, we discuss the practicability of LiNeS. The measurement environment is three patterns networks: one PC for the LiNeS-Net and either two, eight, or fourteen PCs for the students. We put one PC for students among each pattern on the Internet and the others on a LAN where the PC for the LiNeS-Net is located. The PC specifications for the LiNeS-Net include the CPU and memory of Pentium 4 2.8 GHz 512 MB, and the PC specifications for the students are equal to or worse. A measurement tool for the networks is ttcp [13]. After making a pair of PCs, the PC and its paired PC send and receive telecommunication data from each other simultaneously. In this situation, each student downloads data from the paired student. By conducting such telecommunication on all the pairs in each pattern, we measured the network performance when the VPN Gateway has the heaviest workload. Figs. 7 and 8 show the CPU use rates of the LiNeS-Net PC and the traffic speeds of the student PCs, respectively. During telecommunication in all of the patterns, the CPU use rates are 100% and the traffic speeds are in inverse proportion to the scale of the networks. Therefore, we verified that the bottleneck is the CPU performance of the PC for the LiNeS-Net, and from this, we concluded that the bottleneck was caused by the telecommunication data waiting for transfer processing on the LiNeS-Net PC. Comparing the patterns of the two units, the transfer speed on the eight-unit pattern is almost 33% worse and the fourteen-unit pattern is almost 90% worse; the traffic speeds are in inverse proportion to the number of PCs. From this result, using a condition where the LiNeS has 30 students, we extrapolated the traffic speeds on 30 PCs as about 13 KByte/sec. Up to 5 Mbytes of data are used during the LiNeS practice. Based on our calculation, it takes about 386 seconds to transfer the data. However such a situation where all 30 students simultaneously send and receive 5 Mbytes of data is rare. Therefore, this is not a big concern. On the other hand, since the transfer speeds are approximately the same both between the PCs on the LAN and on the Internet and the PCs on the LAN, we concluded that they have little trouble conducting exercises by the Internet by LiNeS.

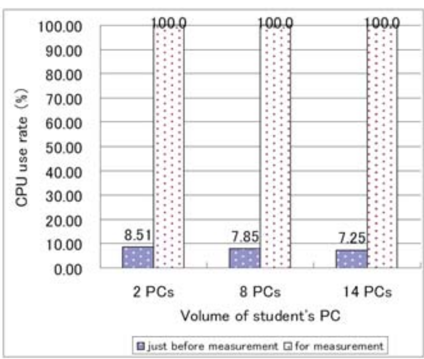


Fig. 7: Measurement of CPU use rate.

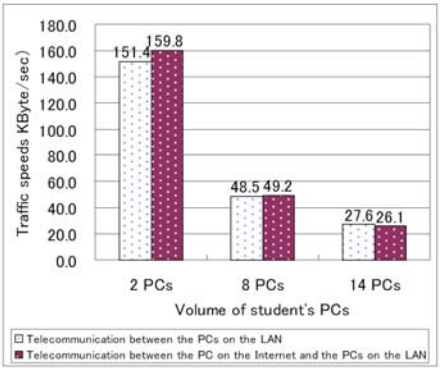


Fig. 8: Measurement of traffic speeds.

7 CONCLUSION

In this paper, we described how to connect virtual networks constructed by students on their own PCs to external networks. Then we discussed their practicability based on the results of our preliminary evaluation experience in which we measured traffic speeds between networks and CPU use rates of the PC running on the VPN Gateway. From the results, we extrapolated the worst performance of LiNeS during use. From the extrapolated results, we concluded that LiNeS can work without big concern. In the future, we would like to develop teaching materials and establish a curriculum for utilizing our system. Acknowledgments This research was partially funded by the Grants-in-Aid for Scientific Research Foundation and the Telecommunications Advancement Foundation.

