13TH ANNUAL EUROMEDIA CONFERENCE
2007

DELFT, THE NETHERLANDS

APRIL 25-27, 2007

Organized by

ETI

Sponsored by

EUROSIS

TTVI

EU-DG INFSO

BELGACOM

GHENT UNIVERSITY

HOSTED BY

DELFT UNIVERSITY OF TECHNOLOGY

EUROMEDIA'2007

FEATURING

THIRTEENTH ANNUAL SCIENTIFIC CONFERENCE
ON WEB TECHNOLOGY, NEW MEDIA
COMMUNICATIONS AND TELEMATICS THEORY
METHODS, TOOLS AND APPLICATIONS
AND D-TV

Leon Rothkrantz

and

Charles van der Mast

APRIL 25-27, 2007
DELFT, THE NETHERLANDS

A Publication of EUROSIS-ETI

Printed in Ghent, Belgium

EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

Editors

Dr. Leon Rothkrantz
Delft University of Technology
Delft
The Netherlands

Dr. Charles van der Mast
Delft University of Technology
Delft
The Netherlands

Workshop Editor

Dr. H. Joachim Nern
TTVI
Dusseldorf, Germany

Programme Committee

All articles published in this Proceedings have been peer reviewed

EUROSIS-ETI Publications are ISI-Thomson and INSPEC referenced

For permission to publish a complete paper write EUROSIS, c/o Philippe Geril, ETI Executive Director, Ghent University, Faculty of Engineering, Dept. of Industrial Management, Technologiepark 903, Campus Ardoyen, B-9052 Ghent-Zwijnaarde, Belgium.

EUROSIS is a Division of ETI Bvba, The European Technology Institute, Torhoutsesteenweg 162, Box 4, B-8400 Ostend, Belgium

Printed in Belgium by Reproduct NV, Ghent, Belgium
Cover Design by Grafisch Bedrijf Lammaing, Ostend, Belgium

EUROSIS-ETI Publication

ISBN: 9789077381328
EAN : 9789077381328

PREFACE

Over the last couple of years, new media has taken an increasing place in many activities in our every day life both as a professional and as a consumer. New media are becoming part of more and more aspects of our work by supporting computer-based tasks and activities, by supporting the user interfaces of embedded software, and by increasing the engagement and trust of users of web-based applications. This results in extending the research focus from traditional human-computer interaction to engineering effective experience with multi-modal interaction between human and artificial actors in a dynamic, social context. We think this is an important development. And in the programme of this EUROMEDIA 2007 conference you can recognize this development.

As with any conference, EUROMEDIA also would not be possible without the help and support of a number of people, and we would like to begin by thanking all the reviewers for their efforts, which have resulted in a truly interesting and varied conference programme. We are also most grateful to Pieter Jan Stappers of Delft University of Technology for presenting the keynote about the application new media to support design conceptualization, especially in industrial design engineering. Thanks also to the session chairs and other delegates who we are sure will guarantee us a lively and thought-provoking conference. Finally, special thanks are due to Philippe Geril, whose continued dedication and hard work as the conference organiser has enabled us to maintain the standard expected of EUROMEDIA events.

We sincerely hope that all of the delegates enjoy the conference, and that other readers of these proceedings will be encouraged to participate in EUROMEDIA events in the future. On behalf of all of EUROSIS, the International Programme Committee, we welcome you to this event and look forward to a successful conference.

Leon Rothkrantz
Charles van der Mast
General chairs

CONTENTS

KEYNOTE SPEECH

WEB BASED APPLICATIONS

WEB ENVIRONMENTS

AUDIO VISUAL APPLICATIONS

CONTENTS

MEDIA DATA COLLECTION

UBIQUITOUS COMPUTING

VEHICLE BASED MULTIMEDIA

WORKSHOP ON DIGITAL TELEVISION & DIGITAL SPECIAL INTEREST CHANNELS

SCIENTIFIC PROGRAMME

KEYNOTE

NEW MEDIA TOOLS TO SUPPORT DESIGN CONCEPTUALIZATION

Pieter Jan Stappers, Daniel Saakes, Aadjan van der Helm, and Gert Pasman
ID-StudioLab, Faculty of Industrial Design Engineering
Delft University of Technology
NL-2628CE, Delft
The Netherlands
E-mail: {p.j.stappers, d.p.saakes, a.j.c.vanderhelm,g.pasman}@tudelft.nl
Website: http://studiolab.io.tudelft.nl/

KEYWORDS
Design tools, Design research, new media

ABSTRACT

The early phases of a design project are characterized by combinations of associative and logical thinking. Interactive visualizations have always played an important role in this. In the design techniques research group, we study this phase, and try to support designers with new media tools. The research is driven by a merging of technology push (exploring media possibilities) and contextual push (studying design practice), and prototypes of new design tools take a central role in this. In this presentation we explain our approach and illustrate this with examples of interactive design tools to support early idea generation.

INTRODUCTION

The design techniques research group has been working for some 15 years on tools to support designers in the early phases of idea generation. In this work, the use of rich expressive computer-supported interactive visual media (which we'll refer to as 'new media' here) has been a central ingredient, together with contextual studies into the way designers work. The tools we develop have two purposes: as an instrument to clarify the current way of working, and as a demonstration of how the current ways of working might be improved upon.

The research is situated in ID-StudioLab in the faculty of Industrial Design Engineering of Delft University of Technology. ID-StudioLab is a multi-disciplinary community doing design research with a human-centered focus. We lay emphasis on making rich use of perceptual and motor skills, to support the creative cognitive processes, especially visual thinking, in idea generation and conceptualization.

In this paper we illustrate our approach by discussing some tools. For reasons of space, we limit ourselves to only the work from our lab, but we are not the only ones in this business, obviously. More balanced literature references are included in our regular papers.

IDEA GENERATION AND CONCEPTUALIZATION

The design process is generally conceived as a succession of phases, which may be iterated. In the first phase, the goal is defined, possibly in the form of a design brief, and initial analysis of the topic is performed, which may result in design requirements constraining the solution, and/or a design vision indicating a desired direction. This continues in generation of ideas for possible (partial) solutions, development of a concept of the product as a whole. The later phases include detailing the concept, working out production schemes and marketing the product.

In our work we focus on the early phases, especially the generation of ideas and concepts, which often includes bits of analysis as well. The design activities here can be very fuzzy as well as logical. Many different concerns are considered, loose ideas are generated, wild associations are evoked, and all these are integrated. In doing this, designers do a lot of visual thinking, and make expressive visualizations, such as sketches, models, and renderings. Design studios have a rich visual culture, as illustrated in Figure 1.

The introduction of computers as the ubiquitous tool for the thinking person in all professions has also had its impact in design. Drawingboards disappeared from design offices and designers, like everybody else, were sitting behind screens, operating keyboards, mice, and pen tablets (Figure 2). Computers were excellent at handing symbols and following logical rules, but lacked many of the informal strengths of traditional media as the sketch-on-the-wall. In our research we have worked at using those other strengths of computers, the media capabilities, to support designers at being creative.

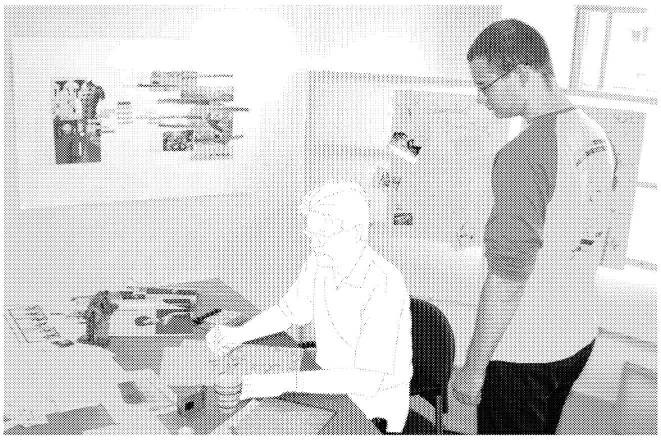


Figure 1: the Visually Rich Design Studio

Approach

The approach in developing the tools involves uniting two opposite and necessary forces: technology push and contextual push. The first is done by playing with the new media, such as the possibilities of using beamers, multiple input devices, different types of sensors, and different types of animations and interactions. These explorations yielded insights on how to use two-handed input techniques (Gribnau, 1999), how sketchy and hi-fi visualizations differ (Stappers & Hennessey, 1999, 2000; Stappers & Hoeben, 2001), and how large and small display sizes support different types of cognitive processes (Stappers, Keller, & Hoeben, 2001; Keller, Stappers, & Hoeben, 2001).

Figure 2: the Visually Poor Design Studio

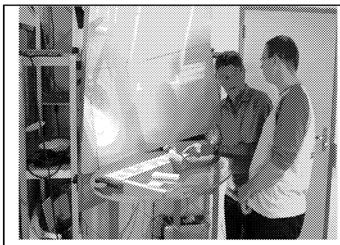
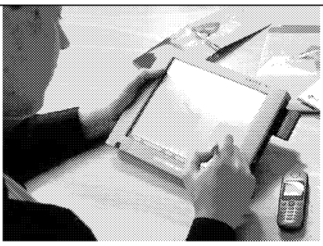
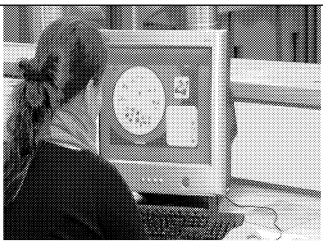
The second is done by studying the context of designers in current practice. In one such study, conducted in 1992 (see Pasman, 2003), it was found that designer's traditional tools were characterized by

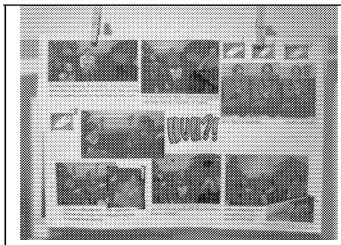
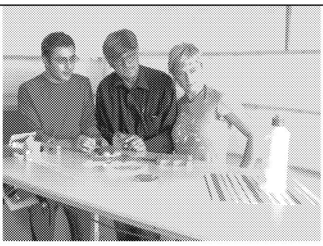
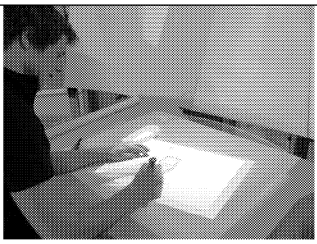
- An inspiring visual environment
- Use of high motor skills (e.g., sketching)
- Rapid shift between different ways of working (thinking, sketching, organizing,)

Computer tools were lacking in these respects. The rich visual environment, where designers and visitors were constantly reminded of and inspired by various aspects of their current and earlier projects, was replaced by closed electronic documents, stored by name in digital folders on a harddisc: good for retrieval, but never accidentally encountered. In a later study (Keller & Pasman, 2006), it was found that designers use two totally unconnected collections of visual materials: a physical one, used for inspiration, and a digital one, used for communication. Very little exchange (scanning or printing) actually occurred between the two.

Both pushes are merged by the development of a tool that fits the context and makes innovative use of the new technology parts. Such tools are developed in the studio, and we try them out on ourselves before we test them with designers in practice (Stappers, 2006). During the development of the tools, we reflect on the findings and decisions on the way, so the output of this research is not just the tool, but (more importantly) guidelines for supporting designers, and creative people in general. We refer to this approach as 'research through design', because in the activity of designing, we confront theories and empirical findings (van der Lugt & Stappers, 2006).

Table 1: Some Media-based Conceptual Design Tools from ID-StudioLab

 **TRI** A platform for exploring design tools using a sketchy variety of Virtual Reality techniques. (Keller, & Stappers, Hoeben, 2000)	 **SketchBook** A digital sketchbook which uses the fluency of real world sketchbooks. (Hoeben, 2001)	 **ProductWorld** An ideation tool that helps designers finding patterns in collections of existing designs by interactive spatial classification (Pasman, 2003)	 **Cabinet** An image collection tool that merges virtual and physical images in one seamless collection. (Keller, 2005)

 **Photoboarding** A technique to capture and retain playacting sessions in a rich and sketchy way, and develop them to storyboards. (Saakes and Keller, 2005)	 **Skin** A technique to play and explore colors, patterns and graphics on physical product shapes. (Saakes, 2006)	 **InstantTemplates** Digital templates of video to support physical drawing of natural two-handed product interactions. (Saakes and Keller, 2005)	 **Iris** A shared digital posting board for screenshots, to enhance situation awareness in distributed studios. (Peeters and Stappers, 2005)

Principles

Many of the design activities in early design are open-ended and involve associative and visual thinking next to logical thinking. But most computer tools supported only the latter. Therefore, the first guiding vision for the group was formulated as designing loose and sketchy tools, 'electronic beermats and napkins' (Stappers & Hennessey, 1999, 2000). This was refined later into principles based on aesthetics (sketchy, loose appearance to support associative thinking), interaction (making use of rich sensory and motor skills), and usability (directed at fitting real-world design processes rather than laboratory activities), which was illustrated by a sequence of 'tiny' tools (Stappers, Keller, Hoeben, 2002). More complete tools were developed into prototypes in PhD projects (e.g., Pasman, 2003; Keller, 2005; Saakes, 2006, 2006a). Table 1 shows an overview of some of the tools that came out of this research in the StudioLab.

AN OVERVIEW BY EXAMPLES

What did we learn from these exercises? Instead of reproducing the formal research paper findings (for that, see the references), we discuss the ingredients and findings through two examples of tools, and two platforms/toolkits

ProductWorld

Research had shown that designers studied existing products to find new solutions on product aesthetics. To support them in this highly visual process, a tool called ProductWorld was developed that allowed designers to categorize sample products from a database, organize them on criteria, and explore computer-generated multi-criteria organizations. It was found that through this process designers 'discovered' in a playful way visual properties of products regarding form, style, meaning etc. which are of great value for the generation of new form concepts, but also generally very difficult to formalize.

In ProductWorld designers can create structures of design knowledge by spatially arranging product samples relative to each other on various similarity criteria (Figure 3). The distances between the samples are taken as a measure for their mutual relationships. Thus samples which are arranged closely together are considered more similar than samples which are placed far apart. The resulting groups of samples can then be given names that typify their characteristics.

Retrieval of product samples is conducted through a dynamic, small, and interactive set of product samples. These are shown as an Multi-Dimensional Scaling configuration, reflecting their similarity relations. The points in this configuration are dynamically updated by moving them to their optimal location whenever the user removes samples, adds new samples, or changes the similarity criterion. Through these three actions, the designer implicitly builds up an understanding of the design knowledge that is embedded in the product samples and thus can be applied into new design situations.

ProductWorld's high degree of interaction encouraged students and designers to get actively engaged with the visual appearance of products (Figure 4). Such an active level of involvement was found to be essential, since hrough it the designer created, evaluated and modified new structures of design knowledge, which could then be applied in the generation and development of new product forms.

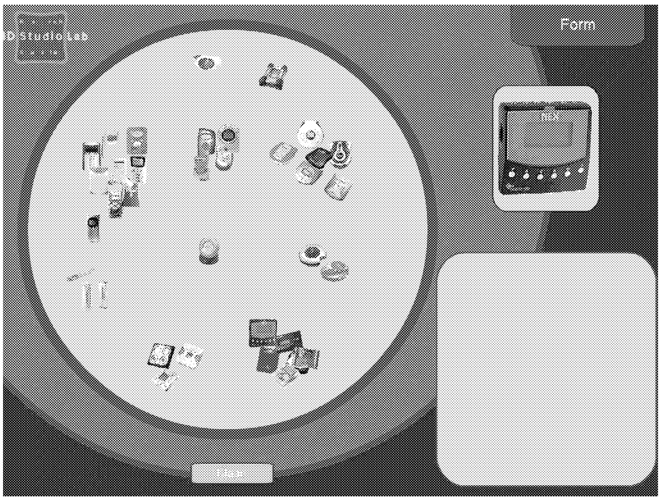


Figure 3. MP3 Players Organized by *Form* in ProductWorld

Figure 4. Using ProductWorld in an Educational Setting

Thus, while the visual thinking was supported by symbolic information 'under the hood', the interface allowed the designer to decide when to explore visually and when to study the logical values. By playing with this balance and shifting their way of thinking, a combination of analytical and creative thinking was created. An important lesson learned from the use of ProductWorld was therefore the notion that its value was not directly in any physical or tangible output, but much more in the experience and insights the designer acquired while interacting with the tool.

SKIN tool – tangible computer visualisation

In the early stages of new product development, consumer products are oftentime designed by multidisciplinairy teams, with stakeholders from engineering, marketing, usability and sometimes the end-users of the products are involved (Saakes, 2006). Creative group workshops support these stakeholders to generate new ideas and facilitate collaboration. These workshops regularly consist of cycles of generating new ideas and exploring ideas into concepts. Facilitators steer the workshops, and sometimes visualizers aid the participants by sketching ideas.

Even though these stakeholders actively participate and contribute to design solutions their input is mediated and primary in oral or written language. This in contrast to regular design meetings where designers make fluid and extensive use of sketching as well as building little prototypes out of paper or foam. Here, participants actively contribute and explore the solution space by doing first.

Skin 2.0 is a novel ideation technique aimed at these creative group workshops. The aim is to engage participants in unmediated intuitive exploring of the solution space by doing-first.

With Skin, groups of designers explore colors, textures and graphics on physical objects such as foam and paper models. Similar to other spatially augmented systems. Skin projects computer generated images with a projector. But, Skin projects materials, textures and colours as flat 2D images, without any tracking or knowledge of the 3D object. This loss of accuracy has two very large advantages: On the one hand, any object can be used, as long as it is white or light coloured. On the other hand, the rough and ambiguous projection that occurs when moving the objects in the projected light and seeing patterns and graphics deform gives rise to serendipity: unexpected new combinations.

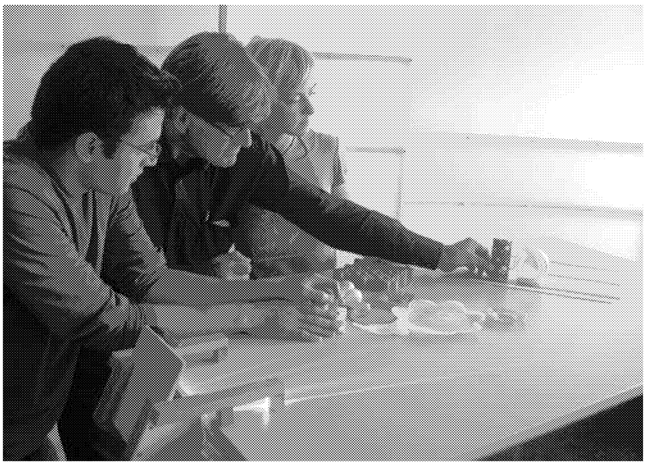


Figure 5. Skin 2.0 is a Physical/augmented Technique to Explore Colors, Patterns and Graphics on Physical Objects.

Skin's projector is mounted on a tabletop in such a way that only the objects on the table are augmented with graphics. See Figures 5, 6, and 7. The surplus light around the object is masked through backlighting. (Saakes, 2006).

Skin has two modes of creating graphics. The first mode is the "browse mode" . With a paddle controller participants can flick through a collection of inspirational images. Rotating the paddle scales and tiles the images.

During workshops in packaging industry we found need for a second mode; namely to add and compose physical artwork. With an attached video camera, physical materials, such as photos found in magazines or fabrics can be added and mixed on top of the digital images.In a series of workshops we found participants actively playing with graphics and generating many new designs. The workshops indicate that the not only the solution space is widened, more concepts are considered in the same amount of time, also the solution space is deepened, the physical/augmented approach provides a better view on a concept. Moreover, the technique might condense the cycles of generating and exploring and so make the process more efficient.

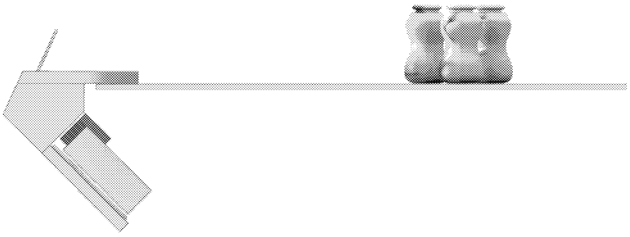


Figure 6 In Skin,Horizontal Projection and Backlighting Enhances the Physical Impression.

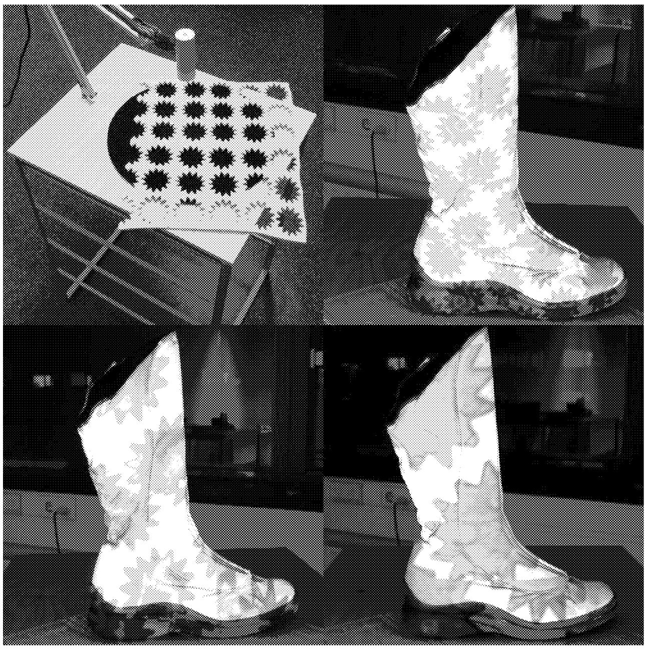


Figure 7 Designs Made by Participants. The Physical Pattern (top left) is Mixed with Digital Images and Projected on the Boot.

TRI – a platform for exploring interaction scales

The TRI setup, short for 'Three Ranges of Interaction', embodies what we learned about how physical scale affects cognitive activities. Its ranges small (fingers and hands, e.g.

pen tablet), medium (arm's length, e.g. table surface) and large (beyond arm's length) were programmed separately on separate hardware, in order to structure the tool designer's thoughts. See Figure 8 and 9. Hitherto, we had seen (and done ourselves) several instances of inappropriate spatial use in Virtual Reality tools (e.g., large CAVE setups used for detailed interactions), which were a mismatch in both technical sense (calibration problems) and a usability sense (using the wrong body actions and perceptual range).

Embodying these findings in an easy-to-program platform allowed design students to rapidly play with technologies, often realizing experience prototypes in less than a day, which allowed them to prototype user interactions in much shorter times than with conventional VR tools. Many small tools and toys were implemented on this platform. Most were short-lived, like a sketch on an envelope, but just as that sketch, helped to progress the development of the interaction concept.

TRI exemplifies that making a user-centered choice of hard- and software elements can lead to different and more appropriate tool designs. Also, TRI's large and medium scale displays were not rectangular on purpose, which forced our students to conceive interaction surfaces as not necessarily rectangular, inviting them to explore wider ranges of design solutions.

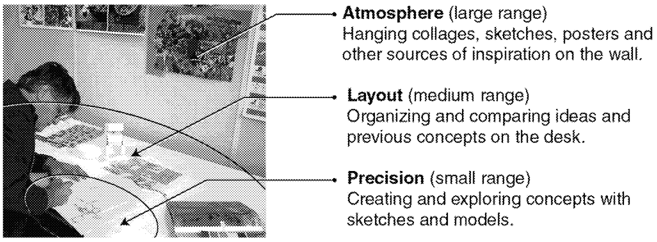

Atmosphere (large range)
Hanging collages, sketches, posters and other sources of inspiration on the wall.
Layout (medium range)
Organizing and comparing ideas and previous concepts on the desk.
Precision (small range)
Creating and exploring concepts with sketches and models.

Figure 8 Three Ranges of Interaction in Real Life

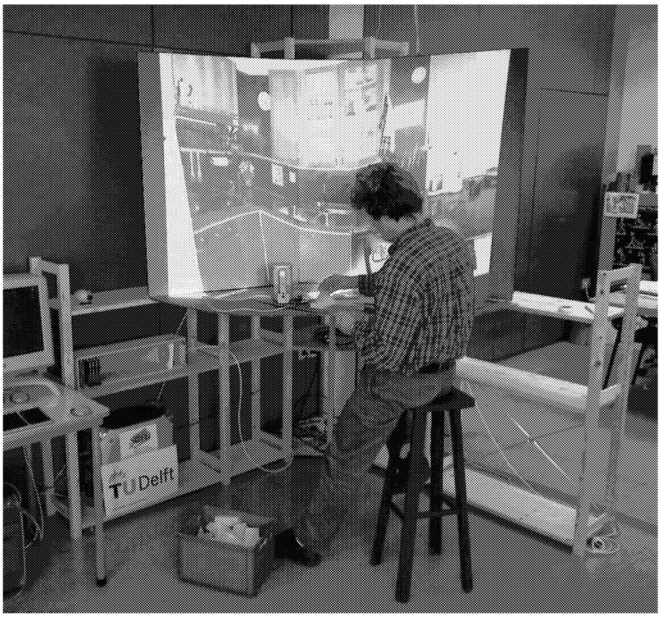


Figure 9 TRI uses interactive images on three scales.

Visual programming and modular sensor systems to support the design of interactive systems

In the field of designing interactive systems, multi disciplinary teams (designers, engineers, end-users, etc.) collaborate in finding a proper solution to the design problem at hand. A typical design process in this context involves getting a feel for what it means to use the interactive system under development. To achieve that, a highly iterative approach is taken that involves making prototypes and testing these. See Figures 10 and 11. The working prototypes provide a means to experience and communicate aspects of the concept for/to all members of the design team.

Because interactive systems have a large technological component traditionally the engineers dominated the early stages of product design, designers or end-users were only consulted in later stages of development. Recent advances in visual programming platforms (Pure Data, Max/MSP, LabView, d-tools) and modular sensor/actuator systems (ICube, Phidgets, Arduino) gave designers the opportunity to get involved early on in the process. See Figure 12.

Figure 10 Designer at work with the tools

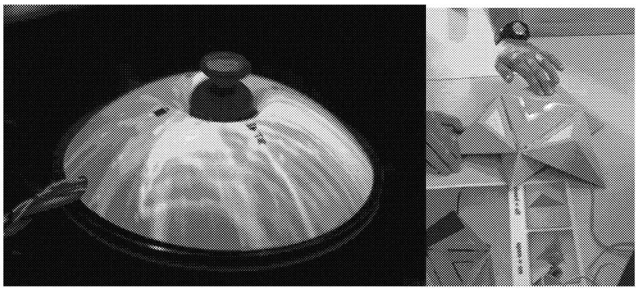


Figure 11 Two Examples of Interactive Music Player

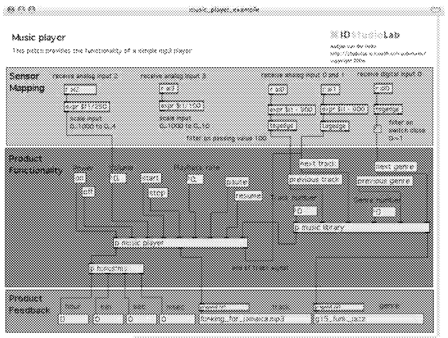


Figure 12 Example of Visual Program used for Designing Interactive Music Players

In recent years visual programming systems have become more widely adopted in the designers community because these systems are easier to use than the traditional language based software development tools. A visual programming environment typically doesn't employ an edit/compile/link/run cycle, instead most systems allow for modifying the program during runtime. This greatly enhances the possibility to explore different configurations of a prototype. Also the visual representation of the program communicates the operation of a program in a more direct way as opposed to language based program representations.
The combination of a visual programming environment with the modular sensor/actuator systems enable designers to be involved early on in the design process of interactive systems and thus enable them to contribute to making interactive products that are more user-friendly.

CONCLUSION

Approximately a decade after computers have prominently taken their place in all design studios, their use beyond symbol manipulators is growing up. The integration of advanced computer graphics hardware with the new operating systems, and the ability of tool designers to use these, opens up possibilities to support the associative and visual thinking styles that are essential for creativity in design. In our research we have explored ways in which these new opportunities can be given form, both in conventional GUI contexts, and in the newer developments in tangible interfacing.

REFERENCES

Gribnau, M. W. 1999 *Two-handed interaction in computer supported 3D conceptual modelling.* Ph.D. Thesis Delft University of Technology

Hoeben, A., & Stappers, P.J. 2001 "Ideas: A vision of a designer's sketchingtool". *Proceedings CHI2001: Conference on Human Factors in Computing Systems.* 199-200.

Hoeben, A, & Stappers, P.J. 2005 "Direct talkback in computer supported tools for the conceptual stage of design". *Knowledge-Based Systems*, 18 (8), 407-413

Keller, A.I. 2005 *For inspiration only:* Designer interaction with informal collections of visual material. Ph.D. Thesis Delft University of Technology

Keller, A.I., Stappers P.J., and Hoeben A. 2000 "TRI: inspiration support for a design studio environment." *DCNET conference.* (http://faculty.arch.usyd.edu.au/kcdc/conferences/DCNet00/.)

Keller, A.I., Hoeben, A., & Stappers, P.J. 2000. "Aesthetics, interaction, and usability in 'sketchy' design tools." *Exchange Online*, 1(1). (http://www.media.uwe.ac.uk/exchange_online/)

Keller, A.I., Pasman, G., and Stappers, P.J 2006 "Collections designers keep: Collecting visual material for inspiration and reference", *Codesign*, 2(1), 17-33.

Van der Lugt, R. and Stappers, P.J. 2006. *Design and the Growth of Knowledge.* Delft: StudioLab Press. (http://studiolab.io.tudelft.nl/symposium/).

Pasman, G. J. 2003 *Designing with precedents.* Ph.D. Thesis Delft University of Technology

Peeters, A., and Stappers, P.J. 2005. "Iris: Supporting workplace awareness by triggering informal interactions with visual material." *Proceedings DPPI: Designing Pleasurable Products and Interfaces*

Saakes, DP, & Keller, A.I. 2005. "Beam me down Scotty: to the virtual and back!" *Proceedings DPPI: Designing pleasurable products and interfaces*

Saakes, D.P. 2006 "Exploring materials: New media in design". *Drawing New Territories, 3rd Symposium of Design Research.* Zurich: Swiss Design Network. 109-123.

Stappers, P. J. 2006. "Creative connections: User, designer, context, and tools." *Personal and Ubiquitous Computing, 10 (2-3),* 95-100

Stappers, P.J. and Hennessey, J.M. 1999. "Towards electronic napkins and beermats: Computer support for visual ideation skills". In: Paton, R.C. & Neilson, E. *Visual Representations and Interpretations.* Springer , Berlin.. 220-225

Stappers, P.J. and Hennessey, J.M. 2000 "Beyond the electronic beermat: Digital devices for discussing design drawings." *Codesigning 2000 Adjunct Proceedings* 143-148

Stappers, P.J. & Hoeben, A. 2001. "Every pixel user-made: Aesthetic consistency in the development of computer-supported conceptual design tools." *Bulletin of the Fifth Asian Design Conference*, October 12-15, Seoul, Korea.

Stappers, P.J., Keller, A.I., Hoeben, A. 2000. "Listen to the noise: 'Sketchy' design tools for ideation." *Exchange Online*, 1(1), (http://www.media.uwe.ac.uk/exchange_online/)

AUTHOR BIOGRAPHY

PIETER JAN STAPPERS did an Msc in experimental physics at the Radboud University Nijmegen, and did his Ph.D. in Delft (1992), exploring the use of Virtual Reality in perception studies. Staying on in Delft, he joined the Design Techniques group, which he headed as full professor since 2003, trying to live up to the characterization by his superviser, that he was 'both playful and solid'.

DANIEL SAAKES was born in Amsterdam and did his M.Sc in industrial design engineering at Delft. He graduated at SARA with a Virtual Reality application in their CAVE setup. Then he worked as a freelance designer bridging the digital and physical through interactive toy concepts and augmented sports products. Now he has returned to Delft to teach 3D Visualisation to product designers and is working on his Ph.D. regarding new media tools. Daniel enjoyes the great outdoors and is an active gamer and hopes to complete guitar hero in expert mode.

AADJAN VAN DER HELM was born in Rotterdam and did a bachelor in Software engineering. He is involved in research and education at the ID-Studiolab, an institute of the Industrial Design Faculty at the Delft University of Technology. He is mostly active in the fields of early prototyping and tangible interaction. He has 20 years experience working with computer technology in a scientific context in the fields of computer graphics, interactive design and art.

GERT PASMAN received a M.Sc. in Mechanical Engineering from the University of Twente in 1989. After completing his military service, he joined the Faculty of Industrial Design Engineering at the Delft University of Technology, from which he obtained a Ph.D. in 2003. Currently he is mostly involved in teaching Interaction Design and Product Design. Gert has a deep fascination for polar exploration, although he himself prefers temperatures far above 20° Celsius.

WEB BASED APPLICATIONS

A WEB-BASED SOLUTION SUPPORTING THE INTEGRATION OF VIRTUAL REALITY ENVIRONMENTS IN LOGISTICS APPLICATIONS

Pietro Colombo
Dipartimento di Informatica e Comunicazione,
Università degli Studi dell'Insubria,
Via Mazzini 5, 21100 Varese - Italy
email: pietro.colombo@uninsubria.it

Emanuele Grosso
NEWLOG Consulting srl,
Piazza Carrobiolo 5, 20052 Monza(MI) - Italy
email: emanuele.grosso@newlog.it

Marco Tarini
Dipartimento di Informatica e Comunicazione,
Università degli Studi dell'Insubria,
Via Mazzini 5, 21100 Varese - Italy
email: marco.tarini@uninsubria.it

KEYWORDS
3DML, Virtual Reality engines, Distributed logistics applications

Abstract

Distributed logistics applications can benefit from Virtual Reality (VR) environments. In the context of warehouses management, virtual warehouses can show, in a visually intuitive way, the positions of stacked goods, the best ways how those positions can be reached, and so on. This work introduces an integration approach of virtual environments in web-based logistics applications. Our solution consists in a simple and sound modelling of 3D virtual warehouses, and in a platform independent navigator that we have specifically built around the structural and behavioural characteristics of the resulting 3D scenes.

INTRODUCTION

Industrial logistics applications can take advantage of the inclusion of a navigable virtual environment. An integrated system can provide additional services to a warehouse managament application, as suggesting the best path to the closest unit of a required stock, visually emphasizing the distribution of goods and their current status, and providing a visual representation of the results of queries on the stocked goods.
Additionally, during the planning activity for a warehouse, a virtual environment could be useful to carry out a feasibility assessment to organize the positions of the stocked goods; similarly, it can be used to test physical optimizations of an existing warehouse.
In order to achieve these goals, we need a 3D model showing the physical structure of the real warehouse and the current positions of the goods. We also need a real navigation tool that provides interaction with the scene, and that dynamically updates it to reflect the real world or planned changes (e.g. movements of goods). Furthermore, in order to support the usage of such tool in a distributed environment, updates are to be triggered also remotely. Finally, this application should be conveniently accessed from common web-browsers, because typical distributed warehouses management systems are interfaced through web applications to take advantage of the World Wide Web (WWW) infrastructure.

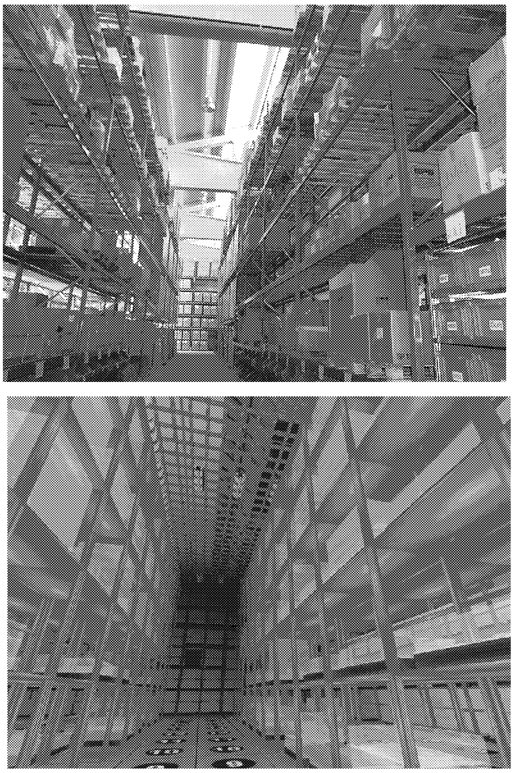


Figure 1: A real-world warehouse (top), and the corresponding 3DML virtual environment (bottom).