References

1. The User-mode Linux Kernel Home Page, `http://user-mode-linux.sourceforge.net/index.html` (accessed March 5, 2009)
2. Tateiwa, Y., et al.: A System for Providing A Training Environment for Linux Networks using Virtual Environment Softwares. IJCSNS International Journal of Computer Science and Network Security 6(8A), 166–173 (2006)
3. Tateiwa, Y., et al.: Development of a System to Visualize Computer Network Behavior for Learning to Associate LAN Construction Skills with TCP/IP Theory, Based on Virtual Environment Software. IPSJ Journal 48(4), 1684–1694 (2007)
4. Tateiwa, Y., et al.: Development of A System For Providing a Training Environment for Network Troubleshooting Based on Virtual Environment Software. Advances in Computer Science and Engineering 1(3), 223–248 (2008)
5. Anisetti, M., et al.: Learning Computer Networking on Open Paravirtual Laboratories. IEEE Transactions on Education 50(4), 302–311 (2007)
6. Welcome to xen_org home of the Xen hypervisor the powerful open source industry standard for virtualization, `http://www.xen.org/` (accessed March 5, 2009)

7. Nakagawa, Y., et al.: Development of a LAN Configuration Support System for Study Using VMware. Transactions of Japanese Society for Information and Systems in Education 24(2), 126–136 (2007)
8. VMware Workstation Virtual Machine Multiple Operating Systems Windows and Linux - VMware, http://www.vmware.com/products/ws/ (accessed March 5, 2009)
9. Netkit, http://www.netkit.org/ (accessed March 5, 2009)
10. Main Page - VNUML-WIKI, http://www.dit.upm.es/vnumlwiki/index.php/Main_Page (accessed March 5, 2009)
11. Jiang, X., et al.: VIOLIN Virtual Internetworking on Overlay Infrastructure. In: Cao, J., Yang, L.T., Guo, M., Lau, F. (eds.) ISPA 2004. LNCS, vol. 3358, pp. 937–946. Springer, Heidelberg (2004)
12. Welcome to OpenVPN, http://www.openvpn.net/ (accessed March 5, 2009)
13. ttcp.c, http://www.netcordia.com/files/ttcp.c (accessed March 5, 2009)

Intelligent Educational System Based on Personal Agents

Marian Cristian Mihăescu, Dumitru Dan Burdescu, Costel Marian Ionascu, and Bogdan Logofatu

Abstract This paper presents an intelligent educational system that uses personal agents for providing personalization and adaptation of e-Learning content to learners. More specifically, we focus on a mechanism for guiding the learner through e-Learning material such so as to maximize the possibility to reach specific goals. The mechanism is designed and implemented using personal agents. The guidance is performed in real time and takes into consideration the actions that were performed by learner himself and also by other learners. An inference engine has been designed to decide if a resource (quiz, definition, chapter, etc.) will be proposed to the learner for study. The e-Learning system is organized as a multi-agent system composed of two kinds of agents: personal learner agent and recommender agent.

1 Introduction

This paper presents advances made within an e-Learning platform called Tesys [17]. This platform has initially been designed and implemented only with core functionalities that allowed involved people (learners, course managers, secretaries) to collaborate in good conditions.

Marian Cristian Mihăescu · Dumitru Dan Burdescu
University of Craiova,
Software Engineering Department, Romania

Costel Marian Ionascu
University of Craiova,
Analysis, Statistics and Mathematics Department, Romania

Bogdan Logofatu
University of Bucharest,
CREDIS Department, Romania

E. Damiani et al. (Eds.): New Direct. in Intel. Interactive Multimedia Sys., SCI 226, pp. 545–554.
springerlink.com

Making Tesys platform user-adaptive represents one of the main goals. Within the platform a mechanism for monitoring and storing all learner's executed actions was implemented. Having this data as input a CLM (Centralized Learner Model) is built and used by a PABS (Personal Agent Based System) in order to match the e-Learning material presentation with user's learning style, preferences, background knowledge and habits. The purpose of this system is to suit each learner's needs, with the purpose of maximizing the subjective learner satisfaction, the learning speed (efficiency) and the assessment results (effectiveness).

The proposed solution makes intensive use of PAs (Personal Agents). For each learner that interacts with the e-Learning system the corresponding PA is created. This agent is one that will interact with the CLM through the PABS system. The PA will represent the learner in front of the e-Learning platform and will try to take the best options that may lead to expected results. In order to obtain good results an important aspect is the employed software architecture of e-Learning platform itself and the followed software development process.

The paper presents in section 2 the main characteristics of Tesys e-Learning platform. In section 3 the employed software architecture is presented and in section 4 the employed software development process is discussed. Section 5 presents the PABS architecture and section 6 presents the conclusions.

2 Tesys Application Platform

We have developed an e-Learning platform that is a collaborative environment for learners, professors, secretaries and administrators. Secretary users manage sections, professors, disciplines and learners. The secretaries have also the task to set up the environment in which professors and learners will work. The main task of a professor is to manage the assigned disciplines. The professor sets up chapters for each assigned discipline by specifying the name and the course document, and manages test and exam questions for each chapter. The platform offers the students the possibility to download course materials, take tests and exams and communicate with other involved parties like professors and secretaries. All users (administrators, secretaries, professors, learners) must authenticate through username and password. If the username and password are valid the role is determined and the appropriate page is presented.

A message board is prepared for professors, secretaries and learners to ensure peer-to-peer communication. This facility is implemented within the platform such that no other service (e.g. email service) is needed.

From software architecture point of view, the platform is a mixture of data access code, business logic code, and presentation code. For development of such an application we enforced the Model-View-Controller [1] (MVC for short) design pattern for decoupling data access, business logic, and data presentation. This three-tier model makes the software development process a little more complicated but the advantages of having a web application that produces web pages in a dynamic manner is a worthy accomplishment. From the software development process point of view we enforced a software cycle development with project planning, requirements definition, software architecture definition, implementation, test, and maintenance and documentation stages. Software development makes intensive use of content management through a versioning system, testing and continuous building infrastructure.

3 Software Architecture in Tesys

The e-Learning platform consists of a framework on which a web application may be developed. On server side we choose only open source software that may run on almost all platforms. To achieve this goal Java related technologies were employed. The model is represented by DBMS (Data Base Management System) that in our case is represented by MySQL [2]. The controller, which represents the business logic of the platform, is Java-based, being built around Java Servlet Technology [3]. As Servlet container, Apache Tomcat 5.0 [4] is used. This architecture of the platform allows development of the e-Learning application using MVC architecture. The view tier is templatebased, WebMacro [5] technology being used. WebMacro is also a Java-based technology the link between view and controller being done at context level. The separation between business logic and view has great advantages against having them together in the same tier. This decoupling makes development process more productive and safer. One of the biggest advantages of not having business logic and view together is the modularity that avoids problems in application testing and error checking. In figure 1 are presented the main software components from the MVC point of view. MainServlet, Action, Manager, Bean, Helper and all Java classes represent the Controller. The Model is represented by the DBMS itself while the Webmacro templates represent the View. The model is built without any knowledge about views and controllers.

The business logic of the application uses Java classes. As it can be seen in figure 1, there are four levels of dependency between classes. The levels are: servlets, actions, managers and beans. Servlets level has so far two of them: Main-Servlet, DownloadServlet, UploadServlet. The platform is currently in use on Windows 2003 Server machine. This platform has three sections and four disciplines at each section. There are defined twelve professors and more than 650 students. At all disciplines almost 2500 questions are edited. In the first month of usage almost 500 tests were taken. In the near future, the expected number of students may be close to 1000. Recording student's activity provides great information regarding user traffic. After six month of usage there are more than 40,000 recorded actions. Our next concern is to employ different methods of analyzing the traffic in order to create usage patterns and keep the application in good shape.

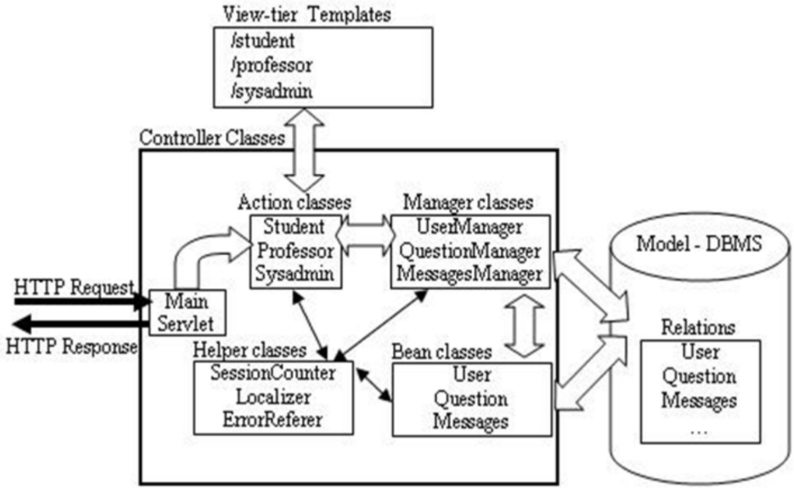


Fig. 1: Software architecture of the platform.