This paper describes a novel approach to facilitate a quick and inexpensive modeling of responsive virtual environments reflecting real (or planned) warehouses, and the design of an ad-hoc navigation tool embeddable in a distributed web-based environment.

MODELLING REAL WORLD SCENES

Purpose of this work is to fullfill the following requirements: (1) to model real warehouses and their behaviour with responsive 3D scenes in a sufficiently faithful way; (2) to interactively navigate and interact with them from a web application.

Several general purpose scene definition languages, as X3D (Web3D Consortium 2004) or it predecessor VRML (Web3D Consortium 1997), coupled with dedicated intepretion engines could be used to meet both needs. Unfortunately, 3D modeling based on such languages is usually a very demanding, expensive and time consuming task. In the general case, manual modeling via CAD systems is error prone and requires large amounts of time by expert designers. Even commercial modellers, as Graphisoft Archicad, that specifically target interiors, do not reduce the modeling effort enough. 3D range scanning techniques can be successully used to capture building interiors, as shown by projects as (Levoy 1999, Stumpfel et al. 2003). The results present a definition and faithfulness level that exceeds our needs, while the acquisition process is far too expensive for our scenario.

Image based modeling techniques (Debevec et al. 1998, Tarini et al. 2000) are more affordable and produce models with a more fitting geometrical complexity, relaying more on textures. Starting from a few pictures showing the interior, 2D features (lines, junctions, corners) are identified, then projective geometry principles are used to invert the projection, thanks to assumptions made on the regularity of the scene. Even this approach is too resource demanding here.

3D models need to be defined through a description at a far higher level, while we have a large tolerance in term of accuracy and faithfulness for the low level details.

The scene modelling task, which also includes the definition of behavioural aspect of the scene, must be as unexpensive and quick as possible.

For all these reasons we resort to a form of modeling that, taking maximal advantage of the simplicity and the modularity of our target (industrial buildings like warehouses), minimizes the efforts for their construction.

THE 3DML LANGUAGE

Our approach is based on the 3DML language (Flatland Online Inc 2003a), a XML (Bray et al. 2006) based language designed to allow an easy, direct description of 3D virtual environments. It was introduced by Flatland Online (Flatland Online Inc 2003a) and focuses on 3D content creation for web applications. 3DML originally targets applications as online games, but it has been used in other contexts, as for a Beer Festival setup (Polovina et al. 2000), for simplified versions of virtual cities (Doyle and Isbister 1999) and for a virtual museum (Virtual Open Air Museum Latvia, http://www.virmus.lv). We will argument that 3DML characteristics make it ideal for virtual warehouse modelling as well.

A 3DML virtual scene (a "spot") is composed by instances of 3D blocks that are disposed over a regular 3D grid. Each instance of a block appears as a 3D polygonized structure with associated textures, colors and other attributes.

The global structure of the spot is defined by its "map", a regular 3D grid of labels (defined slice by slice) that indexes a set of "3D blocks". Each "block" is defined extending an "abstract block" taken from one or more repositories (the "block-sets"), which are part of the 3DML document as well. An abstract block consists of a set of "sub-parts", each of which is composed by polygons and, optionally, has associated default appearance attributes (e.g. textures). Abstract blocks are equally sized and enclosed in a squared bounding box.

An extension of an abstract block can redefine some of the appearance attributes associated with its specific subparts (usually, texture or color). Textures images used in the block-sets are stored inside the block-set itself together with the description of the abstract blocks. Extensions can also add behavioral properties, scheduled by associated events (see next paragraph).

Scripting in 3DML

3DML allows the modeler to embed scripts for defining functions to access the elements of a scene. The supported language for such activity is a Simkin dialect.

Simkin (Whiteside 2000) is an *embeddable scripting language* whose scripts can be inserted in any XML document, thus including 3DML ones. A Simkin extension called Roverscript (Flatland Online Inc 2003b) is specifically designed for the 3DML language. Roverscript defines several elements and services to access the properties of the elements inserted in a spot. Among them, the *Player* element provides methods for changing the position of the avatar in the virtual scene, the *Map* element gives access to the instances of blocks in a given position, and the *Block* element provides services to change to the properties of an instance of block through transformations (translations, rotations, scaling) and overrides of appearance attributes (e.g. texture).

3DML allows one to define, inside a "spot", functions that can be invoked during the scene navigation: they can be triggered by navigation events, by timers, and so on. In the following sections we will extend this schema by allowing for remote functions invocation.

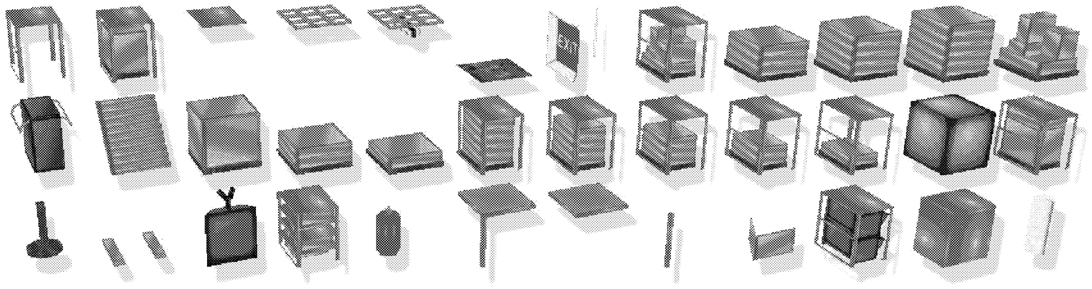


Figure 2: A small subset of the blocks that compose the "block-set" used for a warehouse scene.

APPLYING THE 3DML APPROACH

Given a specific application domain (e.g. warehouses, factories interiors, simplified cities, Doom levels...), building a new environment in 3DML format is fast and straightforward, since the same block-set(s) is shared by the great majority of the scenes in that domain. In particular, in the case of virtual warehouses, the modeling effort spent once to build the needed basic blocks is reused several times for different (but similarly themed) scenes. In most cases, a new virtual warehouse will differ from another for its 3D "map" only, which represents the high-level structure and can be easily built or modified either by a modeler or automatically.

A virtual warehouse is built following the structure of the real one by defining its map (typically sized tenths of blocks in each dimension), extending and instantiating blocks defined by dedicated blockset. Figure 2 shows a subset of the warehouse blockset (composed of over a hundred blocks).

Floor tiles are marked with textures that represent symbols to signal positions according to the standard warehouse shelf coordinate system. Shelves are modeled by using ad-hoc blocks reflecting the current stocking status of their real counterpart. Corridors are empty blocks between shelves. Stairs ramp blocks, leading to different floors, are endowed with scripts to ease their use (e.g. on contact the avatar is dragged to the intended final position, dispensing the user to drive it manually). The 3DML document also encodes an auxiliary GUI that provides direct accesses to predefined places in the scene, commands to change visualization modes and triggers for other context specific functionalities.

The resulting scene, modeled in a short time, is similar in its general structure to the real world warehouse (see fig. 1) enough for all our purposes.

This modeling approach is also web friendly, since the blocksets can be cached in the client side, dramatically reducing the amount of exchanged information during scene downloading in a distributed application.

It should be noted that the repetitiveness of the elements in a scene, that commonly represents a drawback, in the industrial logistics scenario reflects the real world structures (see fig. 3).

At present, available 3DML viewers provide game oriented interaction features, but fall short of communication capabilities that would made them suitable for the logistics scenario. Moreover, activities and animations (defined as Roverscript functions) can be triggered only by *local events* (e.g. generated by GUI, timers or navigation), while we are interested in dealing with *distributed events* that depend on the web-based environment where viewers operate. Specifically, we need to establish two way communication channels between web servers and 3DML viewers (web clients); such channels would allow the servers to issue events to the clients, and to monitor their status. As an example, the servers could dynamically invoke functions that adaptively update the 3D warehouse to reflect changes in the real one.

In addition, available viewers are platform dependent tools (Flatland Online Inc 2003a provides an ActiveX-based plug-in and a Win32-based stand alone application); conversely, in order to deal with the heterogeneous environment of distributed logistics applications, we are interested in platform independent solutions.

As a consequence, in order to exploit the advantages of the 3DML scene modeling approach previously discussed, we choose to develop a novel scene navigator that addresses the listed shortcomings. For such purpose, we focus on scene interactions and other needs associated to the context of logistics applications.

THE 3DML VIEWER

Open3DML is our web-based application specifically designed to support the visualization and navigation of 3D environments described with (a close adaptation of) the 3DML language. Open3DML is based on Java3D and Java Applet, established development technologies for 3D graphics and web-based applications.

Java3D achieves good rendering performances thanks to graphical hardware accelerations, and ensures high portability, hiding platform dependent implementation aspects, and finally provides an unified application interface for the underlying OpenGL or DirectX APIs.

As a result, our solution is platform independent, needs a Java plug-in and a Java3D package installed, and features real time GPU accelerated renderings.

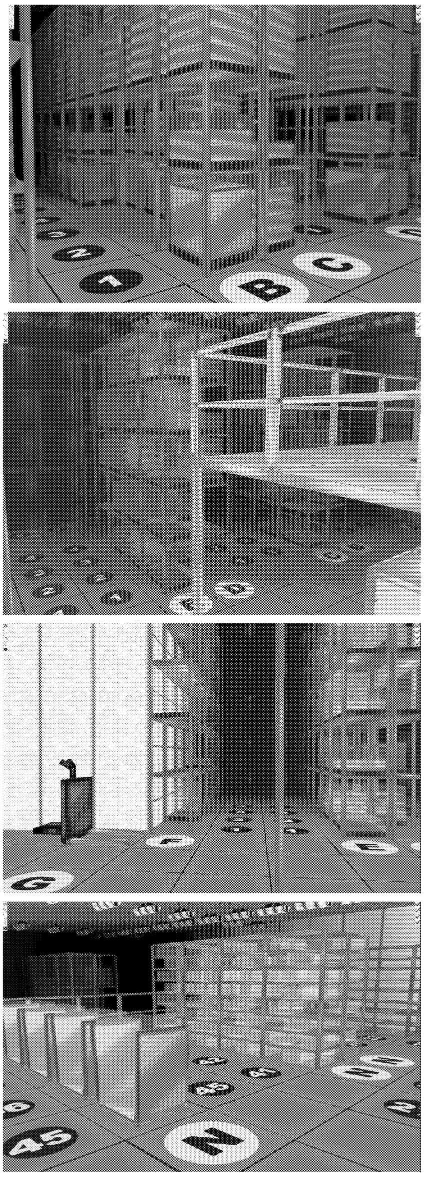


Figure 3: Snapshots of the navigator showing 3DML scenes that model a warehouse.

Taking advantage of the Java Applet technology, Open3DML can be used either as a stand alone tool, or within a web browser as a part of a distributed application.

Architecture

The Open3DML application is structured as a set of modules that supports different activities required by a virtual environment, such as scene loading, scene rendering and user interaction management.

Fig. 4 shows a sketch of the Open3DML architecture and its related modules. They include: (1) a core module providing services for loading a scene from a 3DML document and for rendering activities; (2) a Simkin module that provides inerpretation and execution capabilities for Roverscript functions; (3) a communication manager for the interactions with remote applications, and (4) a user interaction manager.

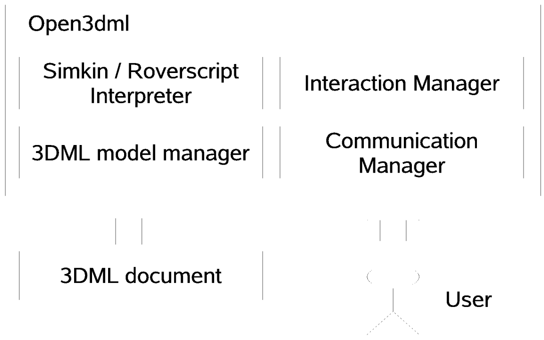


Figure 4: The high level architecture of Open3DML.

Scene loading

Open3DML is conceptually built around the 3DML meta-model. Since 3DML is a XML-based language, we defined its meta-model with an *XML Schema* Fallside and Walmsley 2004 describing the structural relationships among the elements of the language. This choice is motivated by the support for automatic validation, which can be carried out by a lot of parsers. Another reason is ease of development: the Java classes, constituting the core module of Open3DML, are automatically generated from the XML Schema through a template-based code generation approach. We used XML Spy by Altova (Altova Inc 2004) and the related SPL (Spy Programming Language) scripting language (Altova Inc 2004) in order to generate classes used to parse a document (compliant to the aforementioned XML Schema) and to translate the 3DML (XML serialized) elements in Java objects; the latter constitute the internal data structure used to build the 3D scene.

A Java class is dedicated to each element of the meta-model and each of these provides methods to access the properties of an instance of 3DML element.

A 3DML model can be seen as a four level structure (see Fig. 5). The first layer represents the original 3DML document. At the second level we find a DOM tree (Apparao et al. 1998) directly corresponding to the 3DML document. This tree is also the skeleton of the third level, composed by a hierarchical structure of Java objects that represents the elements of the model. The last level consists of the structure of Java3D objects that define the 3D scene. Such objects are preprocessed to optimize their subsequent rendering (see next paragraph).

Navigation and Rendering

Our Java3D based real-time rendering engine takes advantage of the high level structure of the 3DML scene being displayed. Each defined block type can be preprocessed and stored in a display-list or a vertex-buffer-object list, to be efficiently referenced multiple times in the same scene (through Java3D *Link-ShaderGroup* mechanism).

To improve visualization efficacy, our scene can include semi-transparent (alpha blended) elements: luckily it is easy to render the blocks composing the scene in a depth-sorted way thanks to the 3D lattice structure of block references.

Visibility culling and view frustum culling are also made straightforward, as the 3D lattice constitutes a natural decomposition of the scene in cells for precomputed visibility.

Collision detection is implemented in an easy and efficient way: at scene loading time each block is categorized as “impassable”, “floor tile”, “empty space”, and so on; this information is accessed at navigation time, avoiding any further computational effort. This is clearly an approximation, but it is well suited for our purposes. Collision detection is used both for standard collision responses, and as a possible trigger of scripted events (e.g.“step in” trigger).

The scene can also embed animated textures (described in the 3DML document as GIF images): all frames are preloaded in the graphic card memory to be later alternated.

Functions and triggers

Open3DML integrates a SimKin interpreter (Whiteside 2000), an open-source interpreter for RoverScript scripts that are embedded in a 3DML spot. This module is composed of:(1) a parser that analyzes the Roverscript code; (2) a series of Java classes that map the element defined for the base language; and (3) an engine that executes the parsed code and accesses the properties of the instances of the above described classes.

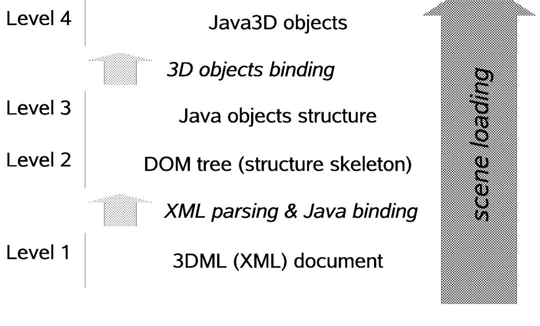


Figure 5: The Open3DML scene loading process

Client-Server communications

Open3DML can be used as a client interface of a distributed application that provides web-services. It is based on Java Applet technology and exploits the communication and transmission infrastructures provided by the WWW. The communication mechanism is based on the HTTP protocol.

Scene loading is the simplest client-server interaction form supported by the tool. 3DML scenes are remote resources that can be accessed through web servers. Open3DML sends a request to a specific URL. In response, the web-server returns a 3DML document that is parsed to instantiate a 3D scene.

Roverscript allows the modeler to define functions that can be invoked as responses to local events. These functions can be seen as services provided by a 3DML scene. As we discussed before, we also need double way communications between Open3DML and the remote application. To achieve this result in a HTTP compliant way, we extended the capabilities of the client side (Open3DML) by defining an interaction mechanism that allows a predefined remote application to require a service. Open3DML continuously sends, with a customizable rate, HTTP *GET* requests to a specific URL. The remote application replies sending either an empty document, or a document containing the invocation of any Roverscript functions defined in the currently loaded 3DML spot. A parser extracts the code from the sent document and passes it to the Simkin interpreter that processes the request. If necessary, after code execution, the result of the computation can be sent back to the remote application via a *POST* request.

As an example, let us consider the implementation of a potentially useful service: “return the shortest path to reach the position of a given pallet”.

Such service is implemented both on client and server side. The client side provides the management of the service invocation. The invocation can be triggered by the interaction of the user with a component of the 3D environment or the GUI. The server side hosts both the service processing activities, namely an implementation of the Dijkstra algorithm (Dijkstra 1959) to compute the shortest path, and a module that translates the output in a sequence of commands whose execution modifies the 3D model so that the shortest path appears as a series of arrows on the floor.

CONCLUSIONS AND FUTURE WORK

This work presented a simple approach to manually model virtual scenes representing industrial warehouses. Our solution minimizes the modeling effort and still produces virtual scenes that are acceptably similar to the real world structures, and additionally embeds meta-information and functionalities. The proposed approach is completed by a platform independent tool to navigate such scenes. Interaction with the scenes is achieved by both local and remote invocations of predefined methods embedded in the scene.

A promising direction, currently being investigated, involves porting Open3DML to self tracking, portable devices to be used inside the physical warehouse. This solution would clearly find many beneficial applications. The porting of Open3DML to hand held devices is eased by its software structure.

Moreover, we think that the general schema presented in this paper can be advantageously adopted in industrial context different from warehouse management. Incidentally, Open3DML, being a cross-platform tool capable to show general 3DML scenes, can be useful to the general 3DML community.

References

Altova Inc, 2004. *Altova XMLSpy 2005 User & Reference Manual.* Vervante.

Apparao V.; Byrne S.; Champion M.; Isaacs S.; Jacobs I.; Hors A.L.; Nicol G.; Robie J.; Sutor R.; Wilson C.; and Wood L., 1998. *Document Object Model level 1.* World Wide Web Consortium (W3C). W3C Recommandation.

Bray T.; Paoli J.; Sperberg-McQueen C.M.; Maler E.; and Yergeau F., 2006. *Extensible Markup Language (XML) 1.1.* World Wide Web Consortium (W3C). W3C Recommendation.

Debevec P.; Yu Y.; and Borshukov G., 1998. *Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping.* In *Rendering Techniques '98, Proc. of the EG Workshop.* Springer, 105–116.

Dijkstra E., 1959. *A note on two problem in connexion with graph. Numerische Mathematik*, 1, 269–271.

Doyle P. and Isbister K., 1999. *Touring machines: Guide agents for sharing stories about digital places.* In *Proc. of AAAI Fall Symp. on Narrative Intelligence.*

Fallside D.C. and Walmsley P., 2004. *XML Schema 1.1 Part 0: Primer.* World Wide Web Consortium (W3C). W3C Recommendation.

Flatland Online Inc, 2003a. *3DML Tag Reference.* Flatland Technical Report, http://www.flatland.com.

Flatland Online Inc, 2003b. *RoverScript: 3DML Scripting Reference.* Flatland Technical Report, http://www.flatland.com.

Levoy M., 1999. *The digital Michelangelo project.* In *Proc. of the Second International Conference on 3D Imaging and Modeling (3DIM99).*

Polovina S.; Khatri B.S.; and Singh S., 2000. *Culture and Web3D: Experiences in Building a Virtual Beer Festival Site in 3DML.* In *Proc. of British Computer Society HCI Cultural Issues in HCI Workshop.*

Stumpfel J.; Tchou C.; Yun N.; Martinez P.; Hawkins T.; Jones A.; Emerson B.; and Debevec P., 2003. *Modelling and Display of Architectural Forms Digital Reunification of the Parthenon and its Sculptures.* In *Proc. of the 4th Int. Symp. on Virt. Reality, Archeology and Intelligent Cultural Heritage (VAST-03).* EG Association, 41–50.

Tarini M.; Cignoni P.; Rocchini C.; and Scopigno R., 2000. *Computer Assisted Reconstruction of Buildings from Photographic Data.* In *Vison Modeling and Visualization 2000 Proceedings.* IOS Press.

Web3D Consortium, 1997. *Virtual Reality Modeling Language (VRML).* ISO/IEC 14772-1.

Web3D Consortium, 2004. *Extensible 3D (X3D).* ISO/IEC 19775.

Whiteside S., 2000. *Simkin for Java.* Technical Report, http://www.simkin.co.uk.

A WEB BASED SOLUTION TO MANAGE DISTRIBUTED DISCRETE EVENT SIMULATIONS

Alberto Coen-Porisini
Pietro Colombo
Ignazio Gallo
Antonella Zanzi

Dipartimento di Informatica e Comunicazione
Università degli Studi dell'Insubria
Via Mazzini 5, 21100 Varese, Italy
E-mail: {alberto.coenporisini|pietro.colombo|ignazio.gallo|antonella.zanzi}@uninsubria.it

KEYWORDS
Web-based distributed simulation, discrete event simulation, open-source SW, simulators integration, 3D visualization.

ABSTRACT

SINPL (Simulator Integration Platform) is an open-source software platform supporting the integration of existing simulators in a distributed Web-based environment and the management of simulation experiments.

In the present work we introduce the platform architecture and we discuss the role of the various modules composing the platform focusing on their visualization capabilities. Such capabilities aim at controlling and analyzing simulation experiments by means of the implemented 2D and 3D graphical user interfaces.

INTRODUCTION

In the simulation field the demand for distributed architectures is mainly motivated by the advantages of reusing existing simulators and modelling complex systems that could be difficult to realize with a single stand-alone application. Web-based simulation systems may be considered the natural evolution of distributed simulations and in the last years several proposals for this kind of system have been made (Kuljis and Paul 2000).

SINPL is an open-source software platform that allows one to carry out distributed simulations. The platform supports the integration of existing heterogeneous discrete event simulators in a Web-based simulation environment.

In a simulation environment, visualization capabilities are important both in modeling and execution activities, and they are used to achieve different goals such as animating modeled processes or allowing one to manage and interact with a running simulation.

In SINPL, the visualization is mainly devoted to show the communication flow among simulators during a running session. As a simulation updates the state of the modeled system, visualization can provide an abstract representation of the state of the on going simulation.

In the present work we introduce the SINPL platform architecture focusing on the simulation execution module and the implemented 2D and 3D visualization functionalities.

The paper is organized as follows: first of all the simulation design process supported by the SINPL platform is introduced; then the architecture of the platform itself is presented, followed by a description of the tools in charge of the simulation execution and control, and by the presentation of the implemented graphical user interface; finally, after a short analysis of the related works, some conclusions are drawn.

SIMULATION DESIGN PROCESS

The simulation design process (Coen-Porisini et al. 2004) supported by SINPL comprises three main activities described in the following.

1. The *Information Model* definition consists in defining the basic elements that represent either the logical components of a system related to a specific application domain. Thus, each domain has its own Information Model. As a result, the SINPL platform can be used in many different application domains by defining the appropriate Information Models.

2. The *Simulation Architecture* design activity aims at building a system by instantiating the elements of the Information Model. The Simulation Architecture provides a logical view of how the different simulation models cooperate by defining both data flow (i.e., which data are exchanged) and control flow (i.e., how the simulators interact).
The semantics of the Simulation Architecture is given in term of a High Level Petri Net (HLPN) (Jensen et al. 1997) in which data are associated with tokens and actions with transitions. A HLPN is associated with every component defined in the Simulation Architecture, and thus the Simulation Architecture itself results in a HLPN obtained by composing the different HLPN associated with the components therein.

3. The *Simulation Experiment* configuration and execution consists in defining how the simulation has to be carried out. This requires first to define the input data needed by the different simulators and then to actually execute the simulation. SINPL manages a simulation experiment executing the HLPN associated with its Simulation Architecture. Whenever a transition is enabled, the associated action is enabled as well, and the input data are taken from the input places and sent to the real simulator;

similarly the output data produced by the simulator are used to update the corresponding marking of the HLPN.

All the aforementioned phases are supported by means of software tools. In particular, Infocreator is the tool supporting the definition of Information Models, while SED is aimed at the definition of Simulation Architectures. Finally, DSC represents the core of the platform and takes care of managing experiments and running simulations. The next section discusses the platform architecture and the different tools composing the platform itself.

PLATFORM ARCHITECTURE

SINPL is a software platform characterized by a heterogeneous architecture. The whole platform is built around the *Distributed Simulation Controller* (DSC) system, which manages the execution and the visualization of a simulation experiment, coordinating the communication among the different simulators involved. DSC exploits the infrastructure of the World Wide Web (WWW) both to provide communication functionalities and to integrate its modules. This distributed application is characterized by a client/server architecture. The client side is composed by two modules, respectively *DSC Manager Web Client* and *DSC Client*, both interacting with a server side module called *DSC Manager* exploiting a HTTP based communication. Figure 1 shows the SINPL architecture from the DSC viewpoint. More specifically, it identifies which modules are composing the client side, which ones the server side, and the communication protocols used.

The server side

The server side of the SINPL architecture is itself a distributed system that includes DSC Manager and all the simulators. Such subsystem is characterized by a point-to-point star architecture, where all the communications among the different simulators modules are mediated by the DSC Manager. The system infrastructure has a hybrid nature: it can be based on the WWW and/or on other kind of infrastructures. DSC Manager, in order to communicate with the simulators, uses established technologies such as *Common Object Request Broker Architecture* (CORBA) (http://www.corba.org) and *Simple Object Access Protocol* (SOAP) (http://www.w3.org/TR/soap). CORBA is an OMG standard that provides integration functionalities for components of a distributed heterogeneous system, assuring independence from operating systems, programming languages and net infrastructures. SOAP is a W3C standard designed to support communication for Web applications. It is based on XML and supports different transport protocols. A description of the utilization of SOAP as communication infrastructure between DSC Manager and the simulators can be found in (Coen et al. 2006).

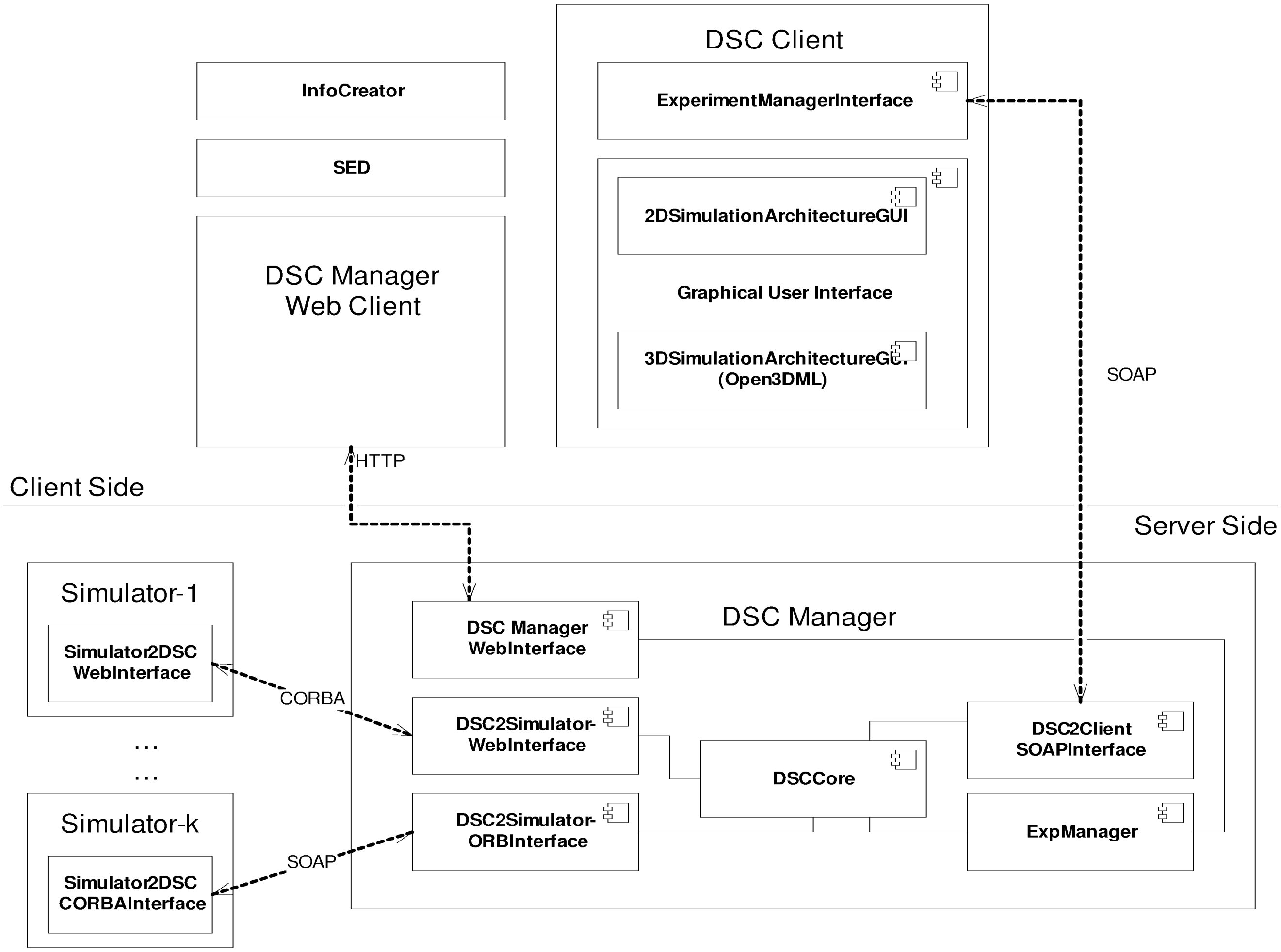


Figure 1: SINPL – Deployment diagram

Both SOAP and CORBA provide mechanisms that allow simulators to share data. A dedicated common software interface has been defined for all the components of a Simulation Architecture in order to enable direct communications among simulators and between simulators and DSC. Moreover, an XML-based data inter-change format has been defined in order to support data exchange. As a consequence each simulator may need a software adapter that implements such interface in order to convert data from/to the common format, and to allow one to supervise and control the simulator execution.

The client side

As aforementioned, the client side is composed of two modules called DSC Manager Web Client and DSC Client, respectively.

DSC Manager Web Client is a SINPL component that provides configuration capabilities allowing one to define and setup simulation experiments. Using such application, a modeller is able to prepare all the required information in order to start the execution of a simulation experiment. These data are associated with the components predefined in a Simulation Architecture.

Moreover *Infocreator* and *System Editor* (SED), two Java-based stand-alone tools, are deployed by means of DSC Manager Web Client exploiting Java Web Start technology (Sun Microsystems 2005). This technology allows one to download, install and execute programs stored on a remote location that can be accessed through a common Web server.

The Infocreator tool supports the definition of Information Models, while SED is aimed at the definition of Simulation Architectures. More specifically, SED provides a graphical editor that allows one to compose Simulation Architectures using the Information Models previously defined. A Simulation Architecture is built instantiating the simulation elements defined by an Information Model and defining how they have to be connected among them through the communication interfaces. SED provides a module, called Petri Net Editor, which allows one to generate and manage the Petri Nets associated with each component of the Simulation Architecture.

The second component (DSC Client) operating on SINPL client side, is a tool designed to manage the execution of simulations and to analyse the results of simulation experiments. In order to provide analysis capabilities, DSC Client integrates different kinds of graphical user tools.

In the rest of the paper, we discuss both the structural characteristics of DSC and the functionalities provided by its components, focusing on the communication, visualization and control capabilities.

THE DSC SYSTEM

A Simulation Architecture is a conceptual model that abstracts away from the characteristics of a heterogeneous distributed software system composed by simulators. A simulator is a software component that operates in the context of a Simulation Architecture receiving and producing data and events and performing some particular tasks.

In the SINPL platform, simulators communication is based on the Blackboard architectural pattern (Buschmann et al. 1996). The Blackboard architecture provides a way for communicating to a collection of independent programs that operate on a common data structure. There is not a predetermined sequence for the activation of the independent programs. Instead, the evolution of the system is determined by the state of the components and is controlled by a central control unit that coordinates the programs.

In the SINPL platform, the execution of the different simulators is managed by the DSC system. DSC is composed of tools that provide events logging, communication and control capabilities. It operates by sending control commands to the simulators and by storing the events and the data they produce. More specifically, executing the HLPN associated with the Simulation Architecture provides control capabilities allowing the synchronization of the different simulators composing a Simulation Architecture. Transitions firing in the Petri Net associated with a simulator represent an internal change of the computational state of the simulator. Instead, transitions firing in the Petri Net connecting different simulators possibly enables the execution of other simulators. The synchronization mechanism supported by DSC exploits an implementation of the Chandy-Misra-Bryant (CMB) protocol (Chandy and Misra 1979) and it is fully discussed in (Carullo et al. 2006).

The SINPL communication system is based on a message passing mechanism that ensures the persistence of the exchanged messages. In this context, messages are events generated by the different components of the Simulation Architecture. The most significant types of events are the following ones:

1. generic control events, as *SimulationStart*, and *SimulationStop*, which are generated at the beginning and at the end of a simulation experiment;
2. data events, as *DataSend* and *DataReceive*, respectively generated whenever a simulator sends or receives data;
3. transition events, as *TransitionRequest*, that is generated by a simulator that would like to change its state. Once received such message, DSC, depending on the current marking of the HLPN, enables the request replying to the simulator with a *TransitionAck*.

All the events are labelled with a Timestamp that specifies the time instant when they have been generated (Carullo et al. 2006) and are stored by DSC. In this way one can examine all the messages exchanged by the simulators and replay the simulation itself starting from any time instant.

The events stored by DSC describe the evolution of the simulators that is, all the data produced and received during an experiment. In order to perform analysis tasks, the events have to be elaborated and displayed. Similarly, we need to show the real-time evolution of all the components involved in an experiment.

Notice that the same experiment can be the target of different analysis tasks. Therefore, events can be displayed exploiting different viewpoints or different visualization modes. In a distributed system, monitoring capabilities should not be implemented by a centralized application. Such requirements and the structural properties of the SINPL platform, whose infrastructure is based on a heterogeneous system, motivate the choice of a distributed solution. Remote applications directly interfaced with DSC can implement all these requirements providing the possibility to simultaneously exploit several viewpoints and to analyse different parts of the whole simulation history without affecting the evolution of the experiment. Similarly, remote management applications can be useful to configure and manage simulation experiments.

Communication, management and monitoring functionalities have been split among different components to optimize the execution of the different tasks related to the management and analysis of simulation experiments. DSC is a Web-based system, expressively designed to satisfy the aforementioned requirements.

DSC Manager is a Web application that operates on the server side of the system. It can be imagined as a shell that provides a Web interface to the DSC system. DSC Manager provides all the functionalities through Web services. Thus, taking advantage of WSDL interfaces and SOAP messages, a remote application is able to directly invoke services provided by DSC Manager. In a typical scenario, once configured a simulation experiment, the modeller may use a Web browser to manage the execution on DSC using the DSC Manager. Such remote Web application can be accessed through DSC Client, a dedicated client application.

DSC CLIENT

DSC Client is a tool used to remotely manage and analyse the execution of simulation experiments. It operates on the client side of the system, accessing the services provided by DSC Manager, which operates on the server side.

This solution allows one to constantly get information on the current state of the on-going simulation experiment, and to manage the execution of an experiment itself.

DSC Client and DSC Manager exploit a profile-based authentication mechanism. Generic users can access monitoring services without any restriction, but only primary users can affect the execution of an experiment. More specifically, only administrators can start or stop the execution of an experiment.

Monitoring services can be accessed both in real-time and after the execution of an experiment. The latter mechanism exploits “log files” that store the events generated by simulators. The real-time monitoring services provided by DSC are based on a synchronization mechanism that uses Timestamp labels associated with the events. Given a certain time instant, the state of an on-going experiment is described by the set of all the events generated since the simulation started. DSC Client exploits an asynchronous interaction mechanism to obtain the current state of the experiment. More specifically, it implements a polling algorithm that constantly sends requests to the server in order to get updates on the experiment state. Such requests specify the Timestamp information associated with the last event received by the client. DSC replies to the client sending all the events that happened since the time instant specified by the received Timestamp. This mechanism allows us to optimise the quantity of data exchanged since all the events are transmitted only once.

DSC Client provides a user-friendly interface to access the DSC services and to analyse the events of a simulation experiment. The DSC Client graphical user interface (GUI) has been implemented using the Java SWING technology, an established standard to build platform independent graphical user interfaces.

The DSC Client GUI has been designed focusing on the optimisation of the visualization capabilities for interacting with the user. Notice that one of the most important features provided by the application is the animated visualization of simulations. Such functionality allows a user to graphically rebuild the execution of an experiment. The GUI structure is quite similar to that of common video-player applications. It provides a slide bar that operates the time line of a simulation execution. The slide bar is directly associated with the queue of events either obtained by DSC or loaded from a log file. As a consequence, users can either monitor an on-going simulation or review an already executed simulation. An example of the DSC Client GUI is shown in figure 2. The screen shot shown refers to the execution of an experiment in the Flexible Manufacturing System (FMS) context. A FMS is composed of several machines connected by means of a transport system. The transport system carries the raw parts to the machines where they are processed. Once the machines have finished their job, the parts are moved back to the load station where they are unloaded. Moreover, the machines use a tool-room as a repository for the tools they need in order to properly work the raw parts (Matta et al. 2004).

A detailed list of all the exchanged events is shown in a dedicated internal frame labeled “Timestamps List”. Such list is temporally ordered according to “timestamps” values. Each item describes the number of events that occurred at a given timestamp.

An internal frame labeled “Events at Timestamp”, lists all the events generated at a given instant. Selecting an item from the list, a user can obtain detailed information on the associated event. The information depends on the type of event, however some fields are common to all types. As an example an analyst can access the event ID or the name of the simulator that generated it. For events related to the distribution of data, a dedicated window shows the XML serialization of the data value. This window is useful for analysis tasks since it allows an analyst to completely rebuild the flow of data among simulators.

All the components involved in an experiment are summarized in a frame labeled “Architecture tree”. Such frame provides a tree view of the components belonging to the Simulation Architecture and of their communication channels. More specifically, one can find the names and the identifiers of all the involved simulators and of the links among their input and output ports.

Visualization in DSC Client

DSC Client provides a dedicated frame showing both 2D and 3D representation of simulation architectures and an animated representation of simulation experiments. In the 2D representation, simulators are shown as boxes labeled with a string identifier and colored dependently on their state. During a simulation experiment, simulators change their computational state (by producing/receiving data) or ask for enabling an internal transition. As a consequence, DSC notifies the user by changing the color associated with the involved simulators. The same mechanism (i.e., by changing the color associated with a connector) is used to notify users that some kind of data have been generated by a simulator. Colors show the computational state of the components and of the communication channels. This visualization might not be expressive enough to carry out analysis task on the experiment, but can help one to obtain an immediate view on the simulation state at a given instant.

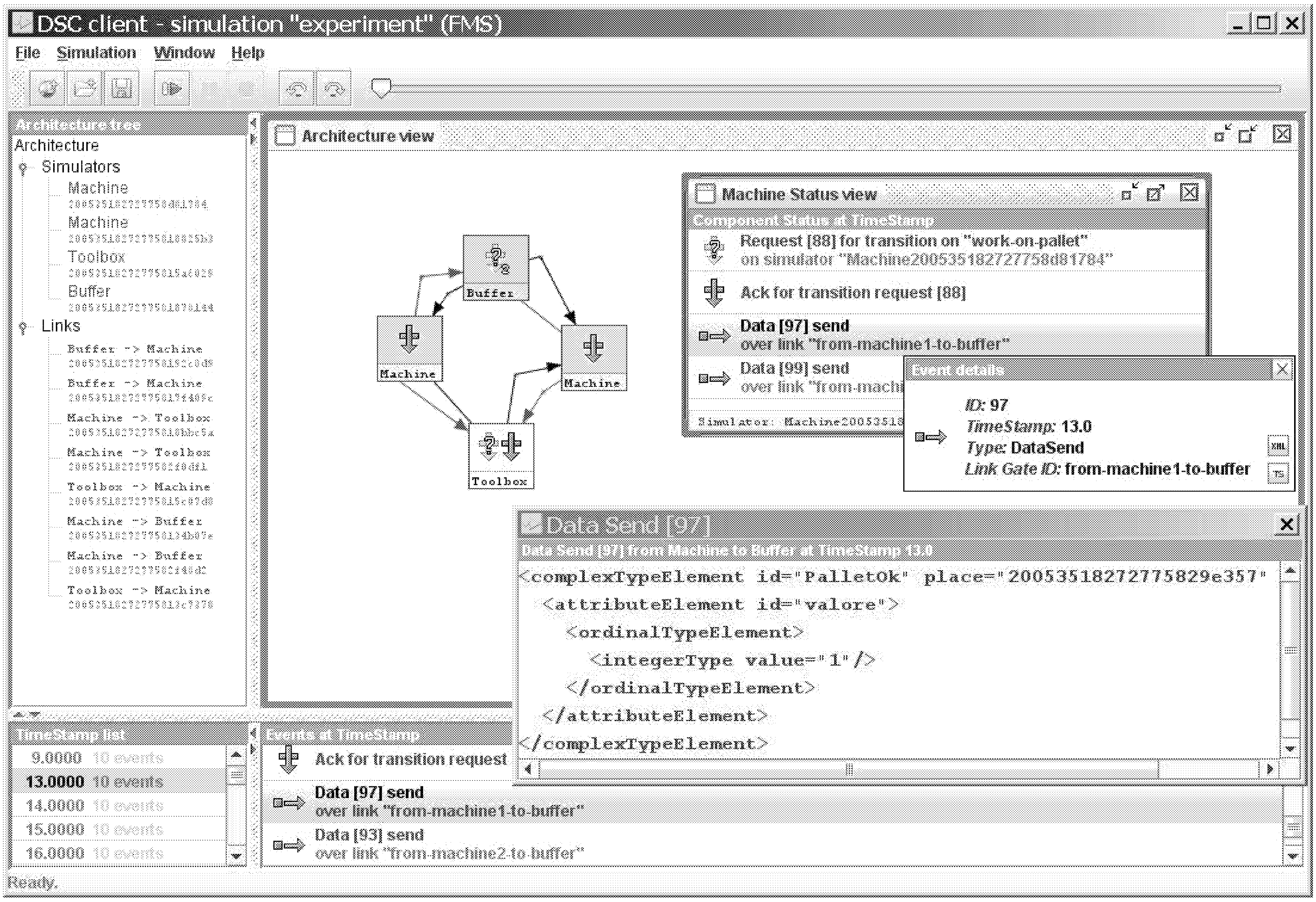
Therefore, the 2D representation mode, which provides a flat view of the System Architecture, even though is able to show all the relevant aspects of a simulation experiment, may not be expressive enough to abstract away from the complexity of the relationships among the simulators. Thus, an architecture composed of several components that share a great quantity of data, may take advantage from a 3D representation. In fact, the complexity of the interconnection among the simulators can be reduced by providing a third dimension to the Simulation Architecture. As an example, consider a couple of simulators that communicates through some tenths of ports; the 2D representation would show a tangle of connections, which would be confusing for the user. On the contrary a 3D representation would depict the connections also along the third dimension. Notice that 2D and 3D representations are simply different views on the current state of the same model and it is up to the user to decide which visualization mode is more suited to his/her purposes.

The 3D simulation architecture provides a schematic representation of the components and an intuitive representation of the connections among them. Simulators maintain the same graphical characteristics introduced by the 2D representation, they are labeled and they show their current state using different colors. Such visualization mode provides additional interaction capabilities. The user can exploit predefined viewpoints to observe the simulation, focusing on particular details related to the components in the foreground. The GUI provides buttons to move from one viewpoint to another. Therefore, the user can navigate the simulation scene coming close to specific simulators or connectors.

A user, clicking on the representation of a simulator obtains a description of its computational state. As a consequence the DSC Client shows a dedicated window (see figure 2) that lists all the events associated with the simulator. The same functionality is implemented for connections. Both 2D and 3D visualization frames provides such interaction capability. An example of the 3D visualization frame is reported in figure 3. The screen shot shown refers to the execution of the same simulation experiment previously described.

The 3DML approach

The 3D visualization capabilities of the DSC Client make use of 3DML (http://www.flatland.com/), a 3D modeling language, and of an open-source application, called Open3DML, specifically designed to support the visualization and navigation of 3D scenes described with 3DML. Open3DML is a general-purpose interpreter for the 3DML language that can be used in different contexts (Colombo et al. 2006).



Architecture
Simulators
Machine 200535182727758d81784
Machine 20053518272775810835b3
Toolbox 2005351827277581 5a6028
Buffer 20053518272775818 78144
Links
Buffer -> Machine 200535182727758192c0d9
Buffer -> Machine 20053518272775819f409c
Machine -> Toolbox 20053518272775810bbc9a
Machine -> Toolbox 2005351827277580f0d81
Toolbox -> Machine 2005351827277581 5c07d0
Machine -> Buffer 20053518272775812 4b07e
Machine -> Buffer 2005351827277580f48d2
Toolbox -> Machine 20053518272775812c7378

Request [88] for transition on "work-on-pallet"
on simulator "Machine200535182727758d81784"

Ack for transition request [88]

Data [97] send
over link "from-machine1-to-buffer"

Data [99] send
over link "from-machi

Simulator: Machine20053518

ID: 97
TimeStamp: 13.0
Type: DataSend
Link Gate ID: from-machine1-to-buffer

Data Send [97]

Data Send [97] from Machine to Buffer at TimeStamp 13.0

```
<complexTypeElement id="PalletOk" place="200535182727758 29e357"
  <attributeElement id="valore">
    <ordinalTypeElement>
      <integerType value="1"/>
    </ordinalTypeElement>
  </attributeElement>
</complexTypeElement>
```

9.0000	10 events
13.0000	10 events
14.0000	10 events
15.0000	10 events
16.0000	10 events

Ack for transition request

Data [97] send
over link "from-machine1-to-buffer"

Data [93] send
over link "from-machine2-to-buffer"

Figure 2: DSC Client GUI with the 2D visualization frame

The 3DML language is a XML (Harold 2004) based language designed to allow direct description of 3D virtual environments. It focuses on 3D content creation for Web applications; examples of use of this language can be found in tourist context like virtual museum (Virtual Open Air Museum Latvia, http://www.virmus.lv) and virtual cities (Isbister and Doyle 1999).

The 3DML modeling approach is based on the approximation of the structure of a scene with a static 3D grid and on the composition of the final scene with predefined 3D objects called “blocks”.

A 3DML virtual scene is composed by instances of “blocks” disposed over a regular 3D grid. Each instance is a 3D polygonized structure with associated appearance attributes (as textures and colors) and behavioral ones (functions provided by the single 3D objects). Each block provides services to change the properties of its instances by means of transformations as translations, rotations, scaling and the overriding of the appearance attributes. Such capabilities allow one to define, inside a 3DML document, services that can be invoked during the scene navigation in order to change the characteristics of the scene.

Applying the modeling approach

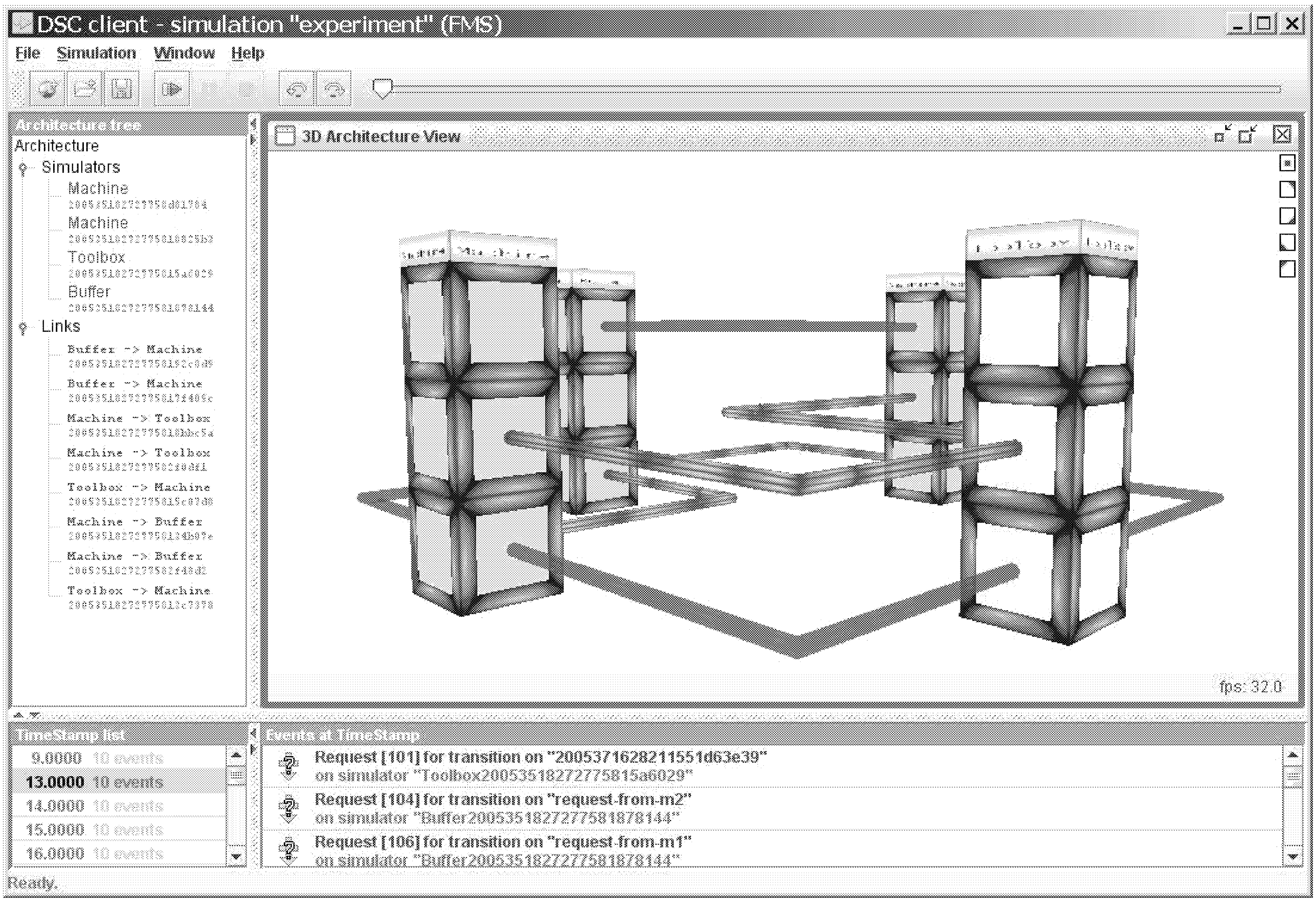
3DML is suitable to define simple scenes that satisfy the aforementioned modeling requirements: a simulation architecture is described as composed by “boxes” (parts of a simulators) and “tubes” that model the connectors defined among the simulators. Such scenes are visualized by the Open3DML engine, which provides also interaction capabilities with the user.

Open3DML has been configured to allow DSC to directly invoke services provided by a scene (like they were local methods of an Open3DML class). As a consequence, DSC interacts with Open3DML to change the characteristics of the currently loaded scene triggering functions provided by the same 3DML scene. Such mechanisms are helpful to deal with animations that keep track of the evolution of the components involved in a simulation experiment.

Exploiting such interaction mechanism, “3D Boxes” change color as a consequence of changes in the computational state of the component that they are modeling. The same mechanism is applied to the “3D tubes” elements as well.

3DML approach assessment

3DML is well suited to deal with the definition of 3D Simulation Architectures. In this context 3D is mainly used to decrease the visualization and management complexity of connections among the simulators. Such characteristics make 3DML an easy and cost effective solution. Simulation scenes exploit the capabilities provided by Open3DML. For example, the 3D scenes modeler can define functions that could be invoked at simulation time to change the characteristics of the scene. Moreover, both Open3DML and DSC Client are open-source solutions written in Java and this allows one to extend their capabilities whenever he/she needs.



9.0000 10 events
13.0000 10 events
14.0000 10 events
15.0000 10 events
16.0000 10 events

Request [101] for transition on "200537162821155 1d63e39"
on simulator "Toolbox20053518272775815a6029"

Request [104] for transition on "request-from-m2"
on simulator "Buffer20053518272775818 78144"

Request [106] for transition on "request-from-m1"
on simulator "Buffer20053518272775818 78144"

Figure 3: DSC Client GUI with the 3D visualization frame

RELATED WORK

In the main field of simulator integration the High Level Architecture (HLA) (Kuhl et al. 2000) provides a framework to describe simulation applications. The main goal of HLA is to facilitate interoperability among simulations and to promote reuse of simulations and their components. HLA describes simulations in terms of federations of federates, where a federation is a simulation system composed of two or more simulator federates communicating through the Run-Time Infrastructure (RTI).

Both HLA and SINPL address the problem of integrating simulators to allow one to execute distributed simulations. However, there are some differences between the two approaches. The more relevant for our goals is that SINPL provides an approach having a higher level of abstraction with respect to HLA because of its integrated design environment. Moreover, the practical use of HLA requires highly skilled people because of its inherent complexity.

Research and development efforts on Web-based simulation include server side simulations, client side simulations, and distributed Web-based simulators. The first two approaches allow one to access already existing simulation environments through the Web, while the third one consists of an extension of the distributed simulation architecture to Web-based network infrastructures.

There are several Java-based simulation libraries that permit the creation of simulation programs as Java applications and applets. Among these are JSIM (Miller et al. 2000a), JavaSim (Little 2001), SimJava (McNab and Howell 1996), DEVSJAVA (Sarjoughian and Zeigler 1998) and Simkit (Buss and Stork 1996).

Researches were done also in the field of distributed simulations on the Web, with components running on different machines. Multiple users can interconnect with the same underlying simulation model through Web browsers from different locations.

Different technologies are investigated to find an ideal substrate for Web-based simulation, for example Remote Method Invocation (RMI) (Page et al. 1997), Web services (Chandrasekaran et al. 2002), component technologies such as Enterprise Java Beans (EJB) and Jini (Miller et al. 2000b).

Visualization in the simulation context is extensively used in modelling, execution and analysis phases. In addition, 3D visualization can be used in order to enhance the expressiveness of the visualization. An example of general-purpose 3D visualization system for animating processes modelled using discrete event simulation tools can be found in (Kamat and Martinez 2004). Training environments (De Lara and Alfonseca 2003) and immersive simulations (Wenzel and Jessen 2001) are simulation fields in which 3D visualization is particularly helpful. Furthermore, the coupling of Java3D and Web technologies has allowed the integration of 3D visualization in Web-based simulation environment (Salisbury et al. 1999).

More recently, Extensible 3D (X3D) mark-up language (Web3D 2006) became the de facto standard for modelling 3D scenes in the Web context. However, the modelling approach supported by X3D is demanding and time consuming for the inherent complexity of the language; in the opposite 3DML, despite providing a limited expressiveness, support a simpler modelling approach and is expressive enough to deal with our modelling context. Moreover, the 3DML modelling approach takes advantage of Open3DML, a Java-based interpreter for 3DML, which appears thinner than other 3D modelling language interpreters and provides very good rendering performances.

CONCLUSIONS

In the present work we introduced a simple and effective design process for distributed discrete event simulations. Moreover, we described the architectural characteristics of the SINPL platform, focusing on its components and the infrastructure used for integration and communication tasks.

We discussed the visualization requirements for the analysis task in the simulation context, describing two different implementations that fulfil them. We introduced a 2D graphical user interface, which operates the visualization by means of coloured connected graphs, and a 3D one, which exploits the rendering of three-dimensional blocks used to depict simulators and connectors. The latter implements the 3DML approach based on the usage of the 3DML modelling language, and of Open3DML, a performing open-source interpreter for this language.

A future enhancement of the 3D representation could concern the visualization granularity degree. More specifically, we aim at supporting a more realistic (less schematic) representation that would require both the visual rendering of the data exchanged by simulators and possibly the animation of the simulation scene. This feature would clearly improve the visualization soundness, allowing the user to immediately understand the state of the several components involved in a simulation experiment. This improvement would provide a deeper exploitation of the capabilities of both the modelling language and the interpreter.

ACKNOWLEDGEMENTS

We would like to thank Riccardo Gerosa for helping us in the design and development of the SINPL platform.

REFERENCES

Buschmann, F.; Meunier R.; Rohnert H.; Sommerlad P.; and Stal, M. 1996. "Pattern-Oriented Software Architecture: A System Of Patterns." John Wiley & Sons Ltd., West Sussex, England.

Buss, A.H. and Stork, K.A. 1996. "Simulation on the World Wide Web Using Java." In *Proceedings of the 1996 Winter Simulation Conference* (Coronado, CA, December 8-11 1996), 780-785.

Carullo, M.; Zanzi A.; Gallo, I; and Coen-Porisini, A. 2006. "An events synchronization approach for integration of simulators in a distributed environment". In *Proceeding of Industrial Simulation Conference – ISC 2006* (Palermo, Italy, June 5-8 2006), 74-78.

Chandrasekaran, S.; Silver, G. ; Miller, A.J.; Cardoso, J.; and Sheth, A.P. 2002. "Web service technologies and their synergy with simulation." In *Proceedings of the 2002 Winter Simulation Conference* (San Diego, CA, December 8-11 2002), 606-615.

Chandy, K.M. and Misra, J. 1979. "Distributed Simulation: A Case Study in Design and Verification of Distributed Programs." *IEEE Transactions on Software Engineering*, 5, 440-452.

Coen-Porisini, A.; Gallo, I.; and Zanzi, A. 2004. "Designing and enacting simulations using distributed components." In

Computer and Information Sciences – ISCIS 2004. Proceedings of the 19th International Symposium (Kemer-Antalya, Turkey, October 27-29 2004). Springer, 706-717.

Coen-Porisini, A.; Gallo, I.; and Zanzi, A. 2006. “Integration of Web-based simulators in the SINPL platform.” In *Proceeding of European Simulation and Modelling Conference – ESM 2006* (Toulouse, France, October 23-25 2006), 259-263.

Colombo, P.; Grosso, E.; and Tarini, M. 2006. "A Web-based solution supporting the integration of virtual reality environments in logistics applications." In *Proceeding of EUROMEDIA 2007* (Delft, The Netherlands, April 25-27 2007).

De Lara, J. and Alfonseca, M. 2003. “Visual Interactive Simulation for Distance Education.” *SIMULATION 79*, No.1,19-34.

Harold, E.R. 2004. “XML Bible 1.1 – 3rd edition”. Wiley Publishing Inc., Indianapolis, Indiana.

Isbister, K. and Doyle, P. 1999. “Touring Machines: Guide Agents for Sharing Stories about Digital Places.” In *Proc. AAAAI Fall Symposium On Narrative Intelligence*.

Jensen, J. 1997. “Coloured Petri Nets. Basic Concepts, Analysis Methods and Practical Use.” Vol. 2, Analysis Methods, Monographs in Theoretical Computer Science, Springer-Verlag (2nd corrected printing).

Kamat, V.R. and Martinez, J.C. 2004. “General-purpose 3D animation with VITASCOPE.” In *Proceedings of the 2004 Winter Simulation Conference* (Washington, D.C., December 5-8 2004), 1691-1697.

Kuhl, F; Weatherly, R.; and Dahmann, J. 2000. “Creating computer simulation systems – An Introduction to the High Level Architecture.” Prentice Hall PTR.

Kuljis, J. and Paul, R.J. 2000. “A review of web based simulation: whither we wander?” In *Proceedings of the 2000 Winter Simulation Conference* (Orlando, FL, December 10-13 2000). ACM, 1872-1881.

Little, M.C. 2001. “JavaSim User's Guide. Public Release 0.3, Version 1.0.” University of Newcastle upon Tyne.

Matta, A.; Tolio, T.; Tomasella, M.; and Zanchi, P. 2004. “A detailed UML model for general flexible manufacturing systems.” *Proceedings of the 4th CIRP International Seminar on Intelligent Computation in Manufacturing Engineering CIRP ICME '04*. (Sorrento, Italy, June 30- July 2), 113-118.

McNab, R. and Howell, F.W. 1996. “Using Java for Discrete Event Simulation.” In *Proceedings of the Twelfth UK Computer and Telecommunications Performance Engineering Workshop*, (University of Edinburgh, UK), 219-228.

Miller, J.A; Seila, A.F.; and Xiang, X. 2000a. "The JSIM Web-Based Simulation Environment," *Future Generation Computer Systems (FGCS),* Special Issue on Web-Based Modeling and Simulation, Vol. 17, No. 2 (October 2000). Elsevier North-Holland, 119-133.

Miller, J.A; Seila, A.F.; and Tao, J. 2000b. “Finding substrate for federated components on the web.” In *Proceedings of the 2000 Winter Simulation Conference*, (Orlando, FL, December 10-13 2000). ACM, 1849-1854.

Page E.H.; Moose, R.L.; Sean, J.; and Griffin, P. 1997. “Web-based simulation in SimJava using remote method invocation.” In *Proceedings of the 1997 Winter Simulation Conference*. (Atlanta, GA, December 7-10 1997). ACM, 468-474.

Salisbury, C.F.; Farr, S.D.; Moore, J.A. 1999. “Web-based simulation visualization using Java3D.” In *Proceedings of the 1999 Winter Simulation Conference*, (Phoenix, AZ, December 5-8 1999), 1425-1429.

Sarjoughian, H.S. and Zeigler, B.P. 1998. “DEVSJAVA: Basis for a DEVS-based collaborative M&Senvironments.” In *Proceedings of the 1998 International Conference on Web-Based Modeling & Simulation* (San Diego, Ca, January 11-14 1998), 29-35.

Sun Microsystems, Inc. 2005. “Java Web Start Overview – White Paper.”

Web3D Consortium. 2006. Extensible 3D (X3D). ISO/ISC 19775.

Wenzel, S. and Jessen, U. 2001. “The integration of 3-D visualization into the Simulation-based planning process of logistic systems”. *SIMULATION 77,* No.*3-4*, 114-127.

BOILING DOWN EMERGENT SELF-ORGANIZING SOUPS TO SOLID MULTIMODAL PERCEPTION

J.C. Stevens
Man-Machine-Interaction Group
Delft University of Technology
Mekelweg 4, 2628 CD Delft
The Netherlands
Netherlands Defence Academy (NLDA)
P.O. Box 10000, 1780 CA Den Helder
j.c.stevens@tudelft.nl

R. Dor and L.J.M. Rothkrantz
Man-Machine-Interaction Group
Delft University of Technology
Mekelweg 4, 2628 CD Delft
The Netherlands
{r.dor, l.j.m.rothkrantz}@ewi.tudelft.nl

KEYWORDS
Audiovisual perception, Auditory scene analysis, Copycat, Emergent systems, Gestalt, Information fusion, Multimodal, Self-organization.

ABSTRACT

Multimodal information fusion helps humans in dealing with ambiguous circumstances and improving their situational understanding and awareness. To mimic human audiovisual perception we propose a model for multimodal fusion based on emergence and self-organization. The proposed model integrates top-down influences with bottom-up pressures of both auditory and visual modalities, and is based on our ongoing research on computational self-organizing models of primitive auditory and visual perception. Ongoing work on the current models of perception supports the plausibility of implementing a successful audiovisual fusion model for coping with real world input. In addition, the model conforms with neuropsychological evidence.

INTRODUCTION

Humans unconsciously utilize audiovisual information fusion continuously. For example, when listening to a speaker, we also tend to look at her lip movements, which help us improve speech recognition by utilizing the complementary information in vision and audition. Not only do we receive more information using multiple senses, but multimodal processing can help us resolve ambiguous information within any single modality. This may enhance our situational understanding and awareness, which, from an evolutionary point of view, helps us to survive. Consequently, multimodal information fusion has been applied in numerous military applications including ocean surveillance, air-to-air defense, battlefield intelligence, surveillance and target acquisition, and strategic warning and defense (Hall 2001). Other non-military applications often focus on enhancing automatic speech recognition with visual features (Rothkrantz et al. 2005), person identity verification (Yacoub et al. 1999) or (multiple) speaker detection (Gatica-Perez et al. 2003) as applied to teleconferencing (Vermaak et al. 2001).

Humans are able to make sense out of the overwhelming amount of data received by their sensory organs. Traditionally, computational models of perception split human perception mechanisms into low and high level perception (Chalmers et al. 1992) regardless of the fact that all neuropsychological evidence points to mechanisms of deeply intertwined levels (Dor 2005). Such approaches tend to equate low-level perception with the primitive processing of the incoming data closer to the sense organs and high-level perception with mechanisms involving mental concepts such as object recognition and understanding. Most work to date in perception (mostly in the field of visual perception) has been targeted at either bottom-up processing (Viola and Jones 2001, Ullman 1996) or higher semantic levels (Mojsilovic and Gomes 2002). The main challenge for future models of perception is the integration of such top-down influences with bottom-up processing (Riesenhuber and Poggio 2000). Neurological evidence (e.g. Damasio 1995) suggests that multimodal fusion is only done at a later stage following the perception of each of the separate modalities. Moreover, though complex two-way communication channels exist between the centers of fusion and each modality, there is little evidence to suggest that any direct communication takes place between any two modality centers. Rather, the fusion mechanism feeds back down into each modality, applying top down pressures on the emerging percepts. It follows then, that in order to move towards a working fusion model, one would need to design an architecture which caters for both separate models of primitive perception for each modality, and for the deep integration with the mechanisms of the fusion model. Subsequently, such a scheme allows for the indirect intra-modality communication. Finally, we argue that any fusion model of perception should provide the infrastructure for the existence of both high-level multimodal concepts and active bottom-up and top-down mechanisms.

In this paper we propose an audiovisual fusion perception model which draws from our ongoing work on emergent self-organizing primitive auditory and visual perception models (Dor 2007, Stevens 2007).

The paper is organized as follows. In the ‘fusion models’ section, we present past work related to fusion models. We follow with ‘Copycat and The Ear’s Mind: emergent self-organization’ – in which we describe our current auditory and visual perception architectures together with the Copycat model (Mitchell 1993). Thereafter we present our proposed fusion model. We end the paper with some conclusions.

FUSION MODELS

Before describing the proposed architecture, a short survey of past fusion models is presented. While by no means exhaustive, this survey may help illuminating those aspects we deem crucial for the design of any fusion model of perception.

The JDL Process Model

The JDL model (Hall 2001) was developed in 1985 by the U.S. Joint Directors of Laboratories (JDL) Data Fusion Group and is the most widely used system for categorizing fusion related functions. Since the JDL model was tailored for the military domain, most applications using the JDL model serve military purposes such as (multiple) target tracking, threat assessment and identification.

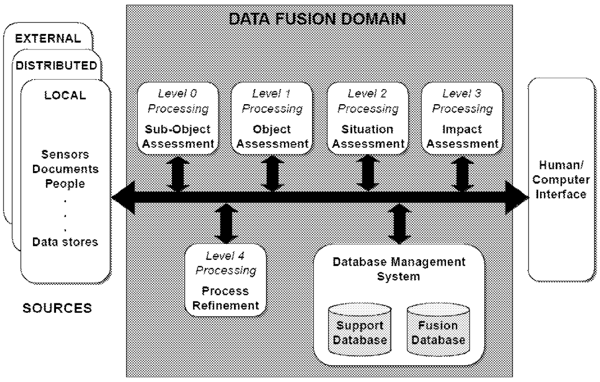


Figure 1: JDL process model
(taken from Hall 2001)

The JDL model, illustrated in figure 1, consists of the five levels:

- Level 0: dealing with sub-object data assessment i.e. the estimation and prediction of signal- or object-observable states on the basis of pixel/signal-level data association and characterization.
- Level 1: handling object assessment i.e the estimation and prediction of entity states on the basis of inferences from observations.
- Level 2: caters for situation assessment i.e the estimation and prediction of entity states on the basis of inferred relations among entities.
- Level 3: processing impact assessment i.e. the estimation and prediction of effects on situations of planned or estimated/predicted actions by the participants.
- Level 4: computing process refinement (an element of resource management) i.e. adaptive data acquisition and processing to support mission objectives.

Although JDL-inspired models have been significantly improved over the last two decades, they are still restricted almost entirely to low-level fusion (level 0 and 1 of the JDL model) while most information fusion is done within a single JDL level (Powell 2002). What makes the JDL model less than ideal for human audiovisual perception (a characteristic feature common to many fusion architectures), is the strict predefined arrangement of non-overlapping processing levels (Chalmers et al. 1992). As described later on, we believe it is essential for any model of perception to allow the ceaseless dynamic interaction among all levels. As a result, levels necessarily overlap, and end up dissolving and mingling with each other to the point where it is no longer relevant to speak of levels in the first place.

The Waterfall Model

The Waterfall model, illustrated in figure 2, is a hierarchical architecture for data fusion consisting of three levels of representation (Harris et al. 1998, Esteban 2005):

- Level 1: Working on the raw data providing the information about the environment.
- Level 2: Extracting and fusing features.
- Level 3: Relating objects to events. Decisions are made based on the gathered information and a-priori information.

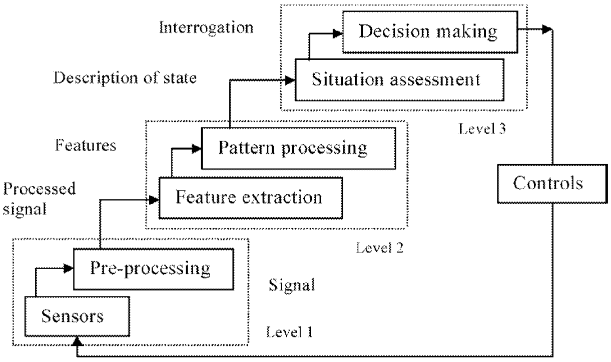


Figure 2: Waterfall model (taken from Esteban 2005)

As the figure clearly shows, the waterfall model is less abstract than JDL. Although it shares with it the static level separation, it goes even further in strictly specifying the direction and sequence of communication between the various levels. Moreover, the predefined feedback path from the highest level to the lowest necessarily restricts the range of top-down pressures the model can handle. Seen in this light, such a scheme may not be regarded as cognitively plausible for the purposes of perception in general, and certainly not for the implementation of multimodal perception in particular.

The Omnibus process model

In (Bedworth et al. 2000) the Omnibus process model is proposed, forming a hybrid of three other models, namely the ‘Boyd loop’, ‘Dasarathy’, and the previously described Waterfall model (Esteban 2005). Although the model has managed to solve some of the problems of its constituent modules in isolation, it still shares with its predecessors the same shortcomings described above making it also less suitable as a (multimodal) model of perception.

The Blackboard Model

The Blackboard model (Engelmore and Morgan 1988, Hou et al. 2000) was designed to allow local sources of knowledge, or experts, to have access to a central blackboard. All communication and interaction between the knowledge sources takes place indirectly using a central blackboard as illustrated in figure 3.

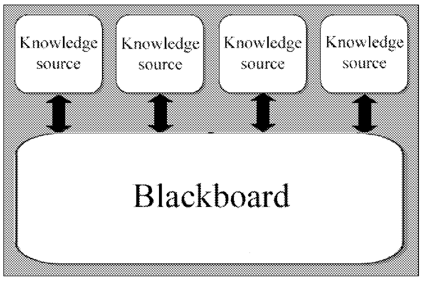


Figure 3: Blackboard model

Blackboard systems can be used in complex domains that require the application of different sorts of knowledge gained from one or more modalities. The Blackboard model is mainly used for high-level (decision-making) fusion. Although the abstract nature of the model does not directly cater for low-level processing, such mechanisms may be internally implemented within each agent as was the case with the Hearsay II speech-understanding project (Engelmore and Morgan 1988).