4 Software Development Process

Software development process and practices have the main goal of building quality software. Project planning represents the first and most general step that needs to be accomplished. This step consists effectively of four stages that need to be accomplished: scope definition, schedule and resource planning, cost estimation and project integration. In this paper we focus on the first and last stages. Scope definition also known as requirements consists of all facilities that the final product has to meet. Project integration, on the other hand, represents the process by which all requirements become real application facilities through design, implementation and testing. Under these circumstances at this stage a clear understanding of the scope of project is the most important goal. The major initial concerns relate to clarifying the requirements, the deliverables, and the organizational framework. Project planning may be seen as a series of stages that define software lifecycle [7].

Requirements definition and analysis phases have as final deliverable the functional specifications document for the system and a requirement analysis report. In this phase developers will resolve ambiguities, discrepancies and specifications that are to be determined. From requirements analysis a preliminary design may be derived that defines the software system architecture and specifies the major subsystems, input/output (I/O) interfaces, and processing modes.

At this step the system architecture defined during the previous phase is elaborated in detail. The development team fully describes user input, system output and I/O files. This step consists of a set of transformations that attempt to understand the exact needs of a software system and convert the statements that represent needs into a complete and unambiguous description of the requirements, documented according to a specified standard. This area includes information about the requirements, activities of elicitation, analysis, and specification. Requirement elicitation provides knowledge that supports the systematic development of a complete understanding of the problem domain. Requirements analysis provides knowledge about the modeling of software requirements in the information, functional, and behavioral concept of a problem. Requirements specification is concerned with the representation of software requirements that result from requirements elicitation and requirements analysis [8,9].

During the implementation (code, unit testing, and integration) phase, the development team codes the required modules using the detailed design document. The system grows as new modules are coded, tested, and integrated. The developers also revise and test reused modules and integrate them into the evolving system. Implementation is complete when all codes are integrated and when supporting documents (system test plan and draft user's guide) are written. Software coding and unit testing are concerned with proving that a correct solution to a problem has been developed. Unit testing is one of many testing activities such as performance testing, integration testing, system testing, and acceptance testing [10, 11]. System testing involves the functional testing of the system's capabilities according to the system test plan. Successful completion of the tests required by the system test plan marks the end of this phase.

In acceptance testing phase an acceptance test team that is independent of the software development team examines the completed system to determine if the original requirements have been met. Acceptance testing is complete when all tests specified in the acceptance test plan have been run successfully [12, 13]. Maintenance and operation begins when acceptance testing ends. At this stage, the system becomes the responsibility of the maintenance and operation group. This last step includes methods, processes, and techniques that support the ability of the software system to change and evolve [14]. Content management is a technology solution that is implemented using specific techniques which ensure wide-scale usability [15] for people involved in the project. Our discussion will focus on content management for web applications. The entire web application can be seen as a web property that is composted of various web assets. Managing content includes the steps to design, create, implement, modify, archive, review, approve and deploy.