Audiovisual Gestalts

The Audiovisual Gestalts (Monaci and Vandergheynst 2006) model describes an algorithm for fusing audio and video data originating from a single physical phenomenon. They define the so-called 'meaningful audiovisual structures' as temporally proximal audiovisual events. Inspired by the research of Desolneux et al. on Gestalt theory and Computer Vision (Desolneux et al. 2003), the authors succeed in fusing the two modalities based on temporal synchrony. Although the model makes use of a single cross-modality grouping-strategy, it has proven capable of detecting cross-modal correlations existing in the audiovisual input in the presence of both distracting visual motion and acoustic noise.

As figure 4 shows, synchronized audio and video data is fed separately into parallel feature detectors, whose outputs are used for the final frame-based fusion. This seemingly simple strategy reveals the potential power of utilizing Gestalt grouping pressures for multimodal fusion. As described later on, our approach regards temporal synchrony as one of myriad grouping pressures that dynamically build (or destroy) structures using extracted features as input. This approach relieves one from the search for the ideal grouping strategy, and puts her in the realm of dynamic self-organization, where strong structures emerge through the processes of cooperation and competition, resulting in a coherent high-level representation of the input.

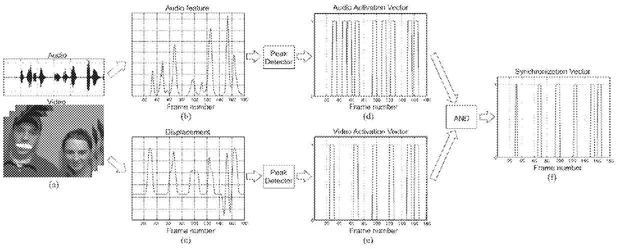


Figure 4: Audiovisual Gestalts
(taken from Monaci and Vandergheynst 2006)

COPYCAT AND THE EAR'S MIND: EMERGENT SELF-ORGANIZATION

Hearsay II, described in the previous section has been the inspiration for Hofstadter's Copycat model of analogy making (Hofstadter 1995, Mitchell 1993). Copycat, in turn, has served as the basic architecture on which our auditory (The Ear's Mind, Dor 2005) and visual perception models are built.

Copycat

The *Copycat* computer program (Mitchell 1993) models the mechanisms of analogy-making in a letter-string micro-domain. Together with other models (e.g. *Seek-Whence*, which tackles linear patterns, *Tabletop* (French 1995), tailored for two-dimensional visual analogies, and *Letter Spirit*, which generates creative font variations), Copycat belongs to a lineage of stochastic sub-symbolic self-organizing cognitive models (Hofstadter 1995).

Copycat is based on the assertion that analogy-making at any given level (e.g. seeing two situations as 'the same' even when no one-to-one correspondences exist among their respective constituent elements) relies on emergent mechanisms at lower levels. Analogy is thus seen as an interpretation of a given arrangement or situation arrived at by lower level activities. Internal pressures at such levels stir the constituent parts (including both atomic features and groups thereof) to form higher-level coherent structures. Consequently, the Copycat implementation allows Subcognitive pressures probabilistically influence the direction of processing. Both context-dependent and context-independent pressures make up a nondeterministic parallel architecture in both bottom-up and top-down directions. Moreover, since the resulting arrangements are not known in advance, such self-organizing systems could not be implemented by deterministic processes working at the same level of abstraction of the outcome itself. Instead, microscopic, nondeterministic, local processes interact with each other with no central control. The macroscopic outcome of such activities is emergent, rather than programmed.

Copycat's seemingly chaotic self-organizing behavior leading to context-sensitive order at higher levels is by no means limited to models of cognition. It takes place in a myriad of systems evolution has come up with such as the flexible resource management behavior of ants and bees, the fluid construction and maintenance of complex structures by termites and the schooling behavior of fish and birds (Marais 1937, Camazine et al. 2001). Microscopic behavior giving rise to flexible ordered macroscopic phenomena can even be seen in natural and chemical systems. Examples of such systems can be found in (Prigogine 1984).

The Ear's Mind

Available data from the field of *Auditory Scene Analysis* (ASA) (Bregman 1990) give compelling evidence to support a theory of perception based on the use of a whole array of active grouping pressures. Incoming sound is broken up into primitive elements (e.g. a sudden increase in the energy level at a given frequency) which are then put together like the pieces of a jigsaw puzzle to form a sensible account of the auditory scene. Or rather, primitive auditory elements actively exert grouping pressures like the constituents of self-organizing systems so that a sensible account of the auditory scene emerges. A mechanism is therefore needed which builds structures of cues based on the resulting coherence of the situation as a whole rather than on any predefined forms of the resulting structures. At the same time, *relations among* elements may have more weight than the elements' respective parameters in stirring the formation of the emergent description.

Constituent elements of any real-world object (e.g. harmonic sinusoids of a monkey call) share parameters and fate in the Gestalt sense. In general, and if not occluded or corrupted by other sounds, they have the tendency to share starting and ending points in time, follow parallel frequency trajectories, undergo the same frequency and amplitude modulations, etc. From an active, emergent perspective, constituent elements may form coherent structures if grouped together by following their shared parameters and fate. It seems that all auditory-capable animals have indeed evolved a repertoire of such active subcognitive auditory-cue grouping pressures. The key ingredient for implementing such mechanisms is the interaction – competition – cooperation among all pressures present at any given moment. This is necessary because of the contradictory and incomplete set of cues present at any real-world input caused, among other things, by occlusions, distortions, and reflections, not to mention transgressions from ideal sound production models. By letting these pressures actively push each other with no centralized interference, structures may emerge, which amount to a reconstruction of the shared fate of the constituent elements.

Seen from a different angle, it is the interpreted relation among constituent objects at different levels of description which brings about the description of a given situation. Perceptual illusions offer good examples of switching alternative interpretations when groups of cues are in disagreement with each other. Likewise, any perceived event will influence the interpretation of new situations and vice versa. This activity allows for the interpretation of vague structures (i.e. weakly self-supporting) to be influenced, or shaped by existing, or familiar ones. In other words, one situation may be *seen* as 'the same' as another. That is, a given situation may be interpreted as being the same as another, even though they share very little at the level of their constituent parts. The ability to fluidly describe and interpret different situations in relation to others suggests that the same underlying mechanisms are in play as those used by higher-level analogy-making.

Example of the Ear's Mind performance

The current implementation was tested with audio fragments of standard psychoacoustic experiments. To illustrate the working of the Ear's Mind, we include a single example of emergent self-organization of primitive auditory features (from track 1 of the ASA-CD, Bregman & Ahad). For a more detailed description of this and other examples, see (Dor 2007).

The input to the Ear's Mind is shown in figure 5. We have added the ovals to illustrate the groups of cues that humans perceive as the lowest level description of the fragment. The triangular cues represent salient first derivative features extracted by the Ear's Mind preprocessor (Dor 2005).

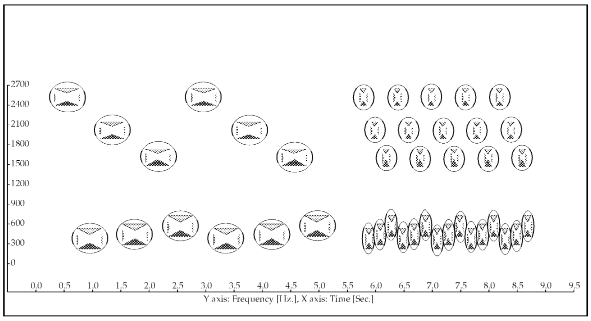


Figure 5: Extracted cues from a fragment of track 1 from ASA-CD (Bregman & Ahad). Ovals enclose what humans perceive as the lowest-level primitive units. These simple tones further segregate into higher-level streams

Since primitive visual and auditory segregation seem to share the same Gestalt principles, it turns out that the ovals in figure 5 enclose not only what humans would hear, but also, looking at the figure, would visually perceive as the most natural way of segregating the lowest level. Consequently, one mistakably tends to evaluate the given segregation as trivial, or straightforward. However, 'good' segregation (in the sense of both local and global coherence) is expected to self-organize itself by emergence from local activities rather than by any global blind processing, or a-priori knowledge of the resulting structure, making the problem challenging to solve even in this simple case. *The Ear's Mind* has no defined absolute measurement, or tuned parameters for the given input, nor does it make use of hard-wired heuristics for judging higher-level topologies (e.g. that shorter bonds should win at all times). In fact, *The Ear's Mind* does not have any knowledge of the desired resulting groups. The expectation is that as more cues are found and bonded, context will push towards building coherent structures, and conflicting bonds will be broken. Figures 6 to 9 contain snapshots from a run sequence and may help illustrating the emerging nature of the Ear's

Mind. These depict a detailed view of the right lower region of figure 5. The arcs represent bonds between the triangular cues in the figure. For clarity, both the cues and dead bonds (those which have been broken by competition along the way) were left out of the figure. At the beginning of the run a fair amount of wrong bonds are built (figure 6). As the run progresses more self-supporting structure emerges, leading more conflicting bonds to lose fights, and resulting in a coherent grouping of the given cues. In figure 9, the surviving bonds have built the desired structure (compare with figure 5).

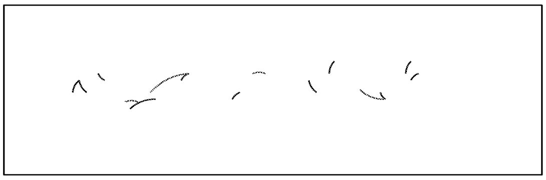


Figure 6: Context test, track 1, lower right region, emergent bond building a

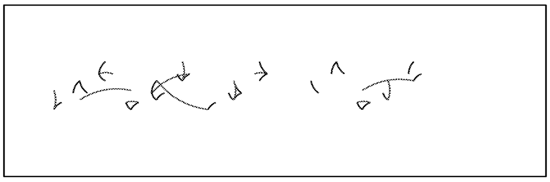


Figure 7: Context test, track 1, lower right region, emergent bond building b

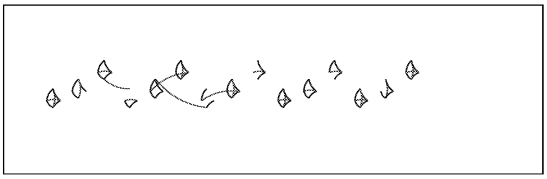


Figure 8: Context test, lower right region, emergent bond building c

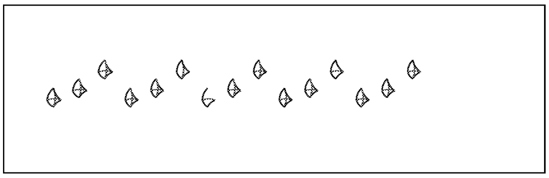


Figure 9: Context test, lower right region, emergent bond building d

For a more detailed discussion on The Ear's Mind architecture, as well as the motivation for its design from the fields of neuroanatomy, psychoacoustics, Gestalt and Auditory Scene Analysis, and the choice of basing The Ear's Mind on the Copycat architecture, see (Dor, 2005).

Visual Perception

Starting with The Ear's Mind we have designed a model for visual perception for implementing and testing visual dynamic grouping. The model is designed to segregate incoming input features into coherent subsets consisting of single objects or structures. Just like in The Ear's Mind, such grouping activity is done by emergence and self-organization using the cooperation and competition of any perceptual grouping pressures deemed relevant, such as (but by no means restricted to) the Gestalt laws of organization. The goal is to mimic human primitive visual perception using extracted features (salient cues) without the use of a-priori knowledge of the resulting higher-level structures, based on Gestalt and emerging context rather than on blind bottom-up image segmentation.

The visual perception model works with the same computational emergent mechanisms as in the Ear's Mind. However, its agents (Codelets) differ greatly from those of Ear's Mind as they are specifically designed to cope with visual features and cues. Furthermore, subsequent tests are planned for assessing the contribution of each of the original visual Gestalt grouping principles. The four currently implemented principles are listed below, and illustrated in figure 10.

- Proximity/Contiguity. Elements create pressures to be grouped together by proximity.
- Similarity. Similar elements create pressures to bond together.
- Good continuation. Perceptual mechanisms tend to preserve smooth continuity in favor of abrupt edges.
- Closure. Out Of several possible perceptual organizations, ones yielding 'closed' figures are more likely than those yielding 'open' ones.

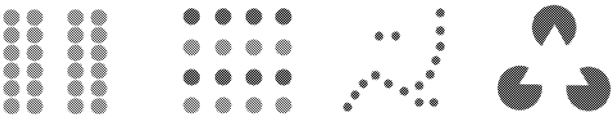


Figure 10: Gestalt principles from left to right, proximity, similarity, good continuation and closure

The assessment of how the Gestalt principles of proximity and similarity compete with one another during the process of organization has been done by Philip Quinlan and Richard Wilton. They have carried out controlled psychophysical experiments to statistically assess human grouping behavior by getting subjects to estimate grouping strength of various elements in visual displays (Quinlan and Wilton 1998). As Quinlan puts it:

> Each of the displays used contained a row of seven colored shapes with a middle target shape and two sets of flanking shapes. For each display subjects were asked to rate the degree to which the target shape grouped with either the left of the right set of

flankers. Across different displays the relationships between the target and the flankers were varied.

Figure 11 illustrates some examples of the actual displays used. More details on the setup of the experiments can be found in (Quinlan and Wilton 1998).

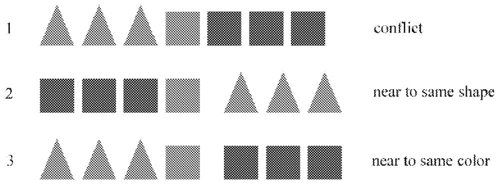


Figure 11: Examples of the actual displays used

In our ongoing work on visual perception we use Quinlan's experiments.

PROPOSED ARCHITECTURE

Multimodal fusion of audiovisual data offers a multitude of potential enhancements in comparison single modalities. One example we have seen earlier (see Audiovisual Gestalts in the 'fusion models' section) is the temporal synchrony fusion correlating synchronous events from the auditory and visual input channels. Such fusion may help in the segregation and tracking of audiovisual objects from a mix of other interfering objects, features, or noise at the input. Another situation where fusion may ease computational perception is in solving inconsistencies, missing information, bad signal to noise ratios, and ambiguities in (one of) the input modalities.

For illustrating purposes, in what follows, we shall briefly expose the proposed architecture using one such example. Abbreviations in square brackets refer to the corresponding items in figure 12: The auditory model of perception is designated with the letters AU (for Audio), the visual model with the letters VI, and the fusion with FU. Likewise, the following abbreviations stand for raw [IN]put, [PRE]processor, [SLIP]net, [WORK]space, and [CODE]rack. These terms, borrowed from the Copycat (Mitchell 1993) and The Ear's Mind (Dor 2005) models, will become clear as we go along.

Starting with the raw input, we have two separate streams entering at [IN-AU] (audio stream) and [IN-VI] (video stream), which are fed into preprocessors [PRE-AU] and [PRE-VI] respectively, where atomic features are extracted (Dor 2005 and Stevens 2007). These features (or cues) are then put on the respective workspaces of each modality, [WORK-AU] and [WORK-VI]. The workspace is the working area in which the features reside and all subsequent structures are built (or destroyed), roughly resembling a working memory. Notice (see figure 12) that each modality has its own workspace, as well as a dedicated fusion workspace, where fused audiovisual objects and structures are built.

Bottom-up Codelets residing on the Coderack [CODE-AU] are launched on the workspace and search for structures. Each Codelet is searching for different sorts of correspondences among features. When found, a Codelet may relate features by building a bond between them, propose to break existing structures, launch other Codelets by posting them on the Coderack (a stochastic stack on which all Codelets wait for their launch biased by their urgency), or activate concepts on the Slipnet [SLIP-AU]. The Slipnet is a network of interrelated concepts, which can be thought of as the long-term memory, where concepts may get activated, and spread activation (biased by a dynamic link length) to other linked (read associated) concepts. When a Slipnet concept gets activated, it may launch a top-down Codelet on the workspace for searching more instances of itself.

In our example, the auditory input might be an isolated human voice. Features on the workspace trigger the formation of structures through the process of self-organization, culminating in a construction and tracking of a coherent perceptual group bound to the Slipnet concept 'human voice' in [SLIP-AU]. The activation of this concept (as well as any other groups) spreads over (biased by the current context-dependent link lengths) to the 'human' concept in the fusion Slipnet [SLIP-FU], resulting in the construction of an instance on the [WORK-FU] workspace. All the while, no successful structures have been built on [WORK-VI] that might correspond to, or link with the human voice on the fusion Slipnet. In our example, this is because the images are too dark to detect the human form by vision alone. Consequently, when no structure on [WORK-FU] can be built with 'human', a fusion Codelet may activate the 'visual-expectation' concept in [SLIP-FU] which will then spread its activation over to [SLIP-VI]. Via this indirect communication (Damasio 1995) a top-down pressure will be launched onto [CODE-VI] to search for instances of human concepts (e.g. face, body, eyes, etc.) in [WORK-VI]. This might even involve top-down pressures all the way down to the alteration of preprocessing parameters.

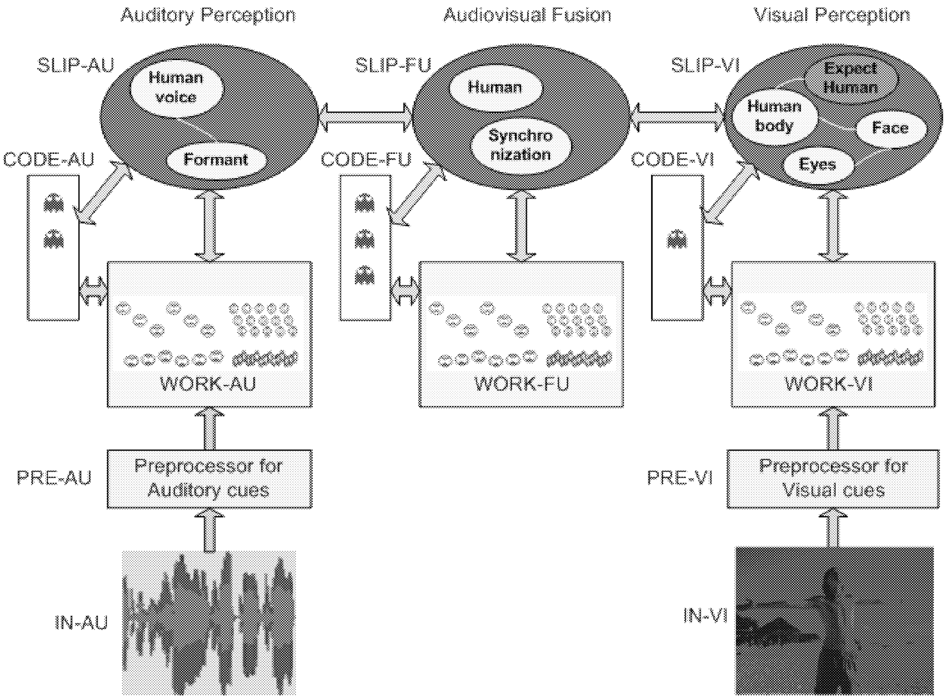


Figure 12: Proposed fusion architecture

If found, visual structures may then bubble up through [SLIP-VI] and [SLIP-FU] back to [WORK-FU], where top-down fusion Codelets may end up building audiovisual relations between the human form and his voice. In the same manner, implementing audiovisual synchrony as a fusion Codelet is feasible. Just as in The Ear's Mind, and its visual counterpart, the proposed fusion architecture supports the cooperation and competition among fusion Codelets, and the construction of higher level structures.

Since the proposed architecture principally shares the design with the two perceptual constituent models, it offers implementation advantages as well as sustaining the modularity of the system as a whole. The entire infrastructure, once implemented, may be tailored to different situations and domains by redefining specific preprocessing schemes, Codelets, Slipnet concepts and their interactions.

CONCLUSIONS

Following a survey of past general models of fusion and the audiovisual Gestalts model, we have briefly described our current models of auditory and visual perception. These are based on the Copycat model, and offer the advantages of emergent behavior and self-organization as plausible mechanisms for the implementation of computational models of perception. Finally, we have described the architecture of the proposed fusion model and exemplified its working using a simple multimodal scenario.

Currently our work consists of extending the two perception models for handling more complex cases, together with the completion of a preliminary proof of concept implementation of the proposed fusion model.

REFERENCES

Bedworth, M., Brien, J. O., and Jemity, M. 2000. The omnibus model: a new model of data fusion? *IEEE Aerospace and Electronic Systems Magazine*, 15(4):30–36.

Bregman, A. S. 1990. *Auditory scene analysis, the perceptual organisation of sound*. 2nd paperback ed. 1999, MIT Press.

Bregman A. S. & Ahad, P. A., *Demonstrations of Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press.

Camazine, S., Deneubourg, J., and Franks, N. R. 2001. *Self-organization in biological systems*. Princeton university press.

Chalmers, D. J., French, R. M., and Hofstadter, D. R. 1992. High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental and Theoretical Artificial Intellige*, 4:185–211.

Damasio, A. R. 1995. *Descartes' Error: Emotion, Reason, and the Human Brain*. Quill.

Desolneux, A., Moisan, L., and Morel, J.-M. 2003. Computational gestalts and perception thresholds. *Journal of Physiology* - Paris, 97:311–324.

Dor, R. 2005. *The ears mind: A computer model of the fundamental mechanisms of the perception of sound*. Technical report 05-16, Delft University of Technology.

Dor, R., Rothkrantz, L.J.M. 2007. *The ears mind: An Emergent Self-Organizing Model of Auditory Perception.* Submitted to the Journal of Experimental and Theoretical Artificial intelligence.

Engelmore, R. and Morgan, T. 1988. *Blackboard systems.* Addison-Wesley.

Esteban, J., Starr, A., Willetts, R., Hannah, P., and Bryanston-Cross, P. 2005. A review of data fusion models and architectures: towards engineering guidelines. *Neural Computing and Applications*, 14(4):273–281.

French, R. M. 1995. *The Subtlety of Sameness: A Theory and Computer Model of Analogy-Making*. The MIT Press.

Gatica-Perez, D., Lathoud, G., McCowan, I., Odobez, J.-M., and Moore, D. 2003. Audio visual speaker tracking with importance particle filters. *In International Conference on Image Processing*.

Hall, D. L. and Llinas, J. 2001. *Handbook on Multisensor Data Fusion*. CRC press, Boca Raton.

Harris, C. J., Bailey, A., and Dodd, T. J. 1998. Multi-sensor data fusion in defence and aerospace. *The Aeronautical Journal*, 102:229–244.

Hofstadter, D. R. and FARG 1995. *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought.* Basic Books, New York.

Hou, P., Shi, X., and Lin, L. 2000. Generic Blackboard Based Architecture for Data Fusion. *In the 26th Annual Conference of the IEEE Industrial Electronics Society (IECON)*, volume 2, pages 864–869.

Marais, E. N. 1937. *The soul of the White Ant.* Methuen & Co. Ltd. London.

Mitchell, M. 1993. *Analogy-Making as Perception: A Computer Model*. The MIT Press.

Mojsilovic, A. and Gomes, J. 2002. Semantic based categorization, browsing and retrieval in medical image databases. *In International Conference on Image Processing.*

Monaci, G. and Vandergheynst, P. 2006. Audiovisual gestalts. *In Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop* (CVPRW'06).

Powell, G. M. 2002. Opportunistic problem-solving in data fusion: The blackboard model. *In 23rd Army Science Conference*.

Prigogine, I. 1984. *Order Out of Chaos, mans new dialogue with nature.* Bantam books.

Quinlan, P. T. and Wilton, R. N. 1998. Grouping by proximity or similarity? competition between the gestalt principles in vision. *Perception*, 27:417–430.

Riesenhuber, M. and Poggio, T. 2000. Models of object recognition. *Nature Neuroscience*, 3:1199–1204.

Rothkrantz, L. J., Wojdel, J. C., and Wiggers, P. 2005. Fusing Data Streams in Continuous Audio-Visual Speech Recognition. In *Text, Speech and Dialogue: 8th International Conference*, Karlovy Vary, Czech Republic.

Stevens, J.C. 2007, *Intelligent Multimodal Information Fusion, An emergent system for primitive audiovisual perception,* White paper, Delft University of Technology.

Ullman, S. 1996. *High-level Vision: Object Recognition and Visual Cognition*. Cambridge, MA: The MIT Press.

Vermaak, J., Gagnet, M., Blake, A., and Perez, P. 2001. Sequential Monte-Carlo fusion of sound and vision for speaker tracking. *In Proc. IEEE Intl. Conf. on Computer Vision.*

Viola, P. and Jones, M. 2001. Robust real-time object detection. *In Second International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling.*

Yacoub, S. B., Abdeljaoued, Y., and Mayoraz, E. 1999. Fusion of Face and Speech Data for Person Identity Verification. *IEEE Transactions on neural networks*, 10(5):1065–1074.

WEB ENVIRONMENTS

FACTORS SHAPING THE USER EXPERIENCE ON UTILITERIAN WEBSITES

Teun Hompe, Joris Leker
valsplat | usability research
Prins Hendrikkade 22, 1012 TM
Amsterdam, the Netherlands
E-mail: teun@valsplat.nl
joris@valsplat.nl

Charles van der Mast
Man-Machine Interaction Group
Delft University of Technology
Mekelweg 4, 2628 CD
Delft, the Netherlands
E-mail: c.a.p.g.vandermast@tudelft.nl

Mark Neerincx
TNO Human Factors
P.O.Box 23, 3769 ZG
Soesterberg, the Netherlands
E-mail: mark.neerincx@tno.nl

KEYWORDS
Technology acceptance model, User experience, Enjoyment, Aesthetics, Interactivity

ABSTRACT

This paper explores factors that influence the user experience when using utilitarian websites. A theoretical model for the user experience of utilitarian websites is proposed and investigated. This model is an extension of the Technology Acceptance Model (TAM). The effects of perceived ease of use, perceived usefulness, perceived enjoyment, perceived visual aesthetics, internet anxiety and internet playfulness on the behavioural intention to use were examined. Most of these factors were adapted from earlier literature. The paper also introduces a new construct, perceived interactivity. A web survey was conducted, yielding a sample size of 147. The results confirmed 8 of the original 14 hypotheses. Perceived enjoyment and perceived visual aesthetics showed to be important factors in shaping the user experience.

INTRODUCTION

Usability is defined as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO 9241-11). Traditionally, human-computer interaction literature focused primarily on the effectiveness and efficiency when using a system, considering satisfaction to be a product of the former two. Reeves and Nass were among the first to show that this view of human-computer interaction was far from complete (Reeves and Nass 1996). Their research showed that “People’s responses to media are fundamentally social and natural.”

The work of Reeves, Nass and others has inspired more and more researchers to empirically study other factors, beyond those associated with the traditional view on usability, affecting information technology use in general, and web use in particular. The interplay of these new and traditional factors is usually called the user experience (often abbreviated as UX). Lindgaard and Dudek (2003) encourage this progress in the HCI field with the following notion: “productivity is probably not the prime motive driving people when they play computer games or surf the Internet in their leisure time. It is therefore reasonable to assume that the notion of ‘user satisfaction’ will rest on quite different criteria in the two environments [...] in either case we believe that ‘user satisfaction’ is the subjective sum of the interactive experience.” (Lindgaard and Dudek 2003, p. 430).

The new factors introduced in user experience research usually cover moods, emotions and feelings, and are usually grouped under the umbrella term ‘affect’. Affect is a set of mechanisms that rapidly evaluate events to provide an initial assessment of their valence or overall value relative to the person. People’s affective responses are often more stable and consistent across individuals than reason-based (cognitive) assessments (Pham et al. 2001).

Another important notion in the context of the user experience is the differentiation between hedonic or utilitarian aspects. The main purpose of a utilitarian website is to increase the user’s task performance and to support efficiency. All of the functionality of utilitarian websites is aligned with task requirements and distraction is kept at a minimum. Hedonic websites aim to provide self-fulfilling value to the user. The term hedonic derives from the word hedonism. Hedonism is the doctrine holding that behaviour is motivated by the desire for pleasure. The dominant design objective of hedonic websites is to encourage prolonged use.

Another perspective on utilitarian and hedonic websites considers the users motivation visiting a particular website. Generally, people are believed to operate based on two kinds of motivations: extrinsic motivations and intrinsic motivations. Extrinsic motivation is defined as the performance of an activity because it is perceived to be instrumental in achieving valued outcomes that are distinct from the activity itself. Intrinsic motivation refers to the performance of an activity for no apparent reinforcement other than the process of performing the activity per se and the interest and enjoyment that accompanies this use.

The aim of this paper is to further *explore* the role of qualities beyond usefulness and usability in shaping the web user experience of utilitarian websites and to *develop* and *empirically validate* a theoretical model based on the Technology Acceptance Model.

THEORETICAL BACKGROUND AND PROPOSED MODEL

One line of research into the user experience has its roots in Technology Acceptance Literature. The Technology Acceptance Model (TAM) was proposed by Davis to explain the formation of attitude and acceptance of information systems (Davis 1989). TAM itself is based on Fishbein and Ajzen 's Theory of Reasoned Action (TRA) (Fishbein and Ajzen 1975).

TRA is built on three core constructs: intention to use, attitude towards behaviour and subjective norm (beliefs). The intention to use is considered to be a function of attitude towards behaviour and subjective norm. A person forms an attitude about a certain object based on certain beliefs. Based on this attitude he or she forms an intention to use the object.

TAM (figure 1) is an adaptation of TRA, tailored to an information technology context, and considers two cognitive beliefs, perceived usefulness and perceived ease of use, to be the main factors driving a user's behavioural intention to use. Several studies in psychology and IT have proven that the behavioural intention to use is a strong predictor of actual use behaviour. Although originally intended for understanding the acceptance of job-related information technology in a work environment, it has proven to be accurate for World Wide Web related technology, more specifically (e-commerce) websites as well. The original TAM also included attitude as a factor explaining intention to use or actual use. Attitude was the only affect related factor in the model. However, attitude was excluded from the original TAM due to its weak mediating effects on the relationship between perceived usefulness and intention to use.

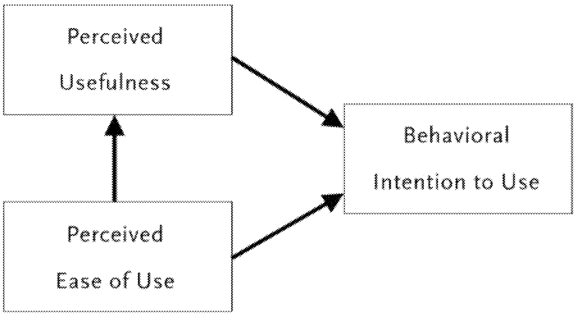


Figure 1: Technology Acceptance Model (Davis 1989).

The theoretical model used in this research will be based on the technology acceptance model. The TAM is chosen as the basis because of it has regularly been used as a basis by other researchers and has proven its value over and over again. Choosing a common model as basis will make it easier to compare the findings of this research with previous findings and with any future findings of researchers also employing the TAM.

The first three hypotheses come from numerous studies on the technology acceptance model. Perceived usefulness (PU) is defined as "the degree to which a person believes that using a particular website will enhance his or her performance in purchasing and/or information seeking". PU has been found to significantly influence the intention to use. Perceived ease of use (PEOU) is defined as "the degree to which a person believes that using a particular website is free of effort". PEOU has repeatedly been shown to be a strong determinant of behavioural intention to use. Furthermore PEOU has shown to have an indirect influence through PU.

H1: *Perceived usefulness of a website has a positive effect on the behavioural intention to use a website.*

H2: *Perceived ease of use of a website has a positive effect on the behavioural intention to use a website.*

H3: *Perceived ease of use of a website has a positive effect on the perceived usefulness of a website.*

In addition to employing previous measures of ease of use and usefulness, antecedents specific to the interaction with websites were sought. Earlier literature on user experience, or similar concepts, was studied to isolate useful constructs. In 1992 TAM was extended with the construct of perceived enjoyment (ENJ) (Davis et al. 1992). ENJ refers to the extent to which the activity of using a computer system is perceived to be enjoyable in its own right aside from the instrumental value of the technology. Various studies have shown that ENJ is a determinant of behavioural intention to use. Besides, PEOU has shown to be a determinant of ENJ.

H4: *Perceived ease of use of a website has a positive effect on the perceived enjoyment of a website*

H5: *Perceived enjoyment of a website has a positive effect on the behavioural intention to use the website.*

Interactivity has not very often been operationalized in a TAM context. However, some research has been done on the effect of interactivity on a user's attitude towards a website or a user's satisfaction with a website. Interactivity has shown to be a determinant of PU, PEOU and ENJ. To measure perceived interactivity an interactivity scale developed by Liu was used (Liu 2003). Items from the active control and synchronicity factors were used. Active control (AC) measures a user's ability to choose information and guide the interaction. Synchronicity (SYNC), or responsiveness, indicates the timing of information exchange.

H6: *Perceived interactivity of a website has a positive effect on the perceived usefulness of the website.*

H7: *Perceived interactivity of a website has a positive effect on the perceived ease of use of the website.*

H8: *Perceived interactivity of a website has a positive effect on the perceived enjoyment of the website.*

Various researchers have extended TAM with constructs aimed at measuring the user's appreciation of the design of an interface. Recent work by Lavie and Tractinsky (Lavie and Tractinsky 2004) showed a promising start in standardizing and operationalising aesthetics. Lavie and Tractinsky identified two dimensions, which were labelled 'classical' (CAEST) and 'expressive' aesthetics (EAEST). The items used to operationalise perceived visual aesthetics (PVA) in this study were therefore adapted from these two dimensions. PVA has shown to be a determinant of PU, PEOU and PE.

H9: *Perceived visual aesthetics of a website has a positive effect on the perceived usefulness of the website.*

H10: *Perceived visual aesthetics of a website has a positive effect on the perceived ease of use of the website.*

H11: *Perceived visual aesthetics of a website has a positive effect on the perceived enjoyment of the website.*

Social cognitive theory states that behaviour is influenced by cognitive, affective and personal determinants. In our research the effect of two of those determinants, internet anxiety (cognitive and affective) and internet playfulness (personal), are studied.

The construct of internet anxiety (ANX) is based on the concept of computer anxiety. Computer anxiety refers to fear of computers, or the tendency of a person to be uneasy or apprehensive towards current or future use of computers.

Internet playfulness (PLAY) refers to the degree of spontaneity a user exhibits when interacting online. The internet playfulness construct is based on the conceptually similar construct of computer playfulness, originally defined as "the degree of cognitive spontaneity in microcomputer interactions."

Both ANX and PLAY have been shown to be determinants of PEOU and ENJ.

H12: *Internet anxiety has a negative effect on the perceived ease of use of the website.*

H13: *Internet anxiety has a negative effect on the perceived enjoyment of the website.*

H14: *Internet playfulness has a positive effect on the perceived ease of use of the website.*

H15: *Internet playfulness has a positive effect on the perceived enjoyment of the website.*

The resulting theoretical model used in this research is displayed in figure 2.

RESEARCH METHODOLOGY

To empirically test the proposed theoretical model and the hypothesis a field study was conducted using a questionnaire to collect data. Items used in this questionnaire were selected form earlier academic literature involving measurements of user experience, preferably in a technology acceptance context. All items were measured using a 7-point Likert scale, were 1 = completely disagree, 4 = neutral and 7 = completely agree.

The stimuli used were three hotel reservation sites (figure 3). Subjects were evenly distributed over these three sites and asked to try and find a couple of suitable hotels in the Brussels (Belgium) area, for two persons, in the weekend of the 14th till the 16th of July. The three sites used were Best-stay.com, Hotelclub.com and Paguna.com. Sites were chosen based on their level of interactivity and visual design.

Subjects used for this research were randomly picked from the valsplat test-subject database. Overall, of the 708 that were distributed, 154 usable questionnaires were received and used for analysis, giving a response rate of 22 percent. All of the subjects had prior experience with the use of the WWW. Sixty-five percent of the respondents were female, and 87 percent have more than four years of experience with the WWW.

The data generated with the survey was coded and analyzed using SPSS version 11.5. In the first step the construct validity, the extent to which a measure accurately reflects the concept that it is intended to measure, of the survey was analyzed using factor analysis. In the next step construct reliability, the extent to which a variable or set of variables is consistent in what it is intended to measure, was examined by computing Cronbach's alpha. In the final step the theoretical model and related hypotheses were tested using multiple-regression.

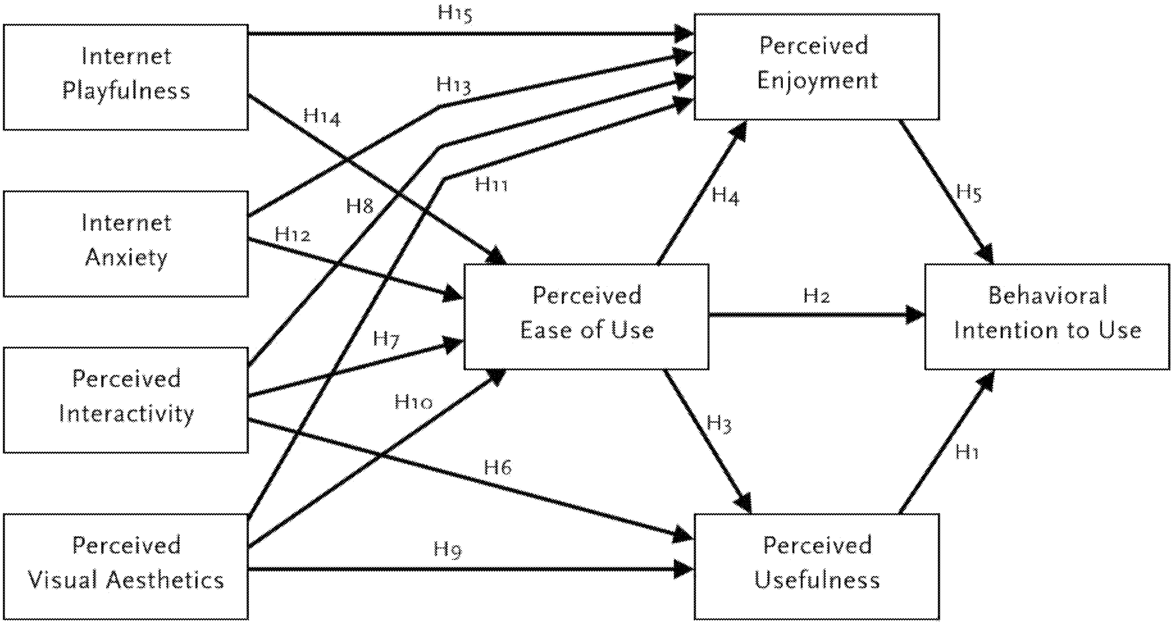
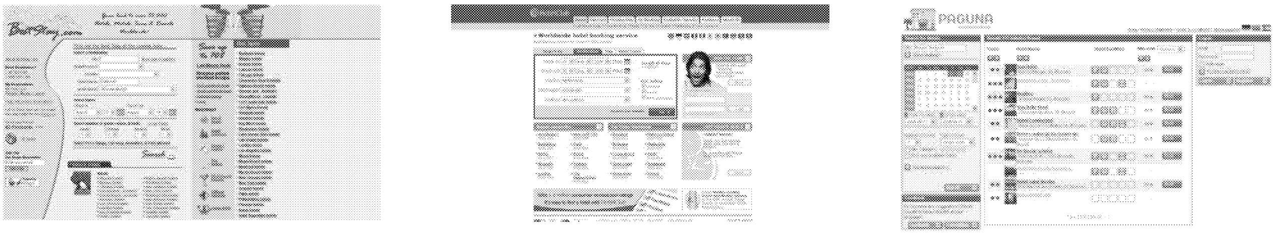


Figure 2: Theoretical Model



Beststay.com Hotelclub.com Paguna.com

Figure 3: Screenshots of stimuli

RESULTS

To evaluate convergent and discriminant validity between the constructs, a confirmatory factor analysis (CFA) was performed. All but one factor loadings exceeded 0.50 on their own construct. Only one cross-construct loading exceeded 0.50. Both items, adapted from Tractinsky's classical dimension of perceived visual aesthetics, were dropped from further analysis. The 10 factors in the CFA solution accounted for 78.5% of the total variation in the data. A close-up of the 10 factor CFA solution displaying the flawed items is showed in table 1.

The reliability of the different constructs is calculated using Cronbach's alpha. Generally, an alpha score greater than 0.6 is considered acceptable for further analysis. However, alpha's exceeding 0.7 are to be preferred. Table 2 shows the alpha scores for all 10 constructs/dimensions used in this research. All variables display acceptable values (>0.80).

Multiple regression analysis was used to test the hypothesized antecedent–consequence relationships in the theoretical model. The intent of this research was to extend TAM by adding different constructs as determinants for perceived ease of use, perceived usefulness and perceived enjoyment. In such research, a stepwise multiple regression analysis technique is recommended examining the contribution of each predictor variables to the regression model. Detailed results of the stepwise regression analysis, and the consequences for the hypotheses, are presented in Table 3.

Eight out of the original 14 hypotheses cannot be rejected on the basis of this empirical data. Some findings are worth mentioning in particular. Behavioural intention to use is explained by a combination of perceived usefulness and perceived enjoyment. Perceived usefulness has the strongest influence on behavioural intention to use (ß=.568), followed by perceived entertainment (ß=.356). These findings support hypotheses 1 and 5. Hypothesis 2 assumed a relationship between perceived ease of use and behavioural intention to use. Results of the regression tests however contradict this hypothesis.

Perceived usefulness is influenced by perceived ease of use and expressive aesthetics. Perceived ease of use has the strongest influence (ß=.717), followed by perceived entertainment (ß=.242). Results show perceived usefulness is not influenced by perceived interactivity.

Perceived enjoyment is determined by classical aesthetics, expressive aesthetics and perceived ease of use. Expressive aesthetics and classical aesthetics have almost equal influence (ß=.365 and ß=.352 respectively). Perceived ease of use has a smaller, but still significant, effect on perceived enjoyment (ß=.218). Perceived interactivity, internet anxiety and internet playfulness do not influence perceived enjoyment.

Perceived ease of use is influenced by classical aesthetics, active control, internet anxiety and synchronicity. Classical aesthetics is by far the strongest determinant of perceived ease of use (ß=.451). Internet anxiety is the next largest predictor (ß=-.183), closely followed by active control (ß=.172) and synchronicity (ß=.166). Internet playfulness did not show to have an effect on perceived enjoyment.

The revised theoretical model is displayed in figure 4. The revised model shows that, except for internet playfulness, all constructs introduced in this research act as determinants of one or more of the original TAM constructs (perceived ease of use and perceived usefulness) and perceived enjoy-

Table 1: Close-up of CFA solution

Item\Factor	ENJ	CAEST	EAEST
CAEST1		.752	
CAEST2		.540	
CAEST3	.497	.435	
CAEST4		.591	
CAEST5		.522	.522
EAEST1			.729
EAEST2			.598
EAEST3			.760
EAEST4			.765
EAEST5			.641

Table 2: Cronbach alpha scores per construct

Construct	Items	Alpha
Perceived Ease of Use	5	.9323
Perceived Usefulness	5	.9429
Behavioural Intention to Use	4	.9648
Perceived Interactivity		
Active Control	4	.8248
Synchronicity	4	.9288
Perceived Enjoyment	4	.9084
Perceived Visual Aesthetics		
Classical Aesthetics	5	.8503
Expressive Aesthetics	5	.8730
Internet Playfulness	5	.8557
Internet Anxiety	4	.8818

Table 3: Results of four regression tests

Dependent variable: Behavioural Intention to Use	R2	R2 change	adjusted R2	ß	Hypothesis
Perceived Usefulness	.622		.619	.568***	H1: supported
Perceived Enjoyment	.700	.078	.695	.356***	H5: supported
Excluded Variables:					
Perceived Ease of Use				.105	H2: rejected
Dependent variable: Perceived Usefulness	**R2**	**R2 change**	**adjusted R2**	**ß**	**Hypothesis**
Perceived Ease of Use	.609		.606	.717***	H3: supported
Expressive Aesthetics	.664	.055	.659	.242***	H9: supported
Excluded Variables:					
Active Control				.036	H6: rejected
Synchronicity				.090	
Dependent variable: Perceived Enjoyment	**R2**	**R2 change**	**adjusted R2**	**ß**	**Hypothesis**
Classical Aesthetics	.409		.405	.352***	H11: supported
Expressive Aesthetics	.502	.093	.495	.365***	
Perceived Ease of Use	.532	.030	.523	.218**	H4: supported
Excluded Variables:					
Active Control				.001	H8: rejected
Synchronicity				.098	
Internet Playfulness				.084	H13: rejected
Internet Anxiety				.043	H15: rejected
Dependent variable: Perceived Ease of Use	**R2**	**R2 change**	**adjusted R2**	**ß**	**Hypothesis**
Classical Aesthetics	.364		.359	.451***	H10: supported
Active Control	.431	.067	.423	.172*	H7: supported
Internet Anxiety	.468	.037	.457	-.183**	H12: supported
Synchronicity	.486	.018	.472	.166*	H7: supported
Excluded Variables:					
Internet Playfulness				.051	H14: rejected

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

ment. The construct of perceived visual attractiveness has the most added value for the model, being a determinant of perceived usefulness, perceived ease of use and perceived enjoyment. The internet anxiety and perceived interactivity constructs add some explanatory power to the model in their relationship with perceived ease of use.

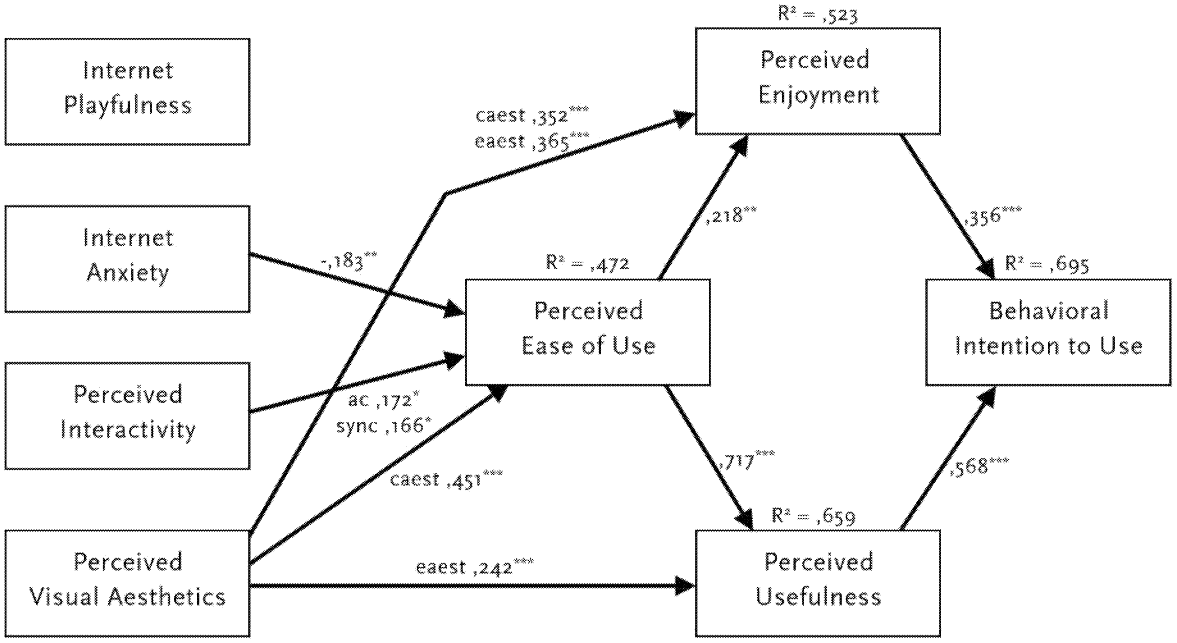
DISCUSSION

In the original TAM, perceived ease of use is considered to be a direct determinant of the behavioural intention to use. The findings in this research suggest that only perceived usefulness and perceived enjoyment are directly connected with behavioural intention to use.

Perceived ease of use was found to have a strong relationship with both perceived enjoyment and perceived usefulness. The originally hypothesized direct effect of perceived ease of use on behavioural intention to use is thus moderated through enjoyment and usefulness.

Van der Heijden found that for hedonic websites, perceived ease of use and perceived enjoyment had the strongest relationship with behavioural intention to use (Heijden 2003). Users of hedonic systems seem to be mostly intrinsically motivated and seem to use the website for no apparent reinforcement other than the enjoyment use per se. The nature of the stimuli used in the context of this research clearly is more utilitarian than hedonic. Users of utilitarian websites can be assumed to be more extrinsically motivated, using the website as means to achieve a goal: in our case to find hotel information and plan a trip.

This suggests that the utilitarian nature of the websites used in this research affected the original relationships in the technology acceptance model. Specifically, perceived ease of use loses its predictive value in favour of usefulness and



*** $p < 0.001$, ** $p < 0.01$. * $p < 0.05$

AC: active control dimension of perceived interactivity; SYNC: synchronicity dimension of perceived interactivity; CAEST: classical aesthetics dimension of perceived visual aesthetics; EAEST: expressive aesthetics dimension of perceived visual aesthetics

Figure 4: Revised Theoretical Model

enjoyment. These findings suggest that intrinsically motivated users of hedonic systems value ease of use over usefulness, while extrinsically motivated users of utilitarian systems value usefulness over ease of use.

In both use contexts, perceived enjoyment has a strong relationship with a user's intention to use. This reinforces all research regarding the important role of enjoyment in the user experience.

This research was among the first to use the two dimensions of perceived visual aesthetics, developed by Lavie and Tractinsky (2004), in a TAM context. The two dimensional scale proved to be valid and reliable for eight out of ten of the original items.

The two dimensions of perceived visual aesthetics were able to explain more than 50 percent of the variance in perceived enjoyment. Furthermore, the classical dimension of perceived visual aesthetics was found to be the most influential determinant of perceived ease of use. Finally, the expressive dimension of perceived visual aesthetics proved to have an influence on perceived usefulness.

Lavie and Tractinsky compared their dimensions with the aesthetic quality of landscape design, order and complexity. "Order may be defined as the degree and kind of lawfulness governing the relations among the parts of an entity … Complexity is the multiplicity of the relationships among the parts of an entity." (Lavie and Tractinsky 2004, p. 288). Good design should be a balance between these two qualities, as "complexity without order produces confusion" and "order without complexity produces boredom" (Lavie and Tractinsky, 2004, p. 288). This theory is perfectly reflected in the findings of this research. The classical dimension is a strong determinant of ease of use, thus implying that orderly design increases usability. Both dimensions proved to be equally influential on the user's perceived enjoyment of a website. These findings suggest that a balance of order and complexity, or classical and expressive aesthetics, should be preferred, especially when recalling the findings regarding the direct influence of enjoyment on behavioural intention to use and the indirect influence of ease of use.

The expressive dimension of visual aesthetics being a weak determinant of perceived usefulness corresponds the discovery of a weak relationship between his construct of perceived visual attractiveness and perceived usefulness (Heijden 2004). This research has further clarified this relationship by showing that the relationship is between the expressive dimension and perceived usefulness. This suggests that perceived creativity and originality in a website's design, qualities that make a site 'stand out', gives a user the impression the site could be more useful.

Perceived interactivity was hypothesized to be of influence on both of the original TAM constructs (perceived ease of use and usefulness) and with perceived enjoyment. Only the hypothesis of a relationship between interactivity and ease of use was supported. This suggests that control over the interaction and responsiveness of the site slightly improves its usability, but does not lead to more enjoyment of more perceived usefulness.

The last two constructs added to this research were related to the character of users, rather than the character of the website. The construct of internet anxiety was added to measure the user's tendency to be uneasy or apprehensive towards current or future use of the internet. Internet playfulness refers to the degree of spontaneity a user exhibits when interacting online. These, or similar, constructs were used in earlier TAM research and relationships with ease of use and enjoyment were reported. This research only found internet anxiety to be a weak determinant of perceived ease of use. This suggests that users who feel more anxious about using the internet perceive websites to be less easy to use.

LIMITATIONS

This research suffers from some limitations. In the first place, the sample used for this research does not match the distribution of the Dutch internet population. The sample is rather young, has a relatively high education and the male-female distributions is skewed. Moreover, the sample appears to have quite a lot of internet experience, using the web for several years and reporting rather lengthy weekly use. This could have caused the sample to be slightly biased toward a high degree of playfulness and low degree of anxiety. Replicating this study with a larger sample might provide a wider range of responses regarding playfulness and anxiety. This research's rejected hypothesis regarding playfulness and anxiety could than be re-examined.

Secondly, there is a survival bias in the sample, because those who were unsatisfied with the website or web survey could have ceased to use it.

Finally, the current research was limited by the use of 'first generation statistics', which can only analyze direct relationships between dependent- and independent variables. Future research should try to profit from 'second generation statistics' using structural equation modelling, which enables researchers to model the relationships among multiple independent and dependent constructs simultaneously. Examples of such techniques are LISREL and PLS.

CONCLUSION

The aim of this paper was to explore the role of qualities beyond usefulness and usability on the web user experience. This study succeeded in this objective by constructing and validating a revised theoretical model of web user experience based on the well-known Technology Acceptance Model. Five new constructs were introduced to the original model, being perceived enjoyment, perceived visual aesthetics, perceived interactivity, internet anxiety and internet playfulness. Four of the new constructs added explanatory power to the model. The model was able to explain 66% of the variance observed on the constructs of perceived usefulness, 47% on perceived ease of use, 52% on perceived enjoyment and 70% behavioural intention to use.

Future research should try to connect these abstract, and subjective, elements of the user experience to more concrete, objectively measurable, features and characteristics of website development and design. The relationship between the construct of 'behavioural intention to use' and tangible usage measures, 'actual use', should be examined.

REFERENCES

Davis, F.D. 1989. "Perceived usefulness, perceived ease of use, and user acceptance of information technology." *MIS Quarterly* 13, No. 3, 318-340.

Davis, F.D.; R.P. Bagozzi; and P.R. Warshaw. 1992. "Extrinsic and intrinsic motivation to use computers in the workplace." *Journal of Applied Social Psychology* 22, No. 14, 1111–1132.

Fishbein, M. and I. Ajzen. 1975. *Belief, attitude, intention and behavior: an introduction to theory and research.* Reading, MA: Addison-Wesley.

Heijden, H.v.d. 2003. "Factors influencing the usage of websites: the case of a generic portal in the Netherlands." *Information and Management* 40, No. 6, 541-549.

Heijden, H.v.d. 2004. "User acceptance of hedonic information systems." *MIS Quarterly* 28, No. 4, 695-704.

Lavie, T. and N. Tractinsky. 2004. "Assessing dimensions of perceived visual aesthetics of web sites." *International Journal of Human-Computer Studies* 60, No. 3, 269-298.

Lindgaard, G. and C. Dudek. 2003. "What is this evasive beast we call user satisfaction?" *Interacting with Computers* 15, No. 3, 429-452.

Liu, Y. 2003. "Developing a scale to measure the interactivity of websites." *Journal of Advertising Research* 43, No. 2, 207-218.

Pham, M.T.; J.B. Cohen; J.W. Pracejus; and G.D. Hughes. 2001. "Affect monitoring and the primacy of feelings in judgment." *Journal of Consumer Research* 28, No. 2, 167-188.

Reeves, B. and C. Nass. 1996. *The Media Equation*. CSLI Publications.

BIOGRAPHY

TEUN HOMPE holds a master's degree in Media and Knowledge Engineering of Delft University of Technology. He currently works as a usability consultant at valsplat. In this position he has carried out usability studies for numerous clients and gained experience testing utilitarian websites for airlines, insurance companies and online retailers.

JORIS LEKER is a founding principal of valsplat and has extensive experience in applied usability research of both utilitarian and hedonic websites for commercial clients.

CHARLES VAN DER MAST has a PHD in Computer Science from Delft University of Technology where he is Associate Professor at the Man-Machine Interaction group at the Department of Electronic Engineering, Mathematics and Computer Science. He teaches courses on Multimodal Interfaces and Virtual Reality, Developing Highly Interactive Systems, Multimedia, Educational Software and Intelligent User Experience Engineering. He co-developed the Bachelor/Master curriculum in Media & Knowledge Engineering. His interests include using various media to improve teaching, VR therapy for phobia treatment and agent support in complex systems.

MARK NEERINCX is head of the Intelligent Interface group at TNO Human Factors, and professor in Man-Machine Interaction at the Delft University of Technology. He has extensive experience in applied and fundamental research. Important results are (1) a cognitive task load model for task allocation and adaptive interfaces, (2) models of human-machine partnership for attuning assistance to the individual user and momentary usage context, (3) cognitive engineering methods and tools, and (4) a diverse set of usability "best practices". He has been involved in the organisation of conferences, workshops and tutorials to discuss and disseminate human factors knowledge.

VU @ SECOND LIFE[*]– CREATING A (VIRTUAL) COMMUNITY OF LEARNERS

Anton Eliëns
FEW
VU University
Amsterdam
eliens@cs.vu.nl

Frans Feldberg
FEWEB
VU University
Amsterdam
jfeldberg@feweb.vu.nl

Elly Konijn
FSW
VU University
Amsterdam
ea.konijn@fsw.vu.nl

Egon Compter
Communicatie
VU University
Amsterdam
e.compter@dienst.vu.nl

KEYWORDS
virtual worlds, community of learners, Second Life

ABSTRACT

In this paper we report on our experiences in creating presence for our university in the Second Life environment. After a brief explanation of our motivation(s), we will describe our approach, which resulted in creating a virtual campus acting both as a portal for information, and, more importantly, as a meeting point, offering the opportunity to create a virtual community of learners, in line with the overall educational policy of our university. We will discuss the merits of Second Life as an educational platform, and indicate relevant research perspectives. To illustrate how the virtual meets the real, an impression will be given of our encounters with the press.

INTRODUCTION

Online virtual worlds have been present for more than 10 years, AlphaWorld[1], for example, was introduced in 1995. However, the recent substantial media attentention for Second Life can be considered as an indication that virtual worlds are no longer the domain of a selective group of fanatic online gamers. For example, the number of registered residents in Second Life increased from 1,8 million at the beginning of December 2006 to over 4 million within a period of less than 3 months. Big companies like Reebok, IBM, Philips, and ABN AMRO organize press meetings to announce their presence in virtual worlds. Even governments, municipalities, and NGOs enter Second Life with an eagerness that is comparable to the *don't miss the boat* feeling recognized at the early days of the internet. Second Life has even been presented as hype. On February 28th 2007, the Vrije Universiteit Amsterdam (in English, our official name is *VU University Amsterdam*) announced its presence in Second Life as the first Dutch university. National and international companies are eager to have their regional headquarters in Amsterdam. The international reputation of Amsterdam with respect to its tolerance for sex and soft drugs has apparently been no hindrance to that. However, when we announced our presence in Second Life as the first Dutch university, news items appeared, in Elsevier[2] among others, which mentioned the senate's (Tweede Kamer) concern with possible irregularities in Second Life immediately after announcing our university's presence in Second Life.

Why does a respectable university, like ours, want to be present in Second Life? And what are the prospects or benefits for an educational institute with a strong research reputation to be present in Second Life? Is it publicity we are after, the momentary attention of the press, taking profit of the (current) hype around Second Life, or are there more sustainable reasons that make such presence worthwhile, from both educational and research perspectives. In the following, we will address these questions, and give an account of the process that led to our presence in Second Life.

The structure of this paper is as follows. First, we explain our motivation(s), and then we will outline the actual building of our virtual campus. We will discuss the potential of Second Life as an educational platform, and after that we will indicate relevant research perspectives. Then we will give a comparative technical overview, and ponder on why Second Life is so successful. Finally, after briefly reporting on our experiences when going live, and some speculative thoughts about future developments, we will present our conclusions.

CREATING PRESENCE IN A PARTICIPATORY CULTURE

In less than a decade after the publication of William Gibson's novel *Neuromancer*, the *metaverse* was realized, albeit in a primitive way, through the introduction of VRML[3], introduced at the Int. Web Conference of 1992. Cf. Anders (1999). The German company *blaxxun*[4], named after the virtual environment in Neil Stephenson's *Snowcrash*, was one of the first to offer

[*] www.vu.nl/secondlife
[1] www.activeworlds.com/worlds/alphaworld
[2] www.elsevier.nl/nieuws/laatste_24_uur/artikel/asp/artnr/140574
[3] www.web3d.org
[4] www.blaxxun.com

a 3D community platform, soon to be followed by *AlphaWorld*, already mentioned in the introduction, which offered a more rich repertoire of avatar gestures as well as limited in-game building facilities. However, somehow 3D virtual communities never seemed to realize their initial promises. Furthermore the adoption of VRML as a 3D interface to the Web never really took off.

The history of Second Life is extensively descibed in the official Second Life guide, Rymaszweski et al. (2007). Beginning 2004, almost out of the blue, *Second Life*[5] appeared with a high adoption and low churn rate, now counting, March 2007, over 4 million inhabitants. Considering the cost of ownership of land, which easily amounts to 200 euro per month rent after an initial investment of 1500 euro for a single piece of land measuring 65,536 square meters, the adoption of Second Life by individuals as well as companies such as ABN-AMRO, Philips and institutions such as Harvard is surprising.

What is the secret of the success of Second Life? We don't know! But in comparison to other platforms for immersive worlds, including MMORPGs such as *World of Warcraft*[6] and *Everquest*[7], Second Life seems to offer an optimal combination of avatar modification options, gesture animations, in-game construction tools, and facilities for communication and social networking, such as chatting and instant messaging. Cf. Utz (2003). Incorporating elements of community formation, commonly denoted as Web 2.0, and exemplified in MySpace, YouTube and Flickr, the immersive appearance, perhaps also the built-in physics and the inclusion of elementary economic principles, seem to be the prime distinguishing factors responsible for the success of Second Life. In addition, the possibility of recording collaborative enacted stories, Davenport (2000), using built-in *machinima*[8] certainly also contributes to its appeal. Later on, after discussing Second Life from a more technical perspective, we will speculate further on the possible reasons for the success and adoption of Second Life as a platform for communication and immersive presence.

What has been characterized as a shift of culture, from a media consumer culture to a participatory culture, Jenkins (2006), where users also actively contribute content, is for our institution one of the decisive reasons to create a presence in Second Life, to build a virtual platform that may embody our so-called *community of learners*, where both staff and students cooperate in contributing content, content related to our sciences, that is.

BUILDING A VIRTUAL CAMPUS

In December 2006, we discussed the idea of creating presence in Second Life. Our initial targets were to build a first prototype, to explore content creation in Second Life, to create tutorials for further content creation, and to analyze technical requirements and opportunities for deployment in education and research.

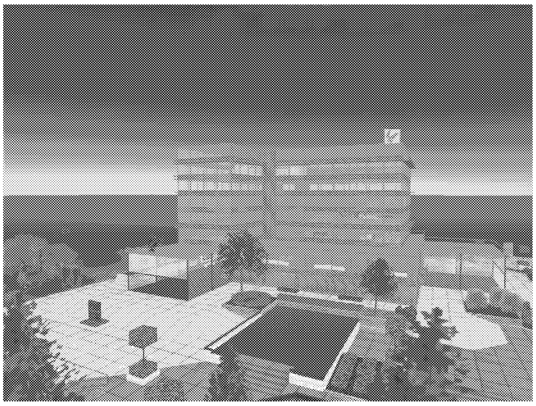


Fig 1. VU Campus – outside view

Two and a half months later, we are online, with a virtual campus, that contains a lecture room, a telehub from which teleports are possible to other places in the building, billboards containing snapshots of our university's website from which the visitors can access the actual website, as well as a botanical garden mimicking the VU Hortus, and even a white-walled experimentation room suggesting a 'real' scientific laboratory. All building and scripting were done by a group of four students, from all faculties involved, with a weekly walkthrough in our 'builders-meeting' to re-assess our goals and solve technical and design issues.

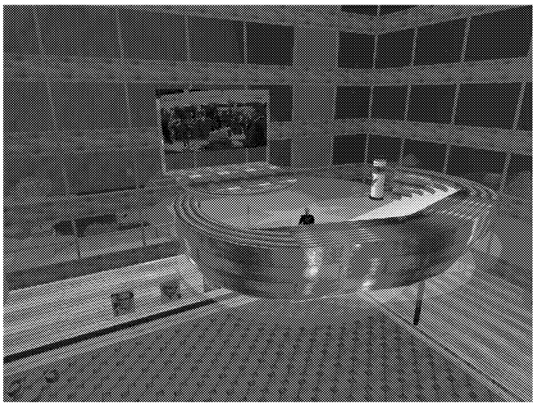


Fig 2. VU Campus – inside view

The overall style is realistic, although not in all detail. Most important was to create a visual impression of resemblance and to offer the opportunity to present relevant infomation in easily accessible, yet immersive, ways. Cf. Bolter & Grusin (2000), Hoorn et al. (2003).

[5] secondlife.com

[6] www.worldofwarcraft.com

[7] everquest.station.sony.com

[8] www.machinima.org

Our virtual campus, see figs. 1 and 2, is meant to serve as an *information portal* and as a *meeting ground*, where students, staff and visitors can meet and communicate, as well as a place were teachers and researchers can conduct experiments aimed at discovering new ways of teaching and doing research.

SECOND LIFE AS AN EDUCATIONAL PLATFORM

The first idea that comes to mind, naturally, is to use Second Life to offer courses online. But, although we do have plans to give lectures (college) on law, probably including the enactment of a particular case, we do consider this approach as rather naive, and frankly we see no reason to include what may be considered an outdated paradigm of learning in our virtual campus, where there might be more appealing alternatives. Similarly, using the virtual laboratory for experiments might not be the best way to offer courses, although, again, we do intend to provide a model of a living cell, allowing students to study the structure, functionality and behavior of organic cells in virtual space.

Considering the success of our multi-disciplinary building team, it seems more worthwhile to take the cooperative effort of building as a model, and switch to a paradigm of learning in which in-game exploration and building plays an important role. It is no secret that many students enjoy gaming, and although some might think that *gaming is a waste of time*, many authors, including Gee (2003) and Vorderer & Bryant (2006), seem to think that gaming and game-related efforts provide a form of *active learning*, allowing the gamer to experience the world(s) in a new way, to form new affiliations, and to prepare for future learning in similar or even new domains.

More importantly, due to intense involvement and the need to analyze game challenges, according to Gee (2003), gaming even encourages *critical learning*, that is to think about the domain in a meta-level as a complex system of inter-related parts, and the conventions that govern a particular domain, which Gee (2003) characterizes as *situated cognition in a semiotic domain.* Without further explanation, we may note here that *semiotic domain* means a *world of meaning* that is due to social conventions and patterns of communication. Cf. Kress & Van Leeuwen (1996).

Observing that both creativity and communication are vital elements of higher education, we envisage to deploy Second Life for a multi-disciplinary honors-track course that will focus on the communication of scientific research, for example the impact of climate change and the various ways we can mitigate or adapt to the potential threats of global warming. In this way we can also contribute to the issue of *media literacy*, or "*mediawijsheid*[9]" as the Dutch Council of Culture calls it, that is making students aware of the impact of the media in presenting controversial issues. In this respect we strongly believe that Second Life does not necessarily lead to another screen-addiction giving access to dubious content, but that it can actually be deployed in a constructive way as an opportunity to stimulate and support active learning.

RESEARCH PERSPECTIVES – VIRTUAL VERSUS REAL

Is decision-making in a virtual environment the same as or similar to decision-making in the real world? And what about investments, and starting a new company? The Second Life economy, powered by Linden dollars and governed by the Lindex-exchange, provides an interesting platform to study decision-making behaviors, for example with a group of students in a course about decision-support systems.

Another way to establish a relation with reality is to provide a *virtual context* to objects existing in actual reality, such as cultural heritage, and for example relate paintings to the world they depict, which must necessarily be re-constructed in a virtual environment as it no longer exists, Rutledge et al. (2000).

In previous work, we did study the construction and deployment of humanoid intelligent agents, Eliens et al (2006), and we looked at ways such agents could provide an explanation in rich media contexts, Eliens et al. (2003), or guidance in finding locations in large virtual worlds, Ballegooij & Eliens (2001). Also did we explore whether virtual replicas of existing buildings, in our case museums, was the best way to provide immersive access to art-related information, Eliens et al. (2007), and actually we concluded that it was not! In one of such virtual replicas, in this case the atelier of the Dutch artist Marinus Boezem, we studied the effectiveness of the use of an intelligent humanoid agent, and we found interesting relationships between the appearance (looks) of the agent, and the trustworthiness of its advice, Hoorn et al. (2004), Van Vugt et al. (2006a). We extended our research efforts into appearances of virtual humans and their effectiveness in virtual worlds like the Sims, Van Vugt et al. (2006b). Furthermore, we studied differences between perceptions of fictitious (i.e. Hollywood) characters versus existing (i.e. real world) characters, Konijn & Bushman (2007). Finally, we examined the role of emotions in establishing effective communication between real and virtual humans, Konijn & Van Vugt (2007).

However, apart from studying patterns of communication, and the way appearance and identity may influence communication (e.g. Konijn & Nije Bijvank (2007)), it

[9] www.cultuur.nl/nieuws.html?nieuws_speeches.php?id=184

seems at this stage more interesting to explore how to enhance communication in a shared virtual world by actually deploying virtual objects, instead of relying on chatting and textual information, and to design tasks that require cooperation in an essential manner. More generally, we would like to deploy Second Life as a platform for *serious games*[10], such as service management games, Eliens & Chang (2007), and we believe that for corporate institutions this might well be the real benefit Second Life has to offer!

Taking, however, a more critical look at Second Life as a platform for serious games, it might appear to be lacking in a number of respects, including (not the least important) security, programmability and robustness. As the failure of many of the early CSCW (Computer Supported Cooperative Work) applications indicates, cf. Churchill et al. (2001), to provide adequate support for collaboration is not easy, since a manifold of issues have to be resolved, such as turn-taking, gaze detection, etcetera. And in addition, for tasks that require strict timing, such as musical improvisation, Eliens *et al.* (1997), synchronization and time-lag have to be taken into account.

Taking these issues into account, we may wonder whether we should adopt Second Life, or rather seek refuge with an open source game engine such as Delta3D[11], or a commercial game engine such as offered by the Steam-powered Half Life 2 SDK[12], cf. Eliens & Bhikharie (2006), which might be more compliant with the extensions required to provide adequate support for serious cooperative games. Interestingly, the Second Life client has recently been given out to open source, and that would allow for many client-side hacks, such as for example multi-modal interaction[13], which in combination with the server-side scripting capabilities may result in powerful extensions.

At this stage, though, it might well be the level of adoption that is decisive in the choice of Second Life as a platform for serious corporate games!

COMPARATIVE TECHNICAL OVERVIEW

From a technical perspective, Second Life offers an advanced game engine that visitors and builders use (implicitly) in their activities. Before discussing how Second Life compares to (a selection of) other game engines and virtual environment frameworks, it is worthwhile to look at an overview of the main functional components of a *game engine*, which according to Sherrod (2006) encompass:

- rendering system – 2D/3D graphics
- input system – user interaction
- sound system – ambient and re-active
- physics system – for the blockbusters
- animation system – motion of objects and characters
- artificial intelligence system – for real challenge(s)

Although it is possible to build one's own game engine using OpenGL or DirectX, or the XNA[14] framework built on top of (managed) DirectX, in most cases it is more profitable to use an existing game engine or 3D environment framework, since it provides the developer with a load of already built-in functionality. In the following table, we give a brief comparative technical overview of, respectively, the Blaxxun Community Server (BlC), AlphaWorld (AW), the open source Delta3D engine (Δ3D), the Half Life 2 Source SDK (HL2), and Second Life (SL).

	BlC	AW	Δ3D	HL2	SL
in-game building	-	+	+/-	-	++
avatar manipulation	+	++	+/-	+	++
artifical intelligence	+	-	+/-	+	-
server-side scripts	+	-	+/-	+	++
client-side scripts	++	-	+/-	+	-
extensibility	+	-	++	+	+/-
open source	-	-	++	-	+/-
open standards	-	-	+/-	-	+/-
interaction	+/-	+/-	++	++	+/-
graphics quality	+/-	+/-	++	++	+
built-in physics	-	-	+	++	+
object collision	-	-	++	++	+
content tool support	+/-	-	++	+	-

Obviously, open source engines allow for optimal extensibility, and in this respect the open source version of the SL client may offer many opportunities. Strong points of SL appear to be *in-game building*, *avatar manipulation*, and in comparison with BlC and AW *built-in physics* and *object collision detection*. Weak points appear to be *content development tool support*, and especially in comparison with Δ3D and HL2 *interaction*. For most types of action-game like interaction SL is simply too slow. This even holds for script-driven animations, as we will discuss in the next section. In comparison with a game as for example Age of Empires III[15], which offers in-game building and collaboration, Second Life distinguishes itself by providing a 3D immersive physics-driven environment, like the 'real' game engines.