As the size of web operation increases in a web development group, different techniques for managing the web property come into play. The approach has to take into consideration the web operations. For a small web site live editing is the right way. As the number of assets grows, and the number of developers increases, it is not practical to edit the production server directly. A staging server is used and runs a copy of the production web site. At the next level, development groups retain the staging server but give each developer an independent area where to commit their changes. This has the benefit that each developer is able to test changes independently. When the number of assets or the web team becomes big enough it is necessary to adopt a content management tool. This will manage the asset in a timely, accurate, collaborative, iterative and reproducible manner. A content management infrastructure consists of subsystems that fulfill the following functions:

content creation and editing, content repository and versioning, workflow and routing, deployment and operations management. The most important one is the repository and versioning subsystem since it provides storage, access, retrieval, indexing, versioning and configuration management of content. The measure of effectiveness in the repository subsystem is done by its ability to store assets reliably, with scalability and excellent performance. Extreme programming (XP) and Gamma/Beck xUnit testing framework [16] has made testing a daily conversation topic among developers. Unit testing is a verification for observable behavior of a programmatic unit. Each unit is tested in isolation by applying a set of one or more inputs to the unit and the outputs are observed in each case. In Java a unit is a class, an input is a method invocation and an output is represented by return parameters.
JUnit [6] is an open source Java testing framework used to write and run repeatable tests. The framework has an elegant design and is mature and easy to use. Software unit testing has to be an integral part of the software development process and all levels of software testing should be run as part of an automatic process.

5 Mult-agent Architecture

Intelligent agents have some or all of the following capabilities: cooperation, reactivity, and adaptability. Figure 2 shows a diagram of agent taxonomy. Cooperative agents communicate with other agents and act according to the results of that communication. Proactive agents initiate actions without user prompting. Adaptive agents, learning from past experience, change how they behave in given situations. Personal agents are proactive and serve individual users. Collaborative agents are proactive and cooperate with other agents.

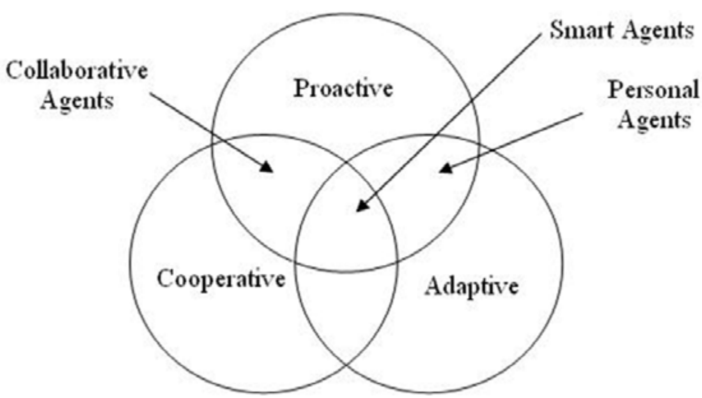


Fig. 2: Diagram of agent taxonomy.

Adaptive personal agents are an ideal technology for finding a user's personalized information. Researchers have developed personal software agents to help manage the increasing amount of available electronic information [18,19,20,21]. Because these agents can initiate tasks without explicit user prompting, they can undertake tasks in the background, such as information search. The user can access the search results by various means, but coupling them with Web technology allows easy access from wherever the user accesses the site. Of equally importance is the fact that some agents learn from experience. In the context of e-communities, learning more about individual members facilitates updating each member's profile; over time, this improves the accuracy of community data, including information about documents, people, and contacts. Personal agents both produce and consume information. By sharing their domain's knowledge with other agents, subject to the privacy limitations imposed on them, they contribute further to community knowledge. Collaborative filtering agents specialize in promoting interaction among community members. These agents benefit both senders and recipients because users can broadcast information to those who are interested in it without annoying other members. Contact-finding agents can locate members with distinct interests or competencies so that members can find experts in a given sub domain or other members with interests similar to their own. Agents can also work on behalf of individual members, shielding them from excess information or protecting expert members from excessive requests, thereby maintaining membership benefits for those who might otherwise become overwhelmed.

Researchers studied the ways in which e-Learning could be practically accomplished. Two technologies seem to have good chances of contributing to successful implementations of e-Learning systems [25]:

- Vortals (vertical or niche portals): They are specialized, dedicated portals that adapt specific learning collaborative strategies aimed at gaining performance and providing needed information. A learning portal should present different interfaces for different types of learners: one interface to a visual learner, one to an auditory learner, and one to a kinesthetic one. As the learner gains experience, frequently selected options should replace initial default choices. The more individualized the portal, the greater its impact;
- Intelligent agents: They are tools that can manage the information overload, serve as academic experts, and create programming environments for the learners. In this way, the learning process is enhanced by having many agents collaborating and competing towards achieving the prescribed goals. Intelligent agents should be able to model the user in order to remember her knowledge, her skills and her learning style.