SCRIPTING IN SECOND LIFE

Second Life offers an advanced scripting language with a C-like syntax and an extensive library of built-in functionality. Although is has support for objects, LSL

[10] games.uscannenberg.org/AWGHome.php

[11] www.delta3d.org

[12] half-life2.com

[13] www.hackdiary.com/archives/000101.html

[14] crosoft.com/directx/XNA

[15] www.ageofempires3.com

(the Linden Scripting Language) is not object-oriented. Cf. Eliens (2000). Scripts in Second Life are server-based, that is all scripts are executed at the server, to allow sharing between visitors. Characteristic for LSL are the notions of *state* and *eventhandler*, which react to events in the environments. As an example of perhaps the most simple script to be found, taken from the online tutorial of CTER[16], look at:

```
default {
  state_entry() {
  llSetText("Do you want to learn scripts?",
            <255,255,255>,5);
  }
}
```

When attached to an object, triggering *state_entry* (in the *default* state), results in displaying the text *"Do you want to learn scripts?"*.

LSL offers a range of built-in types, including *int*, *float*, *list*, and even *vector* and *rotation* (which is a 4-place vector). It provides the standard operators, as well as the usual blocks and scopes. Scripts are attached to objects and must be explicitly activated, for example by right clicking on the object and selecting, for example, the option *teleport*, as in the script below, which may be used for teleporting visitors' avatars:

```
vector target= <162,134,27>; // coordinates
default {
state_entry() {
   llSetText("Info @ VU",<255,255,255>,5);
   llSetSitText("teleport");
   rotation my_rot=llGetRot();
   llSitTarget((target - llGetPos()) /
       my_rot,ZERO_ROTATION / my_rot);
   }

changed(integer change) {
   llUnSit(llAvatarOnSitTarget());
   }
} // end default
```

Selecting the *teleport* option actually results in creating an invisible object on which the avatar *sits*. The object is then transported to the *target* location in about 0.2 seconds. The 0.2 second interval does also apply for other actions, for example rotations to objects, which gives an awkward visual impression, simply because it is too slow. For teleports, however, the 0.2 second interval does suffice.

Among the built-in functions there are functions to connect to a (web) server, and obtain a response, in particular (with reference to their wiki page):

- request – wiki.secondlife.com/wiki/LlHTTPRequest
- escape – wiki.secondlife.com/wiki/LlEscapeURL
- response – wiki.secondlife.com/wiki/Http_response

Other functions to connect to the world include *sensors*, for example to detect the presence of (visitors') avatars, and chat and instant messaging functions to communicate with other avatars using scripts. In addition, LSL offers functions to control the behavior and appearance of objects, including functions to make objects react to physical laws, to apply force to objects, to activate objects attached to an avatar (for example phantom Mario sprites, see section *hold your breath*), and functions to animate textures, that can be used to present slide shows in Second Life.

ADMINISTRATION AND SUPPORT

When building our virtual campus we did experience in practice how difficult it is to manage properties like ownership, access and modifiability rights, and when going live these issues became even more urgent, since malicious visitors may profit from any administrative negligence.

As a reference, we list some of the resources available for developers, which are organized as wiki's, and at the moment of writing still in flux, that is incomplete, but growing:

wiki(s)

- knowledgebase – secondlife.com/knowledgebase
- scripting – wiki.secondlife.com/wiki/LSL_Portal
- main page – https://wiki.secondlife.com/wiki

All in all, administration in Second Life is intricate and in our experience not entirely bug-free. So far we have not understood all the ins and outs of property management and security in Second Life.

Additionally, there are resources that may give developers an idea[17] which direction to take, educators hints[18] on how to set up a course, and more general resources providing building tutorials[19] and an insight[20] in the history of Second Life, explaining among others the growth of the Second Life virtual economy.

A convenient, and to make your world accessible perhaps essential feature is the so-called *slurl*, that allows for access to your Second Life property from a web page. As an example, the *slurl* connecting to the *VU University NL* virtual campus is:

slurl.com/secondlife/VU%20University%20NL/29/151

[16] cterport.ed.uiuc.edu/technologies_folder/SL

[17] www.secondlifeinsider.com

[18] www.simteach.com/wiki

[19] trumpy.cs.elon.edu/metaverse/wiki

[20] www.slhistory.org/index.php/Main_Page

SECOND LIFE SUCCESS FACTORS

Will Second Life become the new (immersive 3D) mass medium of our participatory culture of the 21th century, as once the immersive *panorama* was the propaganda/art medium for the masses in the 19th century? Cf. Grau (2003). In thinking about possible reasons why Second Life is so successful, we observed that Second Life does provide:

- convergence of social networking and content creation
- immersive netweorked 3D environment
- inclusion of elementary economic principles

However, we also see that other factors may contribute to the success of Second Life, such as:

- don't miss the boat effect
- free and easy accessible 3D design tool set
- adoption by big companies like IBM, Reebok, ...
- marketing of Second Life by Linden Lab (?)
- the promise to make (real) money (?)

According to Philip Rosedale, CEO of Linden Lab, (interview in .NET magazine, issue 158, January 2007) the success of SL is due to the fact that (1) it offers a set of capabilities, which are in many different ways superior to the real world, (2) the decision to allow residents to own the intellectual property rights to their creations and (3) because Second Life is full of creative possibilities, and opportunites for innovation.

In order to establish what constitutes the success of Second Life in a more rigorous manner, we must subject Second Life to a (game) *reference model* as introduced in Juul (2005), which we have also applied to (serious) service management games in Eliens & Chang (2007). A first tentative characterization of Second Life according to our reference model would be:

reference model

- *rules* – construct and communicate!
- *outcome* – a second world
- *value* – virtual and real (monetary) rewards
- *effort* – requires elementary skills
- *attachment* – a virtual identity
- *consequences* – transfer to first life

Second Life clearly has a wider scope and more freedom than just gaming. Apart from elementary rules, that more or less require of the (serious) visitor to *construct and communicate*, there are almost no fixed rules, no in-game strategies, but many opportunities for interpersonal contact and the establishment of relations world-wide, affecting (possibly) the Second Lifer's first life (*consequences*).

Whether Second Life will turn out to be a veritable media-supported augmentation of our first life, cf. Zielinski (2006), remains to be seen. Chances are also that Second Life will end up as another item on the *dead media projects*[21] list, to be replaced by an alternative participatory framework or environment.

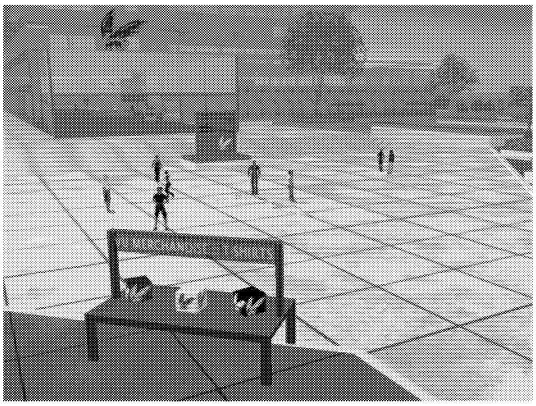


Fig 3. VU @ SL – visitors outside

HOLD YOUR BREATH – GOING LIVE

The 1st of March 2007, we went live. In the evening there was a news item on national televison, RTL4 news, featuring the students showing the virtual campus and our project leader explaining the reasoning behind our presence in Second Life and how to give a course in the virtual classroom. A similar item appeared at AT5, local Amsterdam television, and various newspapers, among which Parool, Telegraaf and Volkskrant, spent a multiple-column article to report on our efforts. As a note, not surprisingly, all items focused on what we have characterized as the naive interpretation of our efforts, exemplifying the old credo *the medium is the message*. To be clear, our intention is not to provide a virtual replica, nor to provide an analogon of the Open University, in Second Life.

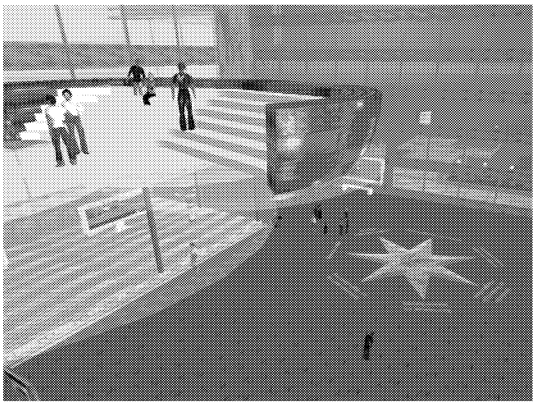


Fig 4. VU @ SL – visitors inside

After the news broadcasts, the number of visitors increased dramatically, having stayed at a modest below

[21] www.cs.vu.nl/∼eliens/media/project-deadmedia.html

100 during the day, see figs. 3 and 4. In the evening, however, just after the news items on the national television, the number of visitors increased rapidly. Since, presently, we do have only one island it appeared to be very difficult to separate internal experimental activities from visitors just asking for additional information, and to exclude potentially malicious visitors. In that evening, we were even surprised by the invasion of an army of Mario Brothers. Hilarious and non-harmful. But enough reason to sit back and limit access to our campus for students and staff only the day after our open day. A few days later, after the first turbulent days following the TV broadcasts, we re-opened our virtual campus to allow visitors to walk/fly around, and enjoy our news items and informative videos. So far, the results exceeded our expectations, the students were praised for the results of their building efforts, and as a team we may continue to think about how to deploy Second Life as a platform for education and research projects.

FUTURE DEVELOPMENT(S)

Virtual or not, ecomomy plays a crucial role in the (past and) future of Second Life, since (www.openthefuture.com):

> ... the internal economy was predicated on the notion that designers could produce in-game objects that they could then sell.

However, the ability to copy mechanically might easily destroy such an economy. In general, it might be questioned whether the (real) economic model of Second Life will hold, or whether an alternative approach which is free from immediate economic constraints, similar to *open source*, will prevail.
In our own educational and research projects we will strive for making Second Life available as a platform for *mediating social awareness*, cf. Vyas et al. (2007) and Vyas et al. (2007b), and actual collaboration, in particular in our university-wide media institute *CAMeRA*, that will coordinate among others our activities in serious game development. Looking what is going on in Second Life, on a global scale, we refer without further comments to the following resources:

- NOAA: 3D weather data visualization[22]
- NOOA: test the water in a virtual world[23]
- CDC: spare change in secondlife[24]
- APPLE: be anyone, set your own agenda[25]
- MMORPG: secondlife as a game[26]

CONCLUSIONS

In this paper we have reported on our experiences in building a virtual campus, giving our university presence in Second Life, and we have delineated the prospects of Second Life as a platform for education and research, embodying our university's credo: to be a *community of learners*. After enjoying our *15 minutes of fame*, however, we need to reflect on what technical requirements must be met to deploy Second Life effectively as a platform for education and research, and, perhaps more importantly, what paradigm of learning to adopt to have real benefit of the potential of Second Life.

ACKNOWLEDGEMENT(S) We thank the students involved, Viola van Alphen (FEWEB), Tom Bestebreurtje (FEW), Elbert-Jan Hennipman (FEW), and last but not least Bjorn de Boer (FSW), for their enthusiasm, creativity, and of course their hard work which led to an amazing result.

REFERENCES

Anders P. (1999), *Envisioning Cyberspace – Designing 3D Electronic Spaces*, McGraw-Hill

Ballegooij A. van and Eliens A. (2001), *Navigation by Query in Virtual Worlds*, In: *Proc. Web3D 2001 Conference*, Paderborn, Germany, 19-22 Feb 2001

Bolter J.D and Grusin R. (2000), *Remediation – Understanding New Media*, MIT Press

Churchill E.F., Snowdon D.N. and Munro A.J., eds, (2001). *Collaborative Virtual Environments – Digital Places and Spaces for Interaction*, Springer

Davenport G.(2000), *Your own virtual story world*, Scientific American, november 2000, pp. 61-64

Eliens A., Welie M., van Ossenbruggen J., Schonhage S.P.C (1997). *Jamming (on) the Web*, In: *Proc. of the 6th Int. World Wide Web Conference — Everone, Everything Connected*, OŔeilly and Associates, Inc. April 1997, pp. 419-426

Eliens A. (2000), *Principles of Object-Oriented Software Development*, Addison-Wesley Longman, 2nd edn.

Eliens A., Huang Z., and Visser C. (2002), *A platform for Embodied Conversational Agents based on Distributed Logic Programming*, In: *Proc. AAMAS 02 Workshop – Embodied conversational agents - letś specify and evaluate them!*, Bologna 17/7/2002

Eliens A., Dormann C., Huang Z. and Visser C. (2003), *A framework for mixed media – emotive dialogs, rich media and virtual environments*, In: *Proc.*

[22] www.secondlifeinsider.com/2006/10/28/3d-weather-data-visualization-in-second-life

[23] www.gcn.com/print/26_04/43147-1.html

[24] www.social-marketing.com/blog/2006/11/cdcs-second-life.html

[25] www.apple.com/games/articles/2005/07/secondlife/

[26] www.mmorpg.com/gamelist.cfm?gameID=83&bhcp=1

TIDSE03, 1st Int. Conf. on Technologies for Interactive Digital Storytelling and Entertainment, Göobel S. Braun N.,n Spierling U., Dechau J. and Diener H. (eds¿), Fraunhofer IRB Verlag, Darmstadt Germany, March 24-26, 2003

Eliens A., Huang Z., Hoorn J.F. and Visser C.T. (2006), *ECA Perspectives - Requirements, Applications, Technology*, In: Z. Ruttkay, E. Andre, W.L. Johnson and C. Pelachaud (eds), *Evaluating Embodied Conversational Agents*, Dagstuhl Seminar Proceedings (04121)

Eliens A. and S.V. Bhikharie (2006), *Game @ VU – developing a masterclass for high-school students using the Half-life 2 SDK*, In: *Proc. GAME'ON-NA'2006*, Sept. 19-21, 2006 - Naval Postgraduate School, Monterey, USA

Eliens A., Wang Y., van Riel C., Scholte T. (2007), *3D Digital Dossiers – a new way to present cultural heritage on the web*, accpeted for the *Int. Web3D Symposium 07*, 15-18 april 2007, Perugia, Italy

Eliens A. and Chang T. (2007), *Let's be serious – ICT is not a (simple) game*, accepted for *FUBUTEC 2007*, April 2007, Delft

Gee J.P. (2003), *What video games have to teach us about learning and literacy*, Palgrave Macmillan

Grau O. (2003), *Virtual Art – From Illusion to Immersion*, The MIT Press

Hoorn J.F., Konijn E.A., Van der Veer G.C. (2003), *Virtual reality: Do not augment realism, augment relevance* , In: *Human-Computer Interaction: Overcoming Barriers*, 4:1, pp. 18-26

Hoorn J., Eliens A., Huang Z., van Vugt H.C., Konijn, E.A., Visser C.T. (2004). *Agents with character: Evaluation of empathic agents in digital dossiers*, Emphatic Agents, AAMAS 2004 New York 19 July - 23 July, 2004

Jenkins H. (2006), *Confronting the Challenges of Participatory Culture: Media Education for the 21th Century*, White Paper, MIT MediaLab

Juul J. (2005), *Half Real – Video Games between Real Rules and Fictional Worlds*, MIT Press

Konijn, E.A. and Bushman, B.J. (2007), *World leaders as movie characters? Perceptions of G. W. Bush, T. Blair, O. Bin Laden, and S. Hussein at the eve of Gulf War II,* , Media Psychology, 9 (1), pp. 157-177

Konijn E.A. and Nije Bijvank M. (2007), *How to become a tough guy? Identity construction through video game play*, In: *Annenberg Workshop on Games for Learning, Development &mp; Change*, Los Angeles, CA, USA

Konijn, E.A. and Van Vugt, H.C. (2007), *Emotions in Mediated Interpersonal Communication: Toward modeling emotion in virtual humans*, In: *Mediated Interpersonal Communicationa*, Konijn, E. A., Tanis, M., Utz, S., Barnes, S. (eds.), Mahwah, NJ.: Lawrence Erlbaum Associates

Kress G. and Van Leeuwen T. (1996), *Reading Images: The Grammar of Visual Design*, Routledge

Rymaszewski M., Au W.J., Wallace M., Winters C., Ondrejka C., Batstone-Cunningham B. (2007). *Second Life – the official guide*, Wiley

Rutledge L., van Ballegooij A., Eliens A. (2000), *Virtual Context - relating paintings to their subject*, Culture Track of WWW9 in Amsterdam, The Netherlands, May 16th, 2000

Sherrod A. (2006), *Ultimate Game Programming with DirectX*, Charles River Media

Utz S. (2003), *Social identification and interpersonal attraction in MUDs*, Swiss Journal of Psychology, 62, 91-101.

Vorderer, P. and Bryant, J. (eds.). (2006), *Playing computer games - Motives, responses, and consequences*, Mahwah, NJ: Lawrence Erlbaum Associate *Playing Video Games*,

Van Vugt, H.C., Konijn, E.A., Hoorn, J.F., Keur, I., Eliens, A. (2006). *Realism is not all! User Engagement with Task-Related Interface Characters*, Interacting with Computers, 2006

Van Vugt, H. C., Hoorn, J. F., Konijn, E. A., de Bie Dimitriadou, A. (2006). *Affective affordances: Improving interface character engagement through Interaction*, International Journal of Human-Computer Studies, 64 (9), 874–888

Vyas D., van de Watering M., Eliens A., van der Veer G. (2007), *Engineering Social Awareness in Work Environments*, accepted for *HCI International 2007*, 22-27 July, Beijing, China

Vyas D. van de Watering M., Eliens A., van der Veer G. (2007b), *Being Social @ Work: Designing for Playfully Mediated Social Awareness in Work*, accepted for *HOIT 2007*, Chennai, India in August 2007

Zielinski S. (2006), *Deep Time of the Media – Towards an archaeology of Hearing and Seeing by Technical Means*, The MIT Press

Intelligent Advertisement for E-Commerce

Prof. Dr.-Ing. Stephan Kassel
Prof. Dr.-Ing. Christian-Andreas Schumann
Dipl.-Inf. Claudia Tittmann
Westsächsische Hochschule Zwickau (FH) – University of Applied Sciences
P.O. Box 201037
D - 08012 Zwickau
E-mail: Stephan.Kassel@fh-zwickau.de

KEYWORDS
Customer Knowledge Management, CRM, Knowledge Management, Data Warehouse, Expert Systems, e-Commerce, Marketing.

ABSTRACT

E-Commerce remains a driving force of business for the next years. But online sales channels typically lack the effects of customer loyalty which is an important success factor for retailers, lowering their overall marketing costs. Customer loyalty can be achieved only by knowledge about the customers behaviour and preferences. These have to be respected in the communication with the customer. Mass mailings and un-specific advertisings on the retailers' web sites are lowering loyalty, whereas specific messages of value for the customer are raising his loyalty. Up to date, knowledge about customers is often limited to the direct sales transactions.

In our approach, a new way of acquiring knowledge about customers is suggested, which allows a deeper insight on his preferences. This is done by coupling the already present information about customer behaviour with external data describing events which can influence this behaviour. Finding correlations between those different kind of data allow a deeper model of the customer which can be utilized for doing more specific marketing actions.

MARKET TRENDS IN E-BUSINESS

In the middle of the 1990s, the concept of E-Business emerged. It was quickly adopted by various companies, leading to an enormous euphoria. After the first crashes of the so-called dot.com companies in 2000, there came a sudden disillusion, and hundreds of e-commerce projects were stopped immediately, leading to a worse development. It was recognized, that there was a lack of sound strategies for establishing a permanent market share and obtaining long term customer loyalty in most of the companies acting on the online markets. In contrary, the enterprises concentrated on acquisition of new customers without sustainable efforts on customer binding. It was like a gold rush, but only few players on the world-wide market survived.

Meanwhile, the second generation of E-Business started: The aim of the companies is once again extraordinary growth, but now this goal is supported by customer loyalty strategies. Companies came to the conclusion that customer loyalty is a key factor for success, a rule that holds in e-business in the same way as in stationary business.

Current studies of German and European market satisfactorily show the existence of success. For example, Europe's E-Commerce forecast shows a rise from 103 billion Euros in 2006 to 263 billion Euros in 2011 (Favier 2006)

In Germany, E-Commerce business is permanently growing. In 2005, 90 percent of the overall business volume was realised in transactions between companies (business-to-business). But even the remaining ten percent of online-retail with private customers (business-to-consumers) increased by 43 percent up to 32 billion Euros. This trend is seen to continue in the next years. For 2009, there is a prognosis of 114 billion Euros for B2C transactions (Bitkom, 2007).
The number of online customers is also increasing. In Western Europe, there are meanwhile 50% of the population online, as shown in figure 1.

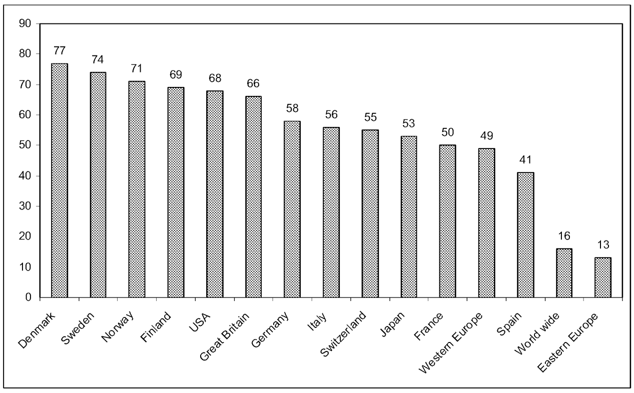


Fig. 1: Internet user per 100 inhabitants 2005 (Bitkom 2007)

So it becomes more and more important to focus on customer behaviour.

CRM AND CKM FOR OBTAINING CUSTOMER LOYALTY

Targeting on influencing customer behaviour is referred to as Customer Relationship Management (CRM). CRM is a philosophy with concerted methods and technologies for optimizing organization and development of customer relationships, leading to a Win-Win situation for both customers and retailers (Foerster et. al. 2002).
The basic principle of CRM lies in the adjustment of the company's business processes on the customers. Critical issues for CRM are mainly insufficient knowledge about the customers as well as deficient consideration of existing business processes and customer processes.

But methods and technologies of knowledge management (KM) could close this gap. Main topic of KM is knowledge acquisition and knowledge utilization in companies and the binding of knowledge as a company resource. Most knowledge acquisition projects are focussed on increasing intra-corporate operations efficiency.

From these facts, it is an important step to transfer the methods and instruments of knowledge management to the processes with customer focus, thus enriching the CRM philosophy with KM (Kolbe et al. 2003).

The synergy of both scientific ideas of CRM and of KM leads to Customer Knowledge Management (CKM). CKM is mainly focussed on solving two fundamental questions:

1. Which knowledge do the customers have?
2. Which knowledge about the customers is known?

The challenge of CKM lies in the efficient collection and analysis of this knowledge to use it by predicting customers' behaviour. Especially one perspective of the behaviour is of major importance, namely customer loyalty. It is known as one of the most important target figures both in offline and online retail.

Companies can benefit from loyalty of customers – especially in E-Commerce - in many ways.

- High rate on repurchase (satisfied customers will come back to the products, brands, or providers)
- Additional sales
- Relatively low sensibility to price changes
- Increasing average length of stay (in online shops)
- More homogenous customer mix
- Innovation initiated by loyal customers

Furthermore, an aspect of customer loyalty not to be underestimated is the reduction of costs (e.g. for acquisition, advertisement).

As customers become more and more sophisticated, the extraction and collection of knowledge on the customers, e.g. about interests, preferences, or buying power is indispensable. Based on this knowledge, it is possible to create customer specific advertisements and offerings. Sprayed advertisements are contradicting customer loyalty.

MARKETING-TRENDS IN E-COMMERCE

Fundamental concepts from classical marketing are also relevant to e-commerce. But there are a lot of unique characteristics of this channel, for example interactivity of customers, or internet as a pull-medium. Fact is that customers are active when buying over internet. This enables to collect knowledge about the customers.

Studying individual **customer behaviour** allows the classification of customers. For example the online bookshop Amazon is permanently practicing in new methods for *studying customer behaviour*. The interests of the customer known by past sales are used to provide targeted offers such as books, CDs, and DVDs specifically chosen for this customer.
Another interesting and important methodology for e-business is **customer segmentation**. The main idea is to categorize customers by comparing their behaviour. Taking again a look at Amazon, two methods they practice in customer segmentation can be shown. On one hand, **cross selling** is used ("customers who bought book A, also bought book B"). Intention of this strategy is to provide overall more products, and services of one company to the customer.

On the other hand, by applying up-selling strategies it is possible to offer higher-class products. Furthermore preferential prices will be offered for additional products and services.

Another effect of internet is that there is more than one channel the customer can use. Online retailers have to react on this and should work **multi-channelling**. With multi-channelling offerings will be provided by several marketing channels to the customer. For example, new book publications will be sent to matching customer groups via email.

Without any difficulties, the customer can do the investigations for cost-efficiently buying products or services on varying online- and offline-channels. This is called **channel hopping**. Therefore, for retailers it is important to manage all these channels. This means, companies should service their customers on every channel that the customer is willing to use.

MISSING PARAMETERS IN E-COMMERCE

The special evolution of marketing methods in e-commerce can not be completed or perfected, because the history of this channel is relatively young.

One of the advantages of online marketing is the accuracy of data gathering, leading to interesting research opportunities.

For example, it is possible to analyse the relationship between customer interests and customer behaviour and externalities. By the term externality we denote events that are not part of the marketing efforts of the retail company. Typical externalities are

- weather forecasts
- events like world cups, big tours of artists, film premieres
- vacation times, holidays
- rare natural phenomenon etc.

The interests of customers are usually not completely static. For sure, there are fundamental interests. But actual interests and behaviours of customers are evolving with their personal experiences. So the above mentioned externalities are influencing the behaviour and interests of customers.

Exactly this relation should be surveyed and analyzed.

CONSIDERING INFORMATION AND KNOWLEDGE ABOUT EXTERNALITIES FOR CKM

Particularly small and medium-sized enterprises (SME) are facing the problem of developing the potential of new sales channels. Therefore some standard software solutions have been built to facilitate fast and simple access to online market places. (cf. Kassel et al. 2005) This is an elegant way to provide access to public market places for retailers with a small existing infrastructure. The interface has to work in two directions. One direction is the transformation of product data from the seller to offers and the automatically placement of the products on the public market. In the other direction the orders from the market place are routed to the seller system for fulfillment.

To be successful on the public market place, learning from the traders which are already experienced is required. Techniques and methods of Knowledge Management and Business Intelligence, e.g. Data Mining or Data Warehousing, can be utilised to generate the essential decision knowledge either explicitly or automatically. Explicit knowledge acquisition from experts in selling goods on public market places (called power sellers) is a first step, which should be chosen according to Sol (2002). This knowledge can be stored in some common knowledge base, which can be seen as a fundament for realizing retailer-specific rule sets. Sellers make decisions by relying on the collected knowledge.

In a recent industry-funded project, an expert system was built using the open source shell Mandarax (Dietrich 2004), which provides an infrastructure for defining, managing, and querying rule-based systems. The main functions of the expert system were the determination of the amount of goods to be offered on the market place as well as the exact point in time for the placement. The rules of the expert system reflected experience knowledge of power sellers. As an important influencing factor for the placement of offers on the market, facts and rules on external incidents like special events were included in the knowledge base from the beginning.

Furthermore the data from all sales activities and transactions were collected and prepared to be utilized. A data warehouse has been built to hold internal data of the market place as well as external data of different data bases delivering information on events which were supposed to be influential for the sales success. This data warehouse, which has been built as a PostgreSQL database, was the first step towards business intelligence, enabling the enterprise to set up a successful customer relationship solution including the analysis of cross selling opportunities, as described by Vitt (2002).

Basically, the extraction and collection of prices over time, auction length, placement time, and external factors like seasons, weather, holidays, were included in the data warehouse. From these basic data, classification numbers were derived to provide compact information on sales success which can be viewed from different perspectives. These classification numbers were directly related to the data describing external events to determine possible correlations between the external events and the internal sales data.

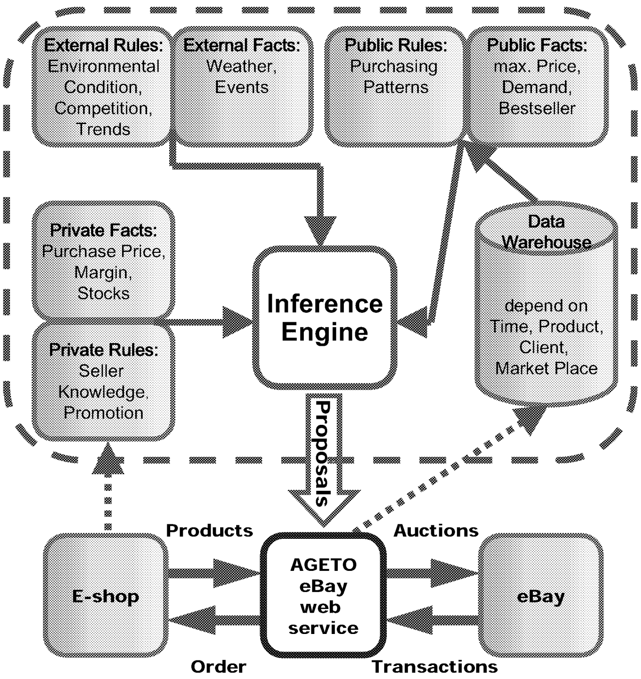

External Rules: Environmental Condition, Competition, Trends
External Facts: Weather, Events
Public Rules: Purchasing Patterns
Public Facts: max. Price, Demand, Bestseller
Private Facts: Purchase Price, Margin, Stocks
Inference Engine
Data Warehouse
depend on Time, Product, Client, Market Place
Private Rules: Seller Knowledge, Promotion
Proposals
Products
Auctions
E-shop
AGETO eBay web service
eBay
Order
Transactions

Fig.2: Expert system architecture (Kassel et al. 2005)

After building the data warehouse and performing initial correlation analysis, the expert system and the data warehouse had been coupled to provide a closed loop application. This enabled the automatic adjustment of the expert system rules to the feedback of the classification numbers, leading to a higher conformance of the expert system with the changing behavior of the customers. Thus, recurring trends could be computed and used for a better prognosis of market activities (Fig.2).

INTELLIGENT ADVERTISEMENT WITH CKM

From the AGETO case study we have learned, that there are ways to automatically adopt offerings on the market to the various external events that are happening in the moment of sales. This principle should now be transferred to the marketing strategy in CKM, leading to a more proficient way of enhancing customer loyalty by providing only those product or service information to the customer that are fitting his interests and his personal life situation. For example, it has to be avoided that customer information gathered in the past is leading to a static classification of the customer which doesn't evolve over the time.

In the first step, a process to achieve knowledge about the customer based on his past behaviour has to be defined. This can excellently be done with e-commerce customers, because it is possible to analyze the complete history of customer interactions. Not only the purchases of the customer are of importance for the qualification, but also the offerings where the customer was looking at some detailed product description which are normally located on special internet pages of the e-shop.

Looking at the customer in this specific way can provide much more information than those that usually can be achieved when the customer looks for goods in offline stores. It is possible to use methods of conceptual clustering to classify the customers.

Even more information about customer preferences can be gained, if these data are interconnected to data describing external events. For example, correlation analysis of customer behaviour and major sports events can provide knowledge of the customer's attitude towards sports and different kinds of sports.

This knowledge is very important, because there is a great up- and cross-selling potential in the hobbies of customers. Customers are willing to pay higher prices for their interests and fun than they are willing to pay for convenience goods they are buying.

But not only the classification and the potential of additional sales are important; customer loyalty can be increased as well. To achieve this, the following second step is important:

Starting with the collection of data concerning external events E, a set R(E) of relevant events can be identified. To determine this set, we need two functions. The first function is a relevance curve

$$r(e,t): E \times T \rightarrow \Re$$

on each specific event e determining the potential that the buying process of customers is influenced by this event at this specific point of time. This relevance curve can be achieved for each individual class of events by statistical measures determining the deviation of sales numbers on event-specific goods (like merchandising articles, special fan equipment etc.) from the long-term average sales numbers of these goods. (Therefore this function delivers a real number)

The second function is a threshold function

$$w(e): \Re \times E \rightarrow \{0,1\}$$

which should be defined on each event e according to business needs. For simplicity, this function can be the same for all events e or the value can be zero for most classes of events. The set R(E) is then determined by the combination of the two functions resulting in

$$R(E) = \{e \in E \mid w(e) = 1\}$$

Now this set can be used to find a subset $C(e_i)$ of the customers which are potentially interested in the event e_i (where $e_i \in R(E)$). Let C(E) be the group of customers that are potentially interested in at least one of the events e_i.

$$C(E) = \bigcup_{e_i \in R(E)} C(e_i)\ ).$$

If you denote with $E(c_j)$ the set of events e which are relevant for customer $c_j \in C(E)$, you have a partial function

$$m(c_j): C \rightarrow E$$

which gives you the set of events for each customer, that can be used for marketing measures.

To provide a useful way of defining the marketing mix for the different customers, some kind of expert system should be used. This has some advantages over hard-wired applications, namely

- the higher flexibility of rule-based expert systems can be used to provide the fitting marketing mix for a specific customer,
- a better understanding on how the marketing measures are used (most expert systems are providing an explanation component to provide information about their reasoning process),
- a more consistent way to deal with great varieties in the set of events that have to be considered for a specific customer (rules can be used hierarchically reducing the complexity of the expert system).

The process results in one of several possible measures:

1. A marketing email can be sent to the customers, which is specifically built for his individual requirements. This email is built according to the classification of the customer and to his personal interests. Therefore it provides useful information

about the customer (from his personal perspective), leading to an increase of his loyalty to the retailer.

2. Based on the knowledge about the customer and his situation, even different views of the web store can be realized by using dynamic components which are chosen at run-time of the application. This includes direct advertising messages as well as different up-selling or cross-selling goods, which can be placed at any specific place on the web page.

Caused by the dynamic contents of the web page as well as the marketing email, the success can directly be measured. This success is the direct feedback loop for further tailoring of future measures.

SUMMARY

Dynamic expert systems coupled with data warehouse concepts can be seen as an important enabler to improve customer loyalty. Customer Knowledge Management (CKM) can be achieved easily for e-commerce, leading not only to enhanced loyalty, but directly to increase sales numbers for the customer. The concept of the AGETO eBay Web Service can be adopted to provide information about the customer combined with information about external events influencing the behaviour of customers depending on time. The major steps of this adoption are sketched and some possible results are delineated.

REFERENCES

Abramowicz, W. 2003. "Knowledge-Based Information Retrieval and Filtering from the Web". Kluwer Academic Publishers.

Aebi, R. 2000. "Kundenorientiertes Knowledge Management". Addison Wesley Verlag.

Bitkom, 2007. Survey about Internet Users, www.bitkom.de.

Deutscher Managerverband e. V. 2003. Die Zukunft des Managements. vdf Hochschulverlag AG. Zürich

Dietrich, J. 2004. "A Rule-Based System for eCommerce Applications". In *Proceedings of Knowledge-Based Intelligent Information and Engineering Systems: 8th International Conference* (Wellington, New Zealand, Sep. 20-25). Springer LNCS 3213 / 2004, Heidelberg, 455-463.

Favier, J. 2006, Europe's eCommerce Forecast: 2006 To 2011, Forrester Research, http://www.forrester.com/Research/Document/Excerpt/0,7211,38297,00.html

Foerster, A. and Kreuz, P. 2002. Offensives Marketing im E-Business, Springer-Verlag, Berlin.

Gadatsch, A. 2002. "Management von Geschäftsprozessen". Vieweg Verlag, Braunschweig/Wiesbaden.