E-Learning intelligent agents are represented by sets of independent software tools that are linked with other applications and database software running within a computer environment. The primary function of an e-Learning intelligent agent is to help a user interact with a computer application that presents a learning domain. Like intelligent agents in general, they should have four main characteristics [26]:

- autonomy: the agent operates by itself, without direct human intervention and holds control over its actions and over its internal state;
- reactivity: the agent perceives the environment in which it is situated, and answers in a timely manner to the changes in that environment;
- pro-activeness: the agent is capable to show a goal-oriented behavior by taking initiatives;
- social ability: the agent can interact with other agents or with humans by some communication language.

Intelligent agents have three main educational potentials [22]: they can manage information overload, they can serve as pedagogical experts, and they can even create programming environments for the learner. While the distance-learning concept provides more convenient virtual access to learners around the world, it also introduces some limitations and shortcomings [24], mainly from communication, collaboration, pedagogy, and course administration perspectives. One of the most important limitations comes from the lack of direct interaction between the teacher and the student. Using intelligent agents in a learning environment, some of these restrictions can be overcome.
Thus, an intelligent agent should be always available for course students, understand and interpret student problems (questions and requests), and have or know a set of actions which may activate according to its recognition of the student needs; and also its responses or decisions should depend on the course program, learner advancement or individual troubles and his special interests [23].
The mechanism of advising learners in real time has to go through several steps. The employed architecture is presented in Figure 3. Firstly, all actions executed by users are saved. These actions represent the raw input data for the mechanism that creates the CLM (Centralized Learner Model). As the time passes, the size of the actions database increases. This fact is very helpful in obtaining more accurate CLMs. Obtaining an accurate CLM with a high degree of predictability is one of the main goals of the employed architecture. The obtained CLM is used as an input for PABS module.

The main purpose of PABS is to create a PA (Personal Agent) for each user that interacts with the platform and a set of RA (Recommender Agent). The PA is instantiated with all the information regarding the user it is representing. PA will interact with appropriate RA in order to predict the behavior of the learner and infer the recommended resource list.

6 Conclusions

It was presented a software architecture that enables usage of Personal Agents and Recommender Agents within an e-Learning environment. The advantages of this approach regard the possibility of assisting user in real time during interaction with the e-Learning material. Besides the architecture itself, an important role is played by the steps that are to be followed in software development process. Following

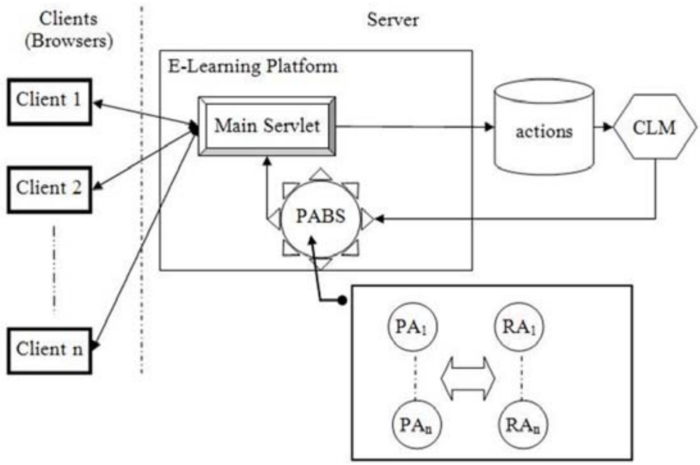


Fig. 3: Architecture of agent enhanced e-Learning platform.

these steps we have created our own development process specific to the platform that we have developed. The architecture of the platform is based on MVC model that ensures the independence between the model, the controller and the view.
Business logic is Java-based and the development process benefits from characteristics of object oriented programming [17] like modularity, reusability and extensibility. Standard technologies like XML allow describing objects through metadata and thus development of applications based on standard predefined objects. In this way interoperability between platforms may be achieved with great benefits in the process of data migration.
The proposed software architecture proved effective and efficient for being integrated within the platform. It is also very appropriate for implementing the complex actions as such type of platform requires. Taking into consideration the complexity of the software architecture which makes intensive usage of different tools and technologies the adopted software development process proved to be a reliable one. Developers that worked on this project had a good learning curve and integrated smoothly into the team.

References

1. Krasner, G.E., Pope, S.T.: A cookbook for using the model-view-controller user interface paradigm in smalltalk-80. In: JOOP, August/September (1988)
2. Yarger, R.J., Reese, G., King, T.: Managing & Using MySQL, 2nd edn. O'Reilly, Sebastopol (2002)
3. Hunter, J.: Java Servlet Programming, 2nd edn. O'Reilly, Sebastopol (2001)
4. Wiggers, C.: Professional Apache Tomcat. Wiley Publishing, Chichester (2003)
5. Faulk, S.: Software Requirements: A Tutorial, Software Engineering. IEEE Computer Society Press, Los Alamitos (1996)

6. Link, J.: Unit Testing in Java: How Tests Drive the Code. Morgan Kaufmann, San Francisco (2002)
7. Bashir, I., Goel, A.L.: Testing object-oriented software: lifecycle solutions. Springer, New York (1999)
8. Davis, A.: Software Requirements: Objects, Functions & States. Prentice-Hall, Englewood Cliffs (1993)
9. Faulk, S.: Software Requirements: A Tutorial, Software Engineering. IEEE Computer Society Press, Los Alamitos (1996)
10. Budgen, D.: Software Design. Addison-Wesley, Reading (1994)
11. Pigoski, T.M.: Practical Software Maintenance. John Wiley, New York (1997)
12. Beizer, B.: Software Testing Techniques. Van Norstrand Reinhold Company, New York (1990)
13. Pfleeger, S.: Software Engineering Theory and Practice. Prentice-Hall, Englewood Cliffs (1998)
14. Lano, K., Haughton, H.: Reverse Engineering and Software Maintenance. McGraw-Hill, New York (1994); Nakano, R.: Web Content Management - A Collaborative Approach. Addison-Wesley, Reading (2002)
15. Beck, K.: Extreme Programming Explained: Embrace Change. Addison-Wesley, Reading (2003)
16. Coad, P., Yourdon, E.: Object-oriented Design. Yourdon Press (Prentice Hall), New Jersey (1991)
17. Burdescu, D.D., Mihăescu, C.M.: Tesys: e-Learning Application Built on a Web Platform. In: Proceedings of International Joint Conference on e-Business and Telecommunications, Setubal, Portugal, pp. 315–318 (2006)
18. Maes, P.: Agents that Reduce Work and Information Overload. Comm. ACM 37(7), 31–40 (1994)
19. Lashkari, Y., Metral, M., Maes, P.: Collaborative Interface Agents. In: Proc. 12th Nat'l Conf. Artificial Intelligence, pp. 444–450. AAAI Press, Menlo Park (1994)
20. Nwana, H.: Software Agents: An Overview. The Knowledge Eng. Rev. 11(3), 205–244 (1996)
21. Crabtree, B., Soltysiak, S., Thint, M.: Adaptive Personal Agents. Personal Technologies J. 2(3), 141–151 (1998)
22. Baylor, A.: Intelligent agents as cognitive tools for education. Educational Technology XXXIX(2), 36–41 (1999)
23. Gadomski, A.M., Fontana, F.: Intelligent Agents in Distance Learning Systems, The UNINFO & eForm Meeting, Roma (May 2001), `http://erg4146.casaccia.enea.it/FADUNINF-5may01.pdf`
24. Jafari A.: Conceptualizing Intelligent Agents For Teaching and Learning, School of Engineering and Technology, Indiana University Purdue University Indianapolis, `http://www.cren.net/know/techtalk/events/jafar.pdf`
25. Hill, L.: Implementing a Practical e-Learning System (2003), `http://www.lakeheadu.ca/~publications/agora/december2002/elearning.html`
26. Wooldridge, M., Jennings, N.R.: Agent Theories, Architectures, and Languages: a Survey. In: Wooldridge, Jennings (eds.) Intelligent Agents. Springer, Berlin (1995)

Author Index