Grebenstein, K.; Schumann, C.A.; Tittmann, C.; Tsering, G.; Weber, J.; and Wolle, J. 2003. "Globale IT-Infrastruktur für die interkulturelle Kommunikation". In *Kommunikation in der globalen Wirtschaft 2003,* Bleich, S.; Jia; and W.; Schneider, F. (Eds.) Peter Lang Verlag, Frankfurt a.M., 135-153.

Hannig, U. 2002. "Knowledge Management und Business Intelligence". Springer Verlag. Berlin – Heidelberg – New York.

Horton, F. W. and Marchand, D. A. 1982 "Information Management in Public Administration". Information Resources Press. Arlington.

Kassel, S. and Grebenstein, K. 2004 "AI-based integration of business intelligence and knowledge management in enterprises", Proceedings Conference AIAI 2004. Toulouse.

Kassel, S.; Schumann, C.-A.; Grebenstein, K.; Tittmann, C. 2005 "A Knowledge-Based Decision-Support-System for e-Commerce", Proceedings Conference Euromedia 2005. Toulouse

Kolbe, L. M.; Österle, H. and Walter Brenner, W. 2003. Customer Knowledge Management. Springer Verlag, Berlin.

Kuppinger, M. and Woywode, M. 2000. "Vom Internet zum Knwoledgemanagement Veränderungen in der Informationsgesellschaft". Carl Hanser Verlag München Wien.

Lehner, W. 2003. "Datenbanktechnologie für Data-Warehouse-Systeme". dpunkt verlag Gmbh. Heidelberg.

Ortmann, G.; Sydow, J. 2001. "Strategie und Strukturation". Gabler Verlag. Wiesbaden.

Puppe, F.; Gappa, U.; Poeck K. 1996. "Wissensbasierte Diagnose- und Informationssysteme". Springer Verlag. Berlin – Heidelberg – New York.

Sachs L. 1992. "Angewandte Statistik". Springer Verlag. Berlin – Heidelberg – New York.

Salcedo L. Market Forecast Report European Commerce, 2003–2009. 2004. JupiterResearch. Jupitermedia Corp.

Sol, H. 2002. "Expert Systems and Artificial Inteligence in Decision Support Systems". Kluwer Academic Publishers.

Vitt, E.; Luckevich, M.; Misner, S. 2002, "Business Intelligence: Making Better Decisions Faster", Microsoft Press, Redmond.

AUTHOR BIOGRAPHY

STEPHAN KASSEL was born in Bad Kreuznach, Germany. He studied computer science at the University of Kaiserslautern. After obtaining his degrees, he worked at several universities in Germany in the areas of Distributed Artificial Intelligence and Information Systems. He earned his Ph. D. at the Technical University of Chemnitz, in 1998. In 1999 he started to work for Intershop, an E-Commerce vendor, in international e-commerce projects. Since 2003, he holds a chair for Information Systems at the University of Applied Sciences in Zwickau, Germany.

AUDIO VISUAL APPLICATIONS

A COMPARISON OF THE ILD AND TDOA SOUND SOURCE LOCALIZATION ALGORITHMS IN A TRAIN ENVIRONMENT

Joost Voordouw, Zhenke Yang, Leon J.M. Rothkrantz, Charles A.P.G. van der Mast
Faculty of Electrical Engineering, Mathematics and Computer science,
Delft University of Technology
Mekelweg 4, 2628CD Delft, The Netherlands
E-mail: Joost@mmi.tudelft.nl, {Z.Yang, L.J.M.Rothkrantz, C.A.P.G.vanderMast}@ewi.tudelft.nl

KEYWORDS

Aggression Detection, Sound Source Localization, Train Compartment, SSL, TDOA, TDE, ITD, ILD.

ABSTRACT

Aggressive behavior in public spaces is an unwanted phenomenon in our society and therefore needs to be eliminated whenever possible. This paper describes a system that enables observation systems to find the location of people that are making noise. The system uses the input of multiple microphones to determine the location of such a person. In this paper two sound source localization (SSL) approaches are investigated and tested. The first approach is based on the signal (also interaural) level difference (ILD) between different microphones, the second approach is based on the time delay of arrival (TDOA). A difference between the proposed system and already existing systems, is the difference in the setup of the microphones. In the proposed system, the microphones are placed in a train compartment.

INTRODUCTION

Aggressive behavior in public spaces is an unwanted phenomenon in our society. When this aggressive behavior escalates, it can distress bystanders or result in damaged properties. This paper investigates the possibilities of detecting possible aggressive situations in train compartments based on sound source localization. By using data from multiple microphones, we want to estimate the location of an aggressive sound source. To test the sound source localization system in a train compartment, experiments have been done with the help of professional stand-up comedians (Figure 1). Most existing surveillance systems do not take sound into account. The systems that do take sound into account, only use the content of one sound signal, and do not combine the different sound signals received at different places. The system described in this paper combines the data of multiple microphones to find the location of a sound source.

Figure 1: Stand-up comedians showing aggressive behavior in a train compartment.

This paper presents two different approaches that are implemented and compared. The two approaches are the TDOA approach and the ILD approach. Both are evaluated in two different settings: an anechoic room and a train compartment.

This paper first gives a short overview of related work. Next, it discusses the TDOA method and the ILD method for sound source localization. Then the details of our application, called *LocalIce*, is explained, followed by the experiments that were done using *LocalIce* and the results.

RELATED WORK

An example of an surveillance system in public places is the system build by Cupillard et al. (Cupillard et al. (2004)). This system uses multiple cameras to perform visual surveillance of metro scenes and consists of three main tasks: (a) motion detection and frame to frame tracking, (b) combining multiple cameras and (c) long term tracking of individuals, groups of people and crowd evolving in the scene. In camera observation systems, only things that are visible to the camera can evoke a response. However, incidents often start with verbal aggression, which is not detectable by visual observation alone. A surveillance system that uses video processing as well as audio signals is described by Velastin et al. (Velastin et al. (2002)). The audio detection is used to detect abnormal sound signatures and to act when such events occur. Another system that can recognize aggressive situations is (Sigard (2007)), which is able to detect verbal aggression. At a location microphones are placed, which are connected to a computer specially designed to perform the sound analysis. In case of detected aggression, a link is made to a central server in the observation room. From there, the nearest observation camera can be activated and an observation team can be alerted. The idea to use audio data to enhance video surveillance, as applied in the examples above, can be optimized by complementing the data with sound source localization. For example video cameras can be directed to focus on the important parts of a scene.

Two different approaches for SSL exist, the first approach (TDOA) uses the fact that microphones located at different distances from a sound source will receive a sound signal at different times (Ben-Reuven and Singer (2003), Dibiase (2000) and Rabinkin et al. (1996)). The second approach is based on the fact that the intensity of the received signal at different microphones decreases when the distance from the microphone to the source increases. A system that is based on TDOA is described in Valin et al. (2003). The paper presents a robust sound source localization method in three-dimensional space using an array of 8 microphones. The system is developed for use with a robot. The hearing sense on a mobile robot is important because it is omnidirectional and it does not require direct line-of-sight with the sound source. Such capabilities can nicely complement vision to help localize a person or an interesting event in the environment. The mobile robot can localize different types of sound sources over a range of 3 meters and with a precision of 3 degrees in real time.

A system based on the differences in the intensity of microphone signals is described by Birchfield and Gangishetty (Birchfield and Gangishetty (2005)). Although interaural level differences has been studied extensively in natural systems, it remains an untapped resource for computer-based systems. The paper investigates the possibility of using ILD for acoustic localization, deriving constraints on the location of a sound source given the relative energy level of the signals received by two microphones. Experimental results show that accurate acoustic localization can be achieved using ILD alone.

TIME DELAY OF ARRIVAL

This section describes the TDOA model for sound source localization. Sound source localization based on TDOA consists basically of two steps. The first step is determining the time delays for selected microphone pairs. The second step is using these time delays to compute a location estimate.

Time Delay Estimation

As sound travels with a constant speed through the air, it takes time to arrive from the source to the microphone location. The farther away a microphone is located from the sound source, the longer it will take for the sound to reach the microphone. Two microphones located at different places will receive the same sound (plus disturbances) at a different time. The TDOA method takes advantage of this 'delayed' arrival to estimate the location of the source. The signal that arrives at microphone i can be modelled as:

$$y_i(t) = s(t - \Delta t_i)/d_i + \xi_i(t), \tag{1}$$

where $y_i(t)$ is the (instantaneous) sound pressure as received by microphone i, $s(t - \Delta t_i)$ is the delayed source signal, d_i is the distance from the source to the ith microphone and $\xi_i(t)$ is the noise received by microphone i. The time delay for a given microphone pair (i, j) is:

$$\Delta t_{ij} = \Delta t_j - \Delta t_i, \tag{2}$$

From Equation (1) and (2) it follows that:

$$y_j(t) = s(t - \Delta t_i - \Delta t_{ij})/d_j + \xi_j(t), \tag{3}$$

Equation (1) and (3) show that the signals from both microphone i and j are scaled and time-shifted versions of the original source signal. Beside the time-shift and scaling also some noise is present in both signals. The difference in time-shift between the signals of microphone i and microphone j is Δt_{ij}. To make the time delay estimation more robust against noise and reverberations, a filter is used. To find the TDOA between two signals, the cross correlation between these two signals is calculated. The cross correlation ($c_{ij}(\Delta t)$) reaches its maximum at the time delay (Δt) where the shifted versions of $s(t)$ align. The cross correlation of signals $y_i(t)$ and $y_j(t)$ is:

$$c_{ij}(\Delta t) = \int_{-\infty}^{+\infty} y_i(t) y_j(t + \Delta t) dt. \tag{4}$$

To reduce the computational complexity of Equation (4) the inverse Fourier transform of the *cross spectrum* ($C_{ij}(\omega)$) (Valin et al. (2003)) can be computed.

$C_{ij}(\omega)$ is the Fourier transform of the cross correlation function and is obtained from the two individual Fourier spectra of $y_i(t)$ and $y_j(t)$:

$$C_{ij}(\omega) = Y_i(\omega) Y'_j(\omega), \tag{5}$$

where $Y_i(\omega)$ is the Fourier transform of $y_i(t)$ and $Y'_j(\omega)$ is the complex conjugate of the Fourier transform of

$y_j(t)$. To get the cross correlation in terms of $Y_i(\omega)$ and $Y_j'(\omega)$, the inverse Fourier transform of Equation (5) is calculated:

$$c_{ij}(\Delta t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} Y_i(\omega)Y_j'(\omega)e^{j\omega\Delta t}d\omega.$$

To minimize the influence of noise and other disturbances on the time delay estimation, both signals can be filtered by a weighting function. The *generalized cross correlation(GCC)* is the cross correlation of the filtered versions of $y_i(t)$ and $y_j(t)$. Denoting the Fourier transform of these filters by $G_1(\omega)$ and $G_2(\omega)$, the GCC is:

$$R_{ij}(\Delta t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left(G_1(\omega)Y_1(\omega)\right)\left(G_2(\omega)Y_2(\omega)\right)' e^{j\omega\Delta t}d\omega,$$

which results after rearranging in:

$$R_{ij}(\Delta t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Psi_{ij}(\omega)Y_i(\omega)Y_j'(\omega)e^{j\omega\Delta t}d\omega, \qquad (6)$$

where $\Psi_{ij}(\omega) = G_1(\omega)G_2'(\omega)$ is the weighting function. $R_{ij}(\Delta t)$ should show a peak for the Δt that maximizes the GCC between $y_i(t)$ and $y_j(t)$. This Δt corresponds to the estimated time delay between microphone i and j:

$$\tilde{\Delta t_{ij}} = \arg\max_{\Delta t} R_{ij}(\Delta t).$$

Weighting Functions

Weighting functions can be used to filter the signals to get a more reliable estimation of the time delay. One weighting function is the *phase transform (PHAT)* weighting function (Ben-Reuven and Singer (2003), Dibiase (2000) and Rabinkin et al. (1996)). This function is proven to be more robust to reverberation than other weighting functions (such as the Maximum Likelihood Weighting Function) Dibiase (2000). The PHAT weighting function normalizes the product of the magnitudes of the captured signals and is defined as follows:

$$\Psi_{12}(\omega) = \frac{1}{|Y_1(\omega)Y_2'(\omega)|} \qquad (7)$$

Substitution of Equation (7) into Equation (6) leads to the following cross correlation function with the weighting function included:

$$R_{ij}(\Delta t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{Y_i(\omega)Y_j'(\omega)e^{j\omega\Delta t}}{|Y_i(\omega)Y_j'(\omega)|}d\omega \qquad (8)$$

The Δt that gives the biggest cross correlation is the estimated time delay. Another useful weighting function is a bandpass filter. This filter suppresses frequencies that are not in the band containing most of the speech data, which is from 300Hz to 8000Hz.

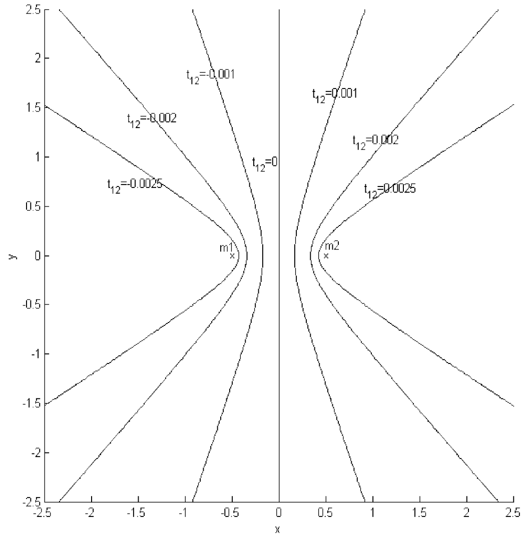


Figure 2: Possible locations for the sound source for a single microphone pair (m_1, m_2) for different Δt_{12}'s.

Location Estimation

Once the time delay is calculated between several pairs of microphones, the actual sound source localization can be performed. The sound source localization estimation used in this paper is based on a least mean square (LMS) method (Dibiase (2000)). The following calculations only take into account two dimensions instead of three dimensions to ease the reading of the formulas. The calculations can easily be extended to the three dimensional case.

For a microphone $i(x_i, y_i)$ and a sound source $s(x_s, y_s)$ the time it takes for sound to travel from the source to the microphone can be calculated by:

$$\Delta t_i = \frac{d(i,s)}{V_{sound}} = \frac{\sqrt{(x_i - x_s)^2 + (y_i - y_s)^2}}{V_{sound}},$$

where V_{sound} denotes the speed of sound and $d(i,s)$ gives the distance between microphone i and the sound source.

When Δt_{ij} is computed, then for Δt_i and Δt_j (see Equation 2) still exist multiple possible solutions. These solutions define a curve of all places where the sound source could be, which fulfills the following equation:

$$\Delta t_{ij} = \Delta t_i - \Delta t_j = \frac{d(p,i) - d(p,j)}{V_{sound}},$$

where p is a point (x_p, y_p) on the curve. Figure 2 is obtained when the possible sound source locations are calculated for different Δt_{ij} and when V_{speed} is set to 340 meters/second $(i = 1, j = 2)$.

By using the data from multiple microphone pairs, multiple curves can be drawn. Theoretically, when no

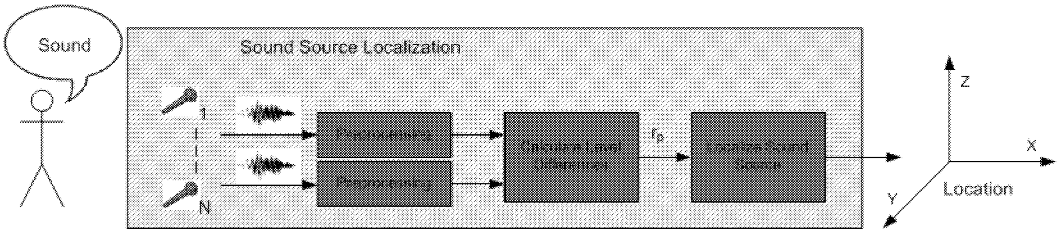


Figure 3: *LocalIce* system overview.

disturbances are present, all these curves should intersect in a single point. However, because of multiple sources of distortion, this is not the case in practice. When having more than two microphone pairs, the corresponding curves will not have a single intersection point and a LMS solution can be applied. For a predefined set of locations (candidate locations) the mean square error is computed. The location resulting in the smallest RMS error is assumed to be the sound source location. The expected TDOA's, for each microphone pair, for a given location q are calculated as follows:

$$\tilde{\Delta} t_{ij}(q) = \frac{d(m_i,q)-d(m_j,q)}{V_{sound}} \qquad \text{for } i,j \in \{1...M, i \neq j\}.$$

The RMS error for a candidate location:

$$E(q) = \sqrt{\sum_{ij=1}^{L} (\Delta t_{ij} - \tilde{\Delta} t_{ij}(q))^2},$$

where L is the number of microphone pairs.

The sound source location is estimated by the candidate location that minimizes $E(q)$:

$$\tilde{q} = \arg\min_{q} E(q). \tag{9}$$

INTERAURAL LEVEL DIFFERENCES

The interaural level differences model uses the level differences in the signals received by the different microphones. This model also consists of two parts. The first part is the estimation of the level differences, the second part is the estimation of the sound source location.

Level Difference Estimation

Depending on the distance from the sound source to the microphones, each microphone receives different sound signals. This effect is caused by the diffusion of the sound energy as it travels through space. The instantaneous sound pressure (i.e. signal) received by the ith microphone can be modelled as:

$$y_i(t) = s(t)/d_i.$$

$y_i(t)$ is the sound pressure as received by microphone i and $s(t)$ is the source signal, d_i is the distance from the source to the ith microphone. These equations ignore the time shift in the signal, caused by the travelling distance. For the interaural level difference localization model it is assumed that these differences are too small to have a significant influence.

A convenient and more usable method to work with the signals received by the different microphones, is by doing calculations on the *effective* sound pressure instead of the instantaneous sound pressure. To calculate the *effective* sound pressure (p_i) for microphone i while having a sound source at a fixed location during time interval $[0, T]$, the following equation is used:

$$\begin{aligned} p_i &= \sqrt{\frac{1}{T}\int_0^T y_i^2(t)dt} \qquad (10) \\ &= \frac{1}{d_i}\sqrt{\frac{1}{T}\int_0^T [s^2(t)]dt} \end{aligned}$$

From Equation (10) it follows that the effective sound pressure is inversely proportional to the distance from the microphone to the sound source: $p_i d_i = p_j d_j$. The ratio of the effective sound pressure between microphone i and microphone j is p_i/p_j (denoted as r_p) and can be written in terms of the distance from a microphone to the sound source:

$$r_p = d_j/d_i \tag{11}$$

When the effective sound pressures for microphone i and microphone j are known, r_p can be calculated.

Location Estimation

According to Equation (11), r_p is not only the pressure ratio, but also the ratio between the two distances from microphone i and j to the sound source:

$$r_p = \frac{\sqrt{(x_j - x_s)^2 + (y_j - y_s)^2}}{\sqrt{(x_i - x_s)^2 + (y_i - y_s)^2}}, \tag{12}$$

with microphone i at location (x_i, y_i) and the sound source at location (x_s, y_s).

When using the calculated r_p of one microphone pair, the sound source is restricted to lie on a circle or a line.

To get an estimation of the position of the sound source when using more than two microphones, a least

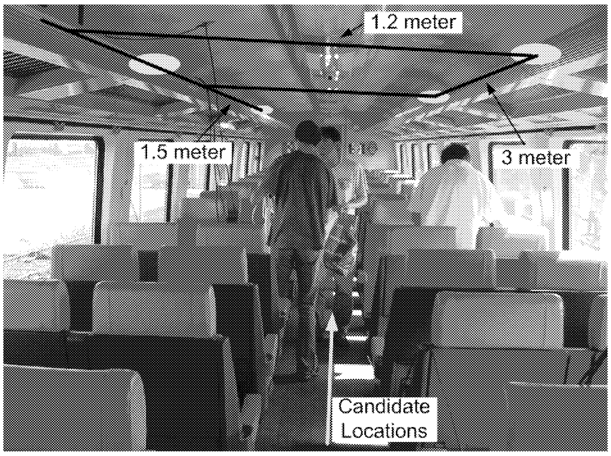


Figure 4: The experiment setup in the train. The brighter ovals are the microphone positions. The distance between two microphones along the length of the compartment is 3 meter. The distance between the left and right side is 1.2 meter. The candidate locations are all in the middle of the passage of the train.

mean square algorithm can be used. This approach first selects a number of candidate locations (q's) and calculates the expected $\tilde{r}_p$'s for all these q's (using Equation (12)). The next step is calculating the sum of the squared differences between the r_p's and the $\tilde{r}_p$'s for each location:

$$S_q = \sum_{p=1}^{L} (\tilde{r}_p - r_p)^2, \tag{13}$$

where L is the number of microphone pairs. The q having the smallest S_q is the estimated sound source location. Also in combination with the ILD method, a bandpass filter can be used to suppress the frequencies that are not relevant. A PHAT filter is not useful in combination with the ILD method.

IMPLEMENTATION

The methods described in this paper are implemented in Matlab, the application was given the name *LocalIce* (code available upon request). Figure 3 shows an overview of *LocalIce* for the ILD method. To ease the testing of the different algorithms *LocalIce* can perform the calculations on prerecorded audio data. The inputs that *LocalIce* needs are:

- The dimensions of the room the data was recorded in.
- The properties of the grid of candidate locations.
- The positions of the microphones in the room.
- The sample size to use in ms.

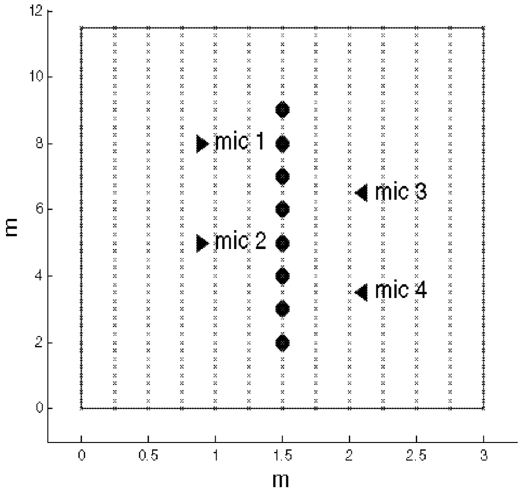


Figure 5: Map (top view) of the part of the train compartment where the experiments were done. The triangles denote the microphones, the circles are the predefined sound source locations. The small crosses are candidate locations.

- The signals that are recorded by these microphones. The number of signals entered needs to correspond with the number of microphones.
- The minimal signal-to-noise ratio (SNR) needed for a sample to be processed.
- A (manually selected) sample of noise as input. Based on this sample, *LocalIce* can calculate the SNR for a sample in a certain time frame. If this SNR is higher than the minimal SNR, position estimation calculation can be performed.

The output of *LocalIce* is a map in which each candidate location is assigned a value. For the ILD method, this value is S_q (Equation 13). For the TDOA method, this value is E_q (Equation 9). For both cases, a lower value means that this candidate location has a higher chance to be the real sound source location.

EXPERIMENTS

The experiments are done in two different environments. One part is done in a common 'Benelux Train' as used by the Netherlands Railways. During the experiment the train is not moving, but the air conditioning is functioning and producing noise. The experiments are concentrated in an area having a length of about 7.5 meter (Figure 4). Figure 5 shows the schematic setup of the four microphones and their directions. In the same figure, the candidate locations and the predefined sound source locations are shown. The test data consists of recordings of a man speaking a Dutch sentence at the

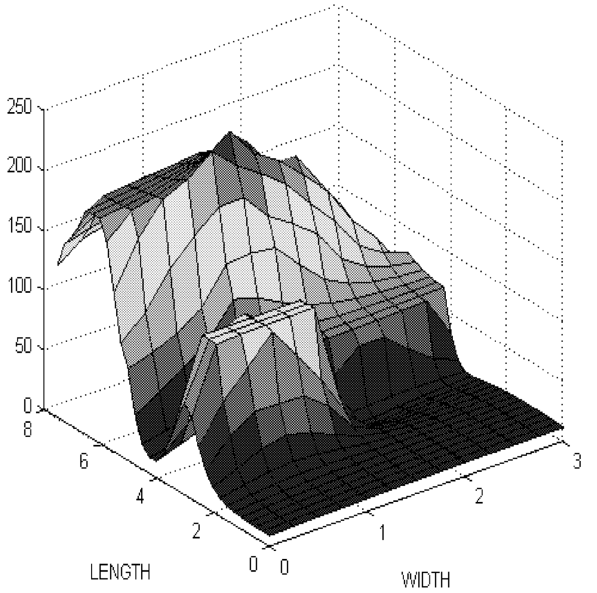


Figure 6: The ILD algorithm applied on measurements in an train compartment using a sample size of 200 ms gives as sound source location (2.0;3.25) where the actual location is (1.5 ; 3.0) and thus has an error of 0.6 meter.

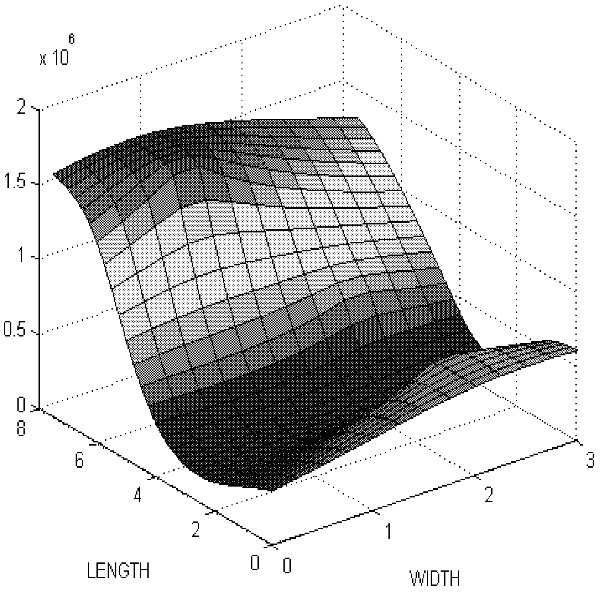


Figure 7: The TDOA algorithm applied on measurements in an train compartment using a sample size of 200 ms gives as sound source location (1.5;3.0) which is the actual location.

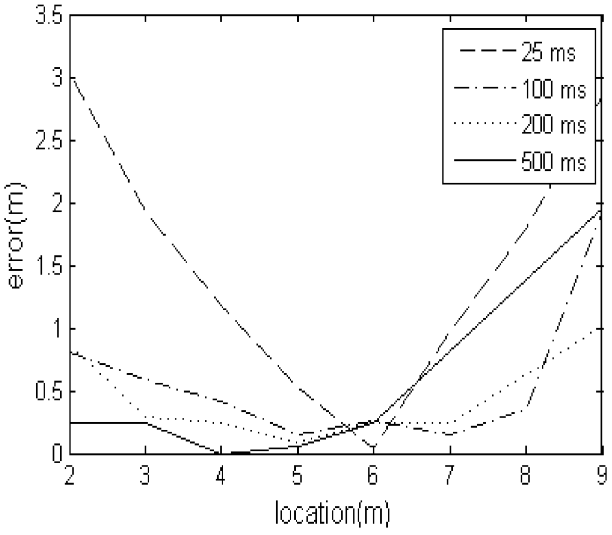


Figure 8: Performance of the TDOA algorithm in the anechoic room for different sample sizes.

predefined locations. These locations are all in the middle of the passage. The distance from the front of the train to the sound source location is increased with one meter for every measurement.

The same experiments were done in an anechoic room (second environment). This is a room that has (almost) no reverberation. These experiments are done to gather reference and comparison material for the data gathered during the train experiment. As a consequence, the microphone setup, the candidate locations and the predefined source locations are identical to the ones used in the train (Figure 5). During the experiments in the anechoic room, a prerecorded sound sample of a man counting twice from one to five is used. This sample is played at the predefined locations by a loudspeaker connected to a laptop. In both experiments the sound sources are in the same plane as the microphones, therefore the height coordinate is ignored.

RESULTS

To compare the performance of the ILD and the TDOA method, the sound files that are recorded during the experiments are given as input to *LocalIce*. To give an impression of the results figure 6, 7, 10 and 11 are examples of the output of *LocalIce*. Figure 10 and 11 show a plot of the result of respectively the ILD and the TDOA algorithm on a single sound sample taken in the anechoic room using a sample size of 200 ms. Figure 6 and 7 show a plot of the result of respectively the ILD and the TDOA algorithm on a single sound sample taken in the train compartment using a sample size of 200 ms. It is interesting to note that the TDOA plots (Figure 6 and 7) are much smoother than the ILD plots (Figure 10 and 11). The peaks that are present in the ILD plots are dis-

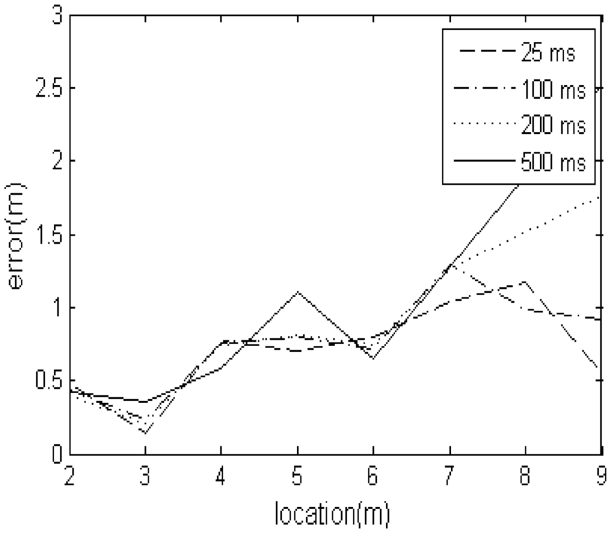


Figure 9: Performance of the ILD algorithm in the anechoic room for different sample sizes.

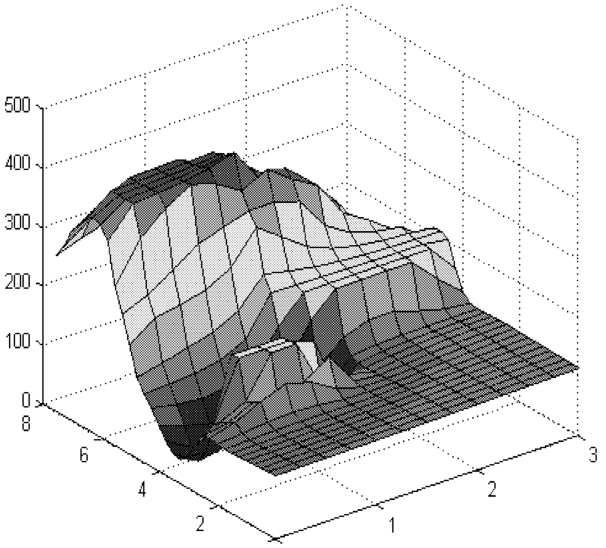


Figure 10: The ILD algorithm applied on measurements in an anechoic room using a sample size of 200 ms gives as sound source location (1.25;3.5) where the actual location is (1.5 ; 3.0) and thus has an error of 0.6 meter.

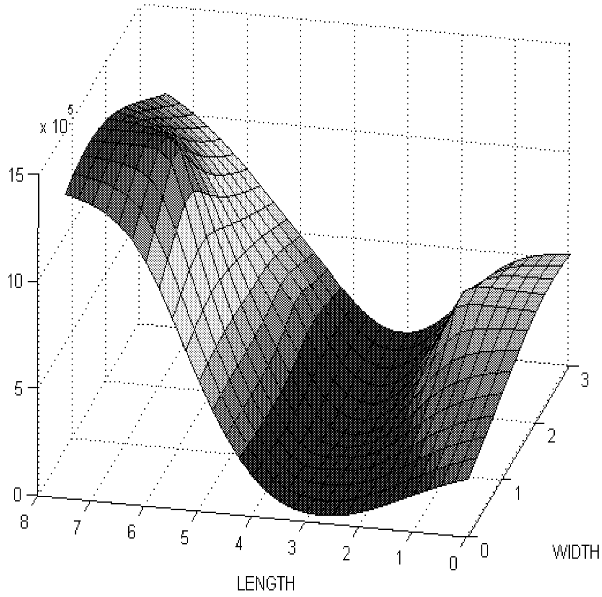


Figure 11: The TDOA algorithm applied on measurements in an anechoic room using a sample size of 200 ms gives as sound source location (1.5;3.0) which is the actual location.

turbances caused by the microphone locations. The ILD algorithm is not able to correct for these disturbances.

The performance of the algorithms is influenced by the choice of the sample size, which can be given as an input to *LocalIce*. Figure 8 shows the performance of the TDOA algorithm in the anechoic room at different locations for different sample sizes. Figure 9 shows this for the ILD algorithm. For both algorithms a sample size of 200 ms is a good choice.

We are particularly interested in the performance of the algorithms in a train compartment. Figure 12 shows the performance of the ILD algorithm in the train compartment at different locations with and without a bandpass filter. Figure 13 shows the performance of the TDOA algorithm in the train at different locations with a PHAT filter, a bandpass filter and without any filter. For both algoritms the use of a bandpass filter improves the performance.

The figures show that the location estimation for the first meters is better than the estimations for the last meters. This probably has to do with the direction of the sound sources and the orientation and position of the microphones. If a microphone is oriented to the direction the sound comes from, the SNR tends to be higher, making the estimation more accurate.

CONCLUSION AND RECOMMENDATIONS

Sound source localization based on TDOA is quite accurate when only little reverberation is present and the length of the samples is around 200 ms. When a sig-

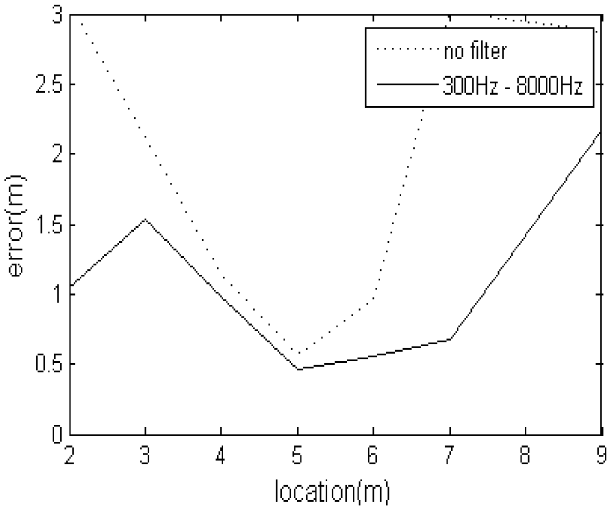


Figure 12: Performance of the ILD algorithm in the train compartment when using a bandpass filter compared to not using a filter.

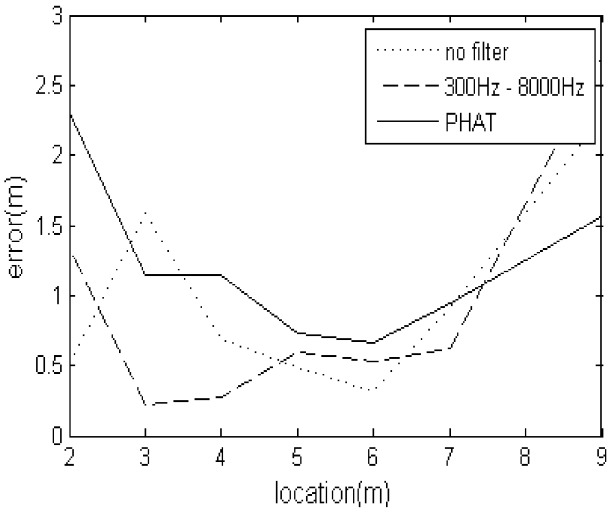


Figure 13: Performance of the TDOA algorithm in the train compartment when using a bandpass filter compared to using a PHAT filter and not using a bandpass filter.

nificant amount of reverberation is present, the performance of TDOA declines. For environments with little reverberation, TDOA performs better than ILD for sample sizes of 200 ms. When there is a lot of reverberation as in the case of the train compartment, it is not possible to give a judgement about the algorithm that should be used. Further research is needed to find out which of the two algorithms performs best under different circumstances.

On the way to automatic aggression detection in trains it would be useful to investigate the following topics. Improving the SSL algorithm, by using filters to decrease the disturbance caused by reverberation or by combining the two discussed SSL methods. The direction of the microphones also will have a significant influence on the performance of the algorithms. Another interesting point is using the content of the audio streams as an aggression indicator. In combination with SSL is this very promising in the fight against aggression in the public transport.

ACKNOWLEDGEMENTS

This research is done as part of an ongoing project at the TUDelft and funded by the MultimediaN project. The experiments are done in collaboration with NS/ProRail.

REFERENCES

Ben-Reuven, E. and Singer, Y. (2003). Discriminative binaural sound localization. In S. Becker, S. T. and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*, pages 1229–1236. MIT Press, Cambridge, MA.

Birchfield, S. and Gangishetty, R. (2005). Acoustic localization by interaural level difference. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, volume 4.

Cupillard, F., Avanzi, A., Brmond, F., and Thonnat, M. (2004). Video understanding for metro surveillance. In *Proceedings of the IEEE International Conference on Networking, Sensing & Control*, Taipei, Taiwan.

Dibiase, J. (2000). *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays.* PhD thesis, BROWN UNIVERSITY.

Rabinkin, D. V., Renomeron, R. J., Dahl, A. J., French, J. C., Flanagan, J. L., and Bianchi, M. (1996). Dsp implementation of source location using microphone arrays.

Sigard (2007). http://www.soundintel.com/products.html.

Valin, J., Michaud, F., Rouat, J., and Létourneau, D. (2003). Robust sound source localization using a microphone array on a mobile robot. In *Proceedings International Conference on Intelligent Robots and Systems.*

Velastin, S. A., Maria Alicia Vicencio-Silva, B. L., and Khoudour, L. (2002). A distributed surveillance system for improving security in public transport networks. *Special Issue on Remote Surveillance Measurement and Control*, 35(8):209–13.

Signal-Coherent Video Watermarking Schemes Based on Visual Cryptography

Cezar Pleşca
IRIT UMR CNRS 5505
ENSEEIHT-Informatique
Toulouse, France
email: plesca@enseeiht.fr

Victor Patriciu
Military Technical Academy
Computer Science Department
Bucharest, Romania
email: vip@mta.ro

Vincent Charvillat
IRIT UMR CNRS 5505
ENSEEIHT-Informatique
Toulouse, France
email: charvi@enseeiht.fr

ABSTRACT

Visual cryptography is a visual form of information concealing. Usually, in a visual cryptography scheme, a secret image is split in two or more shadow images called shares. As for threshold schemes, the secret image is not revealed or recovered unless a minimum number of shares are stacked together. This paper carries out an analysis of a video watermarking method based on visual cryptography. A binary logo, representing the ownership of the host video, is repeatedly split into shares that are inserted into video frames. We investigate the resilience of this scheme. We propose two alternatives to improve it following the principle of coherent watermarks. The first method uses image signatures to gracefully modify a pattern, i.e. the share, along a video scene. The second strategy embeds logo shares into subsequent mosaic scenes. Finally, experimental results show that the proposed schemes provide better resilience against common hostile and non-hostile attacks.

KEYWORDS
Video Watermarking, Visual Cryptography, Visual Hash

1 INTRODUCTION

Digital watermarking was initially introduced in the early 90's as a complementary protection technology since encryption alone shows its limits. Indeed, sooner or later, encrypted multimedia content is decrypted to be presented to human beings. At this very moment, the content is left unprotected and can be perfectly duplicated, manipulated and redistributed at a large scale. Thus, a second line of defense has to be added to address this issue. This is the main purpose of digital watermarking which basically consists in hiding some information into digital content in an imperceptible manner.

Watermarking problem is a complex trade-off between three parameters: fidelity, robustness and capacity [1]. **Fidelity** is related to the distortion that the watermark embedding process is bound to introduce; the inserted watermark should remain imperceptible to a human user. **Robustness** can be seen as the ability of the detector to extract the watermark from some altered watermarked data. **Capacity** is the number of bits encoded by the watermark.

Digital watermarking has first been extensively studied for still images. Today, however, many new watermarking schemes are proposed for other media: audio, video, text and 3D meshes. If the increasing interest concerning digital watermarking during the last years is most likely due to the increase in concern over copyright protection of digital content, it is also emphasized by its commercial potential.

Video watermarking is mostly considered as watermarking a sequence of images. However, even if watermarking images and video is a similar problem, it is not identical. New problems, new challenges show up and have to be addressed. First, there are many non-hostile video processing which are likely to alter the watermark signal. Nonhostile refers to the fact that even content providers are likely to process a bit their digital data for resource management reasons. These video processing operations include: *photometric attacks* (DA/AD conversion, transcoding, format conversion, chrominance resampling), *spatial desynchronization* (display formats change, changes of spatial resolution), *temporal desynchronization* (changes of frame rate) and *video editing* (cut-and-split, transition effects).

Second, resilience to collusion, a problem that has already been pointed out for still images, becomes much more critical in the context of video [4]. Collusion attacks generally refers to a set of malicious users merging their knowledge, e.g. the watermarked data, to produce an unwatermarked content. In the case of video, two types of collusion attacks are possible : inter-video collusion and intra-video collusion. In inter-video collusion, different watermarked versions of the same video are combined to produce an unwatermarked copy of the video. Intra-video collusion attacks exploit the inherent redundancy in the video frames or in the watermark to produce an unwatermarked copy of the video.

This paper focuses on intra-video collusion attacks which have to be considered when analyzing the security of watermarking schemes. The rest of the paper is organized as follows. Section 2 presents a video watermarking scheme based on visual cryptography whose security is evaluated against the intra-video collusion attacks described in section 3. To overcome its security shortcomings, we propose two improvements. The first scheme, based on visual hash functions, is detailed in section 4 and prevents classic collusion attacks. To cope with more complex collusion attacks (registration-based attacks) we designed a second scheme which uses video mosacing to embeds parts of a fingerprinting logo into subsequent panoramic video scenes. Due to lack of space, we present this method into a detailed version of this paper that could be found at http://www.enseeiht.fr/~charvi/papers/visualCryptographyWatermark.pdf. The robustness and the security of these methods are evaluated using simulations and the results corresponding to the first method are presented in section 5. Section 6 concludes the paper and gives some perspectives to this work.

2 RELATED WORK

Many video watermarking algorithms have been proposed in the scientific literature and three major trends can be isolated [1]. The straightforward approach is to consider a video as a sequence of still images and to reuse an existing watermarking scheme for still images. Another point of view exploits the additional temporal dimension in order to design robust video watermarking algorithms. The last trend basically considers a video stream as some data, compressed according to a specific video compression standard whose characteristics are exploited to obtain efficient watermarking schemes.

Both digital watermarking and visual cryptography can involve a hidden image. However, their concepts are different. For visual cryptography, a set of share binary images is used to protect the content of the secret image. In digital watermarking, the hidden image is usually embedded in the host data while preserving the quality of watermarked content.

Some research proposed joint visual cryptography and watermarking algorithms that combine the merits of both approaches [7, 9]. The general principle is to represent a logo image by several different shares and to embed these shares into the host data using watermarking techniques. Most of digital watermarking algorithms can be applied into such a scheme mainly due to the random noise-like nature of the generated shares. In line with this general idea, Houmansadr et al. proposed in [7] a novel video watermarking method based on visual cryptography. The remainder of this section briefly introduces visual cryptography and their method.

Visual Cryptography Visual cryptography is a type of cryptographic scheme to conceal images whose secret can be decrypted without any cryptographic computations [8]. It is a visual variant of the k out of n secret sharing problem. One would produce transparent layers that contain part of the secret. Any k of the n layers stacked on a heap would reveal the secret, but less than k layers do not reveal any information. Contrary to watermarking, there is no host data in visual cryptography. The secret is shared and can be extracted by combining some of the keys (transparent layers). The keys have visual representations.

Let us introduce a basic 2-out-of-2 visual secret sharing (VSS) scheme proposed by Naor and Shamir [8]. Each pixel of a secret binary image is expanded to form two blocks SH_1 and SH_2 (figure 1(a)). A white/black pixel is shared by randomly choosing one of the first/last six rows. Therefore, an $m \times n$ binary image is concealed into two $2m \times 2n$ shares. Regarding the security of this scheme, knowing only one share reveals nothing about its correspondent, and therefore about the secret image. Each 2x2 pixel block could equally correspond to either a white or a black pixel from the secret image. Stacking the shares corresponding to a black pixel results in a 2x2 black block. In the case of a white pixel, half of the reconstructed block are white and the remaining half are black. Even if the stack operation induces a contrast loss of 50%, the secret image is still revealed to human eye without performing any cryptographic computations. Original $m \times n$ secret image can be perfectly recovered using a simple rescaling algorithm from the stacked image (figure 1(b)).

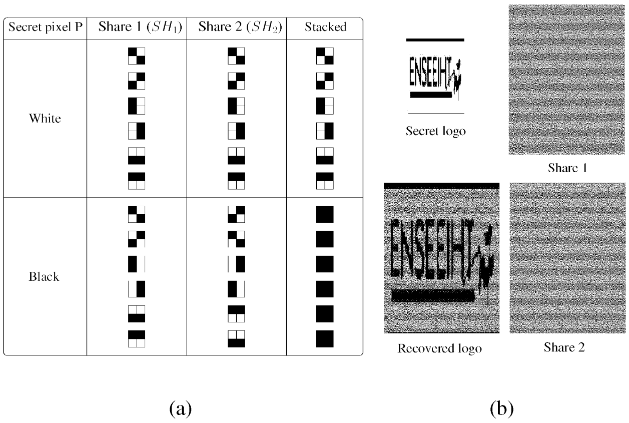


Figure 1. VSS scheme using 2x2 pixels applied to hiding ENSEEIHT logo

2.1 Video Watermarking Method

Houmansadr et al. proposed in [7] a video watermarking method which use the (2,2) VSS scheme presented above.

Embedding First, the video sequence is temporally scrambled. The permutation σ can be obtained using an m-sequence produced by a Linear Feedback Shift Register or more simple, using a modular scrambler. Applying the second method, the frames are temporally scrambled as follows: $\sigma(i) = p \times i \mod L + 1$ where $\sigma(i)$ is the scrambled index of i-th frame, L is the length of the sequence and p an integer number prime relative to L.

Second, the (2,2) VSS scheme is applied to the binary logo to obtain $L/2$ pairs of shares $\{SH_1(k), SH_2(k)\}$, $k \in \{1...[L/2]\}$. The logo and the shares are then transformed from the binary format (0,1) to the signed format (-1/+1), which leads to a zero-mean noise-like share sequence.

Third, the embedding algorithm is performed on the frames of the scrambled video sequence in the following manner: $F_w(i) = F(i) + \alpha SH_j(k)$, $j = i \mod 2$ where $F(i)$ is the i-th frame of the scrambled sequence, $F_w(i)$ is the watermarked frame and α the watermark strength. Finally, the reverse temporal scrambling (i.e. σ^{-1}) is applied to obtain the watermarked video.

Detection First, the watermarked video sequence gets temporally scrambled to place the frames containing correspondent shares adjacently. All frames are then passed through a High Pass Filter (HPF) to strengthen the high frequency spectrum of the watermark sequence (i.e. the shares). After evaluating several HPFs, the authors retained the FFT

(Fast Fourier Transform) filter whose normalized cut-off frequency is 0.31. The filter acts as follows: the FFT transform of the video frame is passed through a masking stage which drops its low-frequency components; then, the inverse FFT transform is applied to obtain the filtered frame.

Second, for $k \in \{1, ...[L/2]\}$ the function *stack* is applied subsequently to filtered frames $HPF(F_w(2k-1))$ and $HPF(F_w(2k))$ yielding $[L/2]$ stacked frames, namely $F_{st}(k)$. The *stack* function retrieves the binary logo from its shares by computing the pixel-by-pixel minimum of the two frames and then retaining the maximum value of each block.It is easy to verify that the stack function performs well by applying it on a row from figure 1(a).

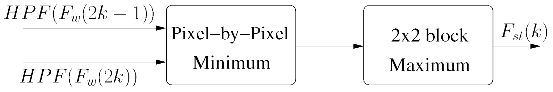


Figure 2. Block diagram of the *stack* function

Third, a measure of correlation to decide whether the video contains the specified logo or not, is necessary. The authors proposed to use the cosine between the stacked frame and the original logo in their $m \times n$ dimensional space:

$$Corr(k) = \frac{F_{st}(k) \cdot Logo}{\sqrt{E(F_{st}(k)) \cdot E(Logo)}}$$

where $Logo$ is the signed logo and $E(X)$ is the energy of X. Finally, an average on these coefficients gives a global correlation measure, namely $Corr$. The closer $Corr$ is to one, the bigger the probability of an watermarked video content. In practice, a threshold Th is chosen and the content is considered watermarked if $Corr \geq Th$. Additionally, other measures such as Linear Correlation Coefficient (LCC) or Bit Error Rate (BER) could be investigated to strengthen this decision.

Observations Our experiments revealed the decrease of correlation coefficient in the case of video sequences containing high-frequency images. This is due to the interference between the shares and the high-frequency spectrum of the frame itself. Inspired by ISS (Improved Spread Spectrum) techniques , we can use the encoder knowledge about the high-frequency spectrum of the signal and enhance detector performance by substracting a part of it, to compensate for the signal interference. Therefore, for each frame F and its corresponding share SH, the equation describing the embedding process becomes: $F_w = F - \beta \cdot HPF(F) + \alpha \cdot SH$ where β is a fidelity parameter, $0 \leq \beta \leq 1$.

We observed that using LCC as a measure of correlation gives better results then the cosine correlation. Additionally, instead of averaging the correlation coefficients, better performance can be obtained by averaging the logo estimations and then computing the correlation with the original logo. In this case, before computing the correlation coefficient, a denoising filter is applied to remove the signal interference. All these observations lead to an increase of approximately 50% in the correlation measure and will be used further.

3 INTRA-VIDEO COLLUSION ATTACKS

The basic idea behind intra-video collusion attacks is the exploitation of the redundancy, either in the host video frames or in the embedded watermark, to estimate the redundant component [2]. Depending on the redundancy, two types of intra-video collusion attacks are possible.

Type I This type of collusion attack exploits the redundancy in the embedded watermark. Due to the imperceptibility constraint, the watermark is embedded in the spatial high-frequency components of the host frames. The difference between a watermarked frame and its spatial low-pass filtered version gives an estimate of the watermark. A refined estimate of the watermark can be obtained by combining the individual estimates obtained from different frames. The estimated watermark is then subtracted from each watermarked frames to get an estimate of the host frames. This attack is known as Watermark Estimation Remodulation (WER) and is effective in visually dissimilar frames embedded with highly correlated watermarks.

Type II This type of collusion attack is possible when visually similar frames are marked with uncorrelated watermarks. Since such watermarks are in the temporal high-frequency band, they can be removed by temporal low-pass filtering the watermarked frames. This attack is known as Frame Temporal Filtering (FTF) attack and is effective in static scenes where uncorrelated watermarks are embedded.

Frame Temporal Averaging A special case is the frame temporal averaging attack (FTA) applied on sliding-window w. In this cas, the attacked frame becomes:

$$\hat{F}_k = \frac{1}{w} \sum_i F_i \text{ , } 0 \leq |i-k| \leq w/2$$

Signal-Coherent Watermarks To sum up, when a frame-by-frame approach is enforced, two major embedding strategies are usually observed: either the same watermark is embedded in all video frames or a different independent watermark is inserted in each video frame. These strategies show their vulnerabilities against collusion attacks and therefore, alternative strategies need to be found.

A basic principle has been enounced in [5] so that intra-video collusion is prevented. The watermarks inserted into two different frames of a video should be as similar, in terms of correlation, as the two frames are similar. In other terms, if two frames are similar, the embedded watermarks should be highly correlated. On the contrary, if two frames are really different, the watermark inserted into those frames should be unalike. **In other terms, the introduced watermark has to be coherent with the host signal**.

An approach to obtain signal-coherent watermarks can be done by using key-dependent image signatures [6]. The goal is to obtain binary strings related with the host content, i.e., image signatures should be as correlated as the associated images. They are used to generate a watermark pattern which change smoothly with the changes in the video signal.

4 MODULATING BY FRAME SIGNATURE

Hash functions are frequently called message digest functions. Their purpose is to extract a fixed-length bit-string from a message (computer file or image) of any length. In cryptography, hash functions are typically used in digital signatures schemes for authentication and integrity purposes. Hash cryptographic functions are "infinitely" sensitive in the sense that a small perturbation of the message will result in a completely different bit-string. In applications involving image authentication, the requirements on what should be a digest of an image are somewhat different. Distortion introduced by lossy compression or typical image processing does not change the visual content of the image and henceforth does not make the image non-trustable. It would be useful to have a mechanism that would return approximately the same bit-string for all similar looking images, yet, at the same time, two completely different images would produce two uncorrelated hash strings. Such a mechanism is provided by visual hash functions.

Visual Hash Functions Fridrich et al. [6] proposed a method based on the following observations: a low-frequency DCT coefficient that has a small or large absolute value cannot be made large or small, respectively, without introducing visible changes in the image. Using a secret key K, N random matrices are generated, with entries uniformly distributed in $[0, 1]$. Then, a low-pass filter is applied to each random matrix to obtain N smooth patterns P_i, $i = \overline{1, N}$. All patterns are then made DC-free by substracting the mean from each of them. The image I is projected on each pattern P_i and the absolute value of the each projection is compared with a threshold Th to obtain N bits b_i: $|I \cdot P_i| < Th \rightarrow b_i = 0$, $|I \cdot P_i| \geq Th \rightarrow b_i = 1$. This method has been tested on real imagery and its robustness was established against typical non-hostile image processing operations such as recoloring, brightness adjustment, filtering, lossy compression or small noise addition. The remainder of this section describes our first improvement based on this method.

Video Preprocess One should note that intra-video collusion attacks could not operate along subsequent scenes without introducing annoying visual artifacts. Therefore, our idea is to split the video into scenes and to prevent collusion inside each scene. Scene changes are detected by applying the histogram difference method. The watermarking logo is split into two shares $\{SH_1, SH_2\}$ (cf. section 2.1) which are then embedded into subsequent scenes.

Embedding The embedding algorithm is depicted in figure 3. Consider a frame F, its corresponding share SH and the secret owner key K. The frame is first decomposed into several blocks whose number is specified as a scheme parameter. The embedding procedure of each block B is the following:

1. Select the corresponding share block BSH.
2. Compute the visual hash of block B, called $h(B, K)$.
3. Spread $h(B, K)$ to fit the size of block BSH.
4. Modulate BSH by the spread $\alpha \cdot h(B, K)$.
5. Add the modulated share block to the original block B.

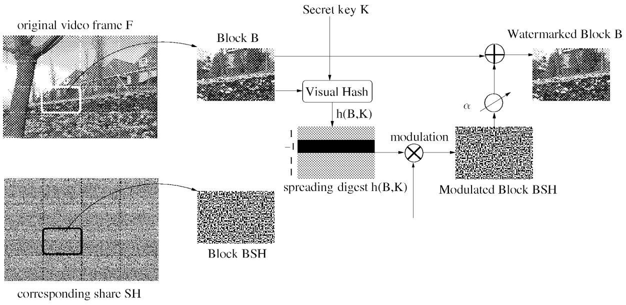


Figure 3. Embedding a share modulated by the frame signature

The last step, not represented in the figure is the compensation of the signal interference as explained in 2.1. Summing up, the resulting watermarked block can be written as follows: $B_w = B - \beta \cdot HPF(B) + \alpha \cdot h(B, K) \otimes BSH$. Finally, the watermarked blocks are merged to obtain the watermarked frame. Inside each scene, we obtain a sequence of watermarked frames whose marks (i.e. the modulated share) evolve smoothly with the dynamics of the scene.

Detection On the detector side, the scenes are rediscovered using the same method described previously. Inside a given scene, the visual hash of each frame is computed in the same block-by-block manner. The frame is passed through the FFT filter described in section 2.1 to obtain a rough estimation of the share embedded in the current scene. These estimations are then averaged to obtain the scene corresponding share. For each pair of subsequent scenes, their shares are stacked and the result is compared to the original logo (using linear correlation) as described in section 2.1.

5 EXPERIMENTAL RESULTS

Type I Collusion Attack In [3], Doërr et al. show that first type of collusion attacks (WER) succeeds only if very similar watermarks are embedded into dissimilar frames (e.g. a dynamic scene). The two methods analyzed here do neither exhibit such a behavior, nor the video content is highly dynamic. Therefore, both methods are expected to survive the WER attack. The average correlations after WER attack are: a) TSS : 0.882 b) MSFS : 0.931.

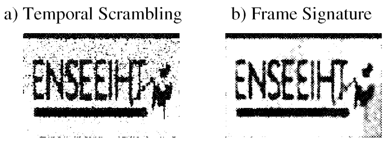


Figure 4. Recovered logos after WER: a) TSS b) MSFS

Type II Collusion In the method proposed in [7], watermarks embedded in neighboring frames are shares from different instantiations of VSS scheme or even the two shares from the same VSS run. Hence, they are uncorrelated and completely independent of the host frames. As shown in section 3, such a blind frame-by-frame strategy with respect to host data is vulnerable against intra-video collusion attacks. We presented in the section 4 an improvement to bypass

FIR Filter	[0 1 0]	[1/6 4/6 1/6]	[1/3 1/3 1/3]	[1/2 0 1/2]
TSS	0.913	0.853	0.432	0.350
MSFS	0.920	0.911	0.898	0.885

Table 1. Average LCC for FTA attack with different FIR filters

this security lack. Along a static scene, similar blocks from subsequent frames will get (almost) the same signature, and therefore the same watermark. Conversely, when two corresponding blocks from two frames are dissimilar, their signatures are completely different and the resulted watermarks are uncorrelated. Henceforth, the signal-coherent watermarks are expected to survive both types of collusion attacks. In order to validate our proposal, we compare the resilience of the two methods described previously against collusion attacks. The methods are abbreviated as follows: **TSS** (Temporal Scrambling Shares) and **MSFS** (Modulating Shares by Frame Signature). Four temporal FIR (Finite Impulse Response) filters were used in FTA collusion attack (table 1). The coefficient of f associated with the current frame F_k defines its weight when applying the filter. The first filter (e.g. [0 1 0]) corresponds to original, non-attacked watermarked video. If 4/6 from the original watermarked frame is conserved in the replacement frame (second filter), the logo survives the attack when TSS method is used. Conversely, in only 1/3 from the watermarked signal is retained, the LCC goes beyond 0.5 and the attack succeeds. MSFS method survives the attack due to its signal-coherence property. The recovered logos for [1/3 1/3 1/3] are depicted in the figure 5.

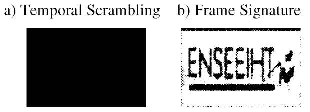


Figure 5. Recovered logos after FTA attack using [1/3 1/3 1/3] filter

Towards motion-aware attacks Frame Temporal Filtering attacks can be applied in static scenes and becomes effective when uncorrelated watermarks are embedded in strongly correlated frames. Nevertheless, in the case of dynamic scenes (moving camera/objects), temporal filtering will cause severe visual degradations to video content. This occurs since filtering operation does not take into account the motion between subsequent frames. Consequently, a more effective, motion-aware attack has to be considered for dynamic scenes.

Non-Hostile Attacks The robustness of the two methods is evaluated under non-hostile video processing operations such as M-JPEG compression, rescaling and noise corruption. The average results are depicted in the following tables.

Quality Factor	90	80	75
TSS	0.850	0.713	0.520
MSFS	0.905	0.827	0.610

Table 2. Average LCC for MJPEG compression

Scaling Ratio	0.5	1.25	2
TSS	0.689	0.886	0.692
MSFS	0.644	0.901	0.712

Table 3. Average LCC for rescaling attack

Noise Amplitude	1	3	5
TSS	0.929	0.910	0.868
MSFS	0.931	0.923	0.910

Table 4. Average LCC for different amplitude noises

6 CONCLUSION

This paper carries out an experimental analysis of a video watermarking method based on visual cryptography. The method starts by temporally scrambling the video sequence using a secret permutation. The watermarks embedded into the scrambled frames are shares of the owner's mark (a visible logo) and the embedding process is done in the spatial domain. Watermarks embedded in neighboring frames are shares that are uncorrelated (since completely independent) with the host frames. Such blind frame-by-frame strategies with respect to host data show vulnerability against intra-video collusion attacks.

We investigate the resilience of this scheme and following the principle of coherent watermarks, propose two alternatives to improve it. The first method uses image signatures to gracefully modify a pattern, i.e. the share, along a video scene. The second strategy, not detailed in this document, embeds logo shares into subsequent mosaic scenes. Finally, experimental results show that our scheme provide better resilience against common hostile and non-hostile attacks.

REFERENCES

[1] G. Doerr and J. Dugelay. A guide tour of video watermarking. *Signal Processing: Image Communication, Special Issue on Technologies for Image Security*, 18(4):263–282, 2003.

[2] G. Doerr and J. Dugelay. Security pitfalls of frame-by-frame approaches to video watermarking. *IEEE Transactions on Signal Processing*, 52(10):2955–2964, 2004.

[3] G. Doerr and J.-L. Dugelay. New intra-video collusion attack using mosaicing. In *Proceedings of the International Conference on Multimedia and Expo*, pages 505–508, 2003.

[4] G. Doerr and J.-L. Dugelay. Collusion issue in video watermarking. In *Security, Steganography, and Watermarking of Multimedia Contents VII. Proceedings of the SPIE.*, volume 5681, pages 685–696, 2005.

[5] G. Doerr, J.-L. Dugelay, and D. Kirovski. On the need for signal-coherent watermarks. *IEEE Transactions on Multimedia*, 8(5):896–904, 2006.

[6] J. Fridrich. Robust hash functions for digital watermarking. In *Proceedings of the The International Conference on Information Technology*, pages 178–183, 2000.

[7] A. Houmansadr and S. Ghaemmaghami. A novel video watermarking method using visual cryptography. In *IEEE International Conference on Engineering of Intelligent Systems*, pages 1–5, 22-23 April 2006.

[8] Moni Naor and Adi Shamir. Visual cryptography. *Lecture Notes in Computer Science*, 950:1–12, 1995.

[9] Hsien-Chu Wu, Chwei-Shyong Tsai, and Shu-Chuan Huang. Colored digital watermarking technology based on visual cryptography. In *Proceedings of the International Conference on Nonlinear Signal and Image Processing*, page 6, 2005.

Interactive augmentation of photographs depicting prehistoric engravings

Christophe Dehais
Vincent Charvillat
Jean Conter

IRIT - ENSEEIHT
2, rue Camichel - 31071 Toulouse - France
email: christophe.dehais@enseeiht.fr

April 5, 2007

KEYWORDS

Augmented Reality, Tangible Human Computer Interaction, Cultural Heritage

ABSTRACT

This paper presents a complete setup that eases the access to prehistoric features in an interactive fashion. High resolution photographs of engraved walls are combined with manual drawings made by experts. To make the interaction very intuitive, images of engraved panels are projected onto an electro-magnetic tracking board that is able to report the position and state of up to three electronic stylus. The user moves a stylus to reveal augmentations interactively by pointing directly the image. At a software level, several problems had to be overcome, from precisely registering line drawings of the engravings made manually by an expert onto the corresponding photographs, to the design of a usable and fault tolerant software. A complete system has been realized and successfully tested with a relatively wide audience, composed in particular of children. This work demonstrates the interest and effectiveness of 2D Augmented Reality in the context of cultural heritage.

1 INTRODUCTION

Public access to prehistoric sites, especially closed ones like caves is often restricted : first some of these sites are hardly accessible without special equipment, and building the infrastructure to allow safer and easier access is not always possible. Then lighting such a site without damaging its features is often challenging, since mural paintings degrade with light exposure and heat. Finally increasing human activity may cause variation of thermal conditions and humidity that can irreversibly damage the walls (for example by causing the development of moisture). However, there is a sustained interest among the public in discovering these sites. In such a context Virtual Reality, Augmented Reality and Mixed Reality applications can prove very efficient in providing an experience close to or even richer than a real visit (Azuma et al. 2001).

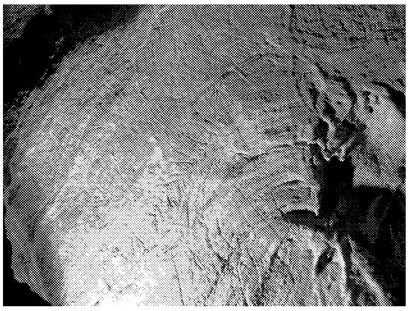


Figure 1: A photograph showing a piece of the incised panel. Recognizing the mammoth shape can be quite difficult for a non expert viewer.

The data and images used in our setup come from the Gargas caves in the French Pyrenées (Gargas 1906), which contain some very rare paintings of stencilled hands and many engraved walls representing animals. The paintings and engravings are estimated to date from the Gravettian Period (from about 27,000 to 22,000 years ago).

The engraved panels are very difficult to appreciate or understand without the help of an expert outlining each shape present in a particular zone. The figure 1 should convince the reader that an undocumented view of such a wall feels indeed very cluttered and hard to appreciate. The fact that multiple meaningful shapes overlap in the same region makes the recognition even more difficult. That means that only providing a reconstructed view (of any form) is not sufficient for presenting such data.

We propose a system that helps a user recognize shapes interactively on an engraved panel. This is made by highlighting the shape in a high resolution photograph of the wall. These augmentations are updated in a natural way as the user guesses the location of the engravings with a pointing device.

We claim that the system actually helps train the eye of the visitor at recognizing such shapes among the clutter. We believe such a system improves the experience of the visitor by providing a complementary approach to the cave's features,

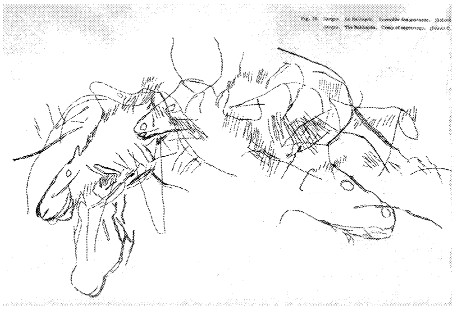


Figure 2: Incised panel drawn by C. Barrière.

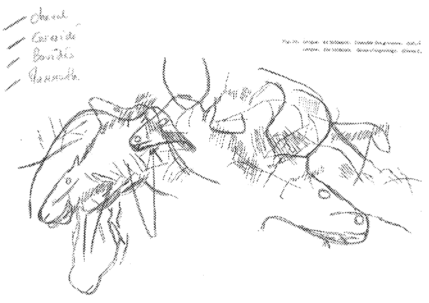


Figure 3: Annotated drawings with highlighted bovids, horses, mammoth etc.

besides of a real visit.
The remaining of this paper is organised as follow: section 2 presents the data we had at hand to realize this project. Section 3 explains our solution to the problem of registering the hand drawn figures into the photographs. Finally section 4 details the hardware and software aspects of our setup and the feedback gathered during a public session.

2 PHOTOGRAPHS AND DRAWINGS AVAILABLE

The Gargas cave organization provided us with several data. First of all, we have been authorized to capture a large set of natural photographs showing a group of engravings under different pose and illumination conditions. The engravings compose an incised panel with a small horse head, a large head of a reindeer, an auroch with its thin but very visible horns, then superimposed a whole mammoth. The place is difficult to access and the panel is very hard to read because the engravings overlap a lot (see figure 1).
In order to help the reading, the drawings made by the prehistorian Professor Claude Barrière (Barrière 1976) were also at our disposal. Figure 2 shows a reprint of those drawings. This figure can be further annotated to provide a meaningful representation and a possible interpretation among several hypotheses (see figure 3).
We performed several preprocessing and processing steps on these resources :

- the prehistorian drawings have been digitalized and enhanced,
- the lines of interest have been segmented and enhanced thanks to the annotations. Therefore individual drawings of bovids, horses and a mammoth can be used,
- the corresponding region of interest inside each photograph has also been extracted by image cropping,
- the non-rigid registration between the obtained natural images and the expert drawings can then be processed as explained in the next section.

3 NON-RIGID REGISTRATION

In this work we use *Radial Basis Mappings* (RBM) to register the prehistorian's drawings within the photographs. Since the data to register are hand drawn and since the 3D shape of wall is unknown, the deformations between the drawing plan and the camera plan can only be modelled by a non-rigid transformation.

3.1 Radial Basis Mappings

A $\mathbb{R}^2 \rightarrow \mathbb{R}$ Radial Basis Function (RBF) f is defined by a basis function ϕ and an $l+3$-vector of coefficients $\mathbf{h}^T = (w_1,...,w_l,\lambda,\mu,\nu)$ and a set of l centres $\mathbf{q_k}$ as :

$$f(\mathbf{x}) = \lambda x + \mu y + \nu + \sum_{k=1}^{l} w_k \phi(||\mathbf{x} - \mathbf{q_k}||) \qquad (1)$$

It consists of a linear part parameterized by (λ,μ,ν) and a non-rigid part, a sum of l weighted terms with coefficients w_k of the basis function applied to the distance of $\mathbf{x}$ and the centres $\mathbf{q_k}$. One of the possible choices for the basis function is the Thin-Plate Spline $\phi(\eta) = \eta^2 log(\eta)$.
We can then construct a $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ RBM m mapping a 2d point $\mathbf{x}$ to a point $m(\mathbf{x})$ defined by two RBF f^x and f^y sharing their centres :

$$m(\mathbf{x}) = \begin{pmatrix} f^x(\mathbf{x}) \\ f^y(\mathbf{x}) \end{pmatrix} = \bar{\mathsf{A}}\mathbf{x} + \mathbf{t} + \sum_{k=1}^{l} \begin{pmatrix} w_k^x \\ w_k^y \end{pmatrix} \phi(||\mathbf{x} - \mathbf{q_k}||) \quad (2)$$

where $\mathsf{A}_{2\times3} = (\bar{\mathsf{A}}_{2\times2}\ \mathbf{t})$ defines an affine transformation represented by 2 rows of 3 parameters (λ^x,μ^x,ν^x), (λ^y,μ^y,ν^y) by generalization of equation 1. When the centres are fixed, the whole set of parameters is encapsulated in an $(l+3)\times 2$ matrix $\mathsf{h} = (\mathbf{h^x}\ \mathbf{h^y})$ partitioned into a non-rigid and a rigid part as $\mathsf{h}^T = (\mathsf{W}^T\ \ \mathsf{A})$.

3.2 Linear estimation from point correspondences

We can estimate a RBM from n matched features (or manual landmarks) : $\tilde{\mathbf{x}}_\mathbf{i} \leftrightarrow \tilde{\mathbf{x}}'_\mathbf{i}$. Given the mammoth image (see figure 1), the correspondences we used are shown in figures 4 and 5. The red crosses sample both the drawing lines in figure 5 and the corresponding visual features in figure 4.
Nevertheless, we do not use each feature location as a deformation centre. Our goal is to automatically select :

- the correct extent of non-rigidity that is needed (in other words the unknown number l of deformation centres),
- a good subset of l features locations $\tilde{\mathbf{x}}_\mathbf{i}$ to be used as the deformation centres $\mathbf{q_k}$.

Such an automatic selection can be performed using a model selection approach (Charvillat and Bartoli 2005) described in the next section. Given the correspondences and the selected deformation centres, we minimize a least-squares transfer error J :

$$J = \frac{1}{n} \sum_{i=1}^{n} \|m(\tilde{\mathbf{x}}_\mathbf{i}) - \tilde{\mathbf{x}}'_\mathbf{i}\|^2. \tag{3}$$

We rewrite J as :

$$J = \frac{1}{n} \left\| \begin{pmatrix} \mathsf{K} & \mathsf{P} \end{pmatrix} \mathsf{h} - \begin{pmatrix} \tilde{\mathbf{x}}'_\mathbf{1} \ldots \tilde{\mathbf{x}}'_\mathbf{n} \end{pmatrix}^T \right\|^2 \tag{4}$$

where the i-th row of $\mathsf{P}_{n\times3}$ is $(1, \tilde{\mathbf{x}}_\mathbf{i}{}^T)$ and the (i,k)-th entry of $\mathsf{K}_{n\times l}$ is $\phi(\|\tilde{\mathbf{x}}_\mathbf{i} - \mathbf{q_k}\|)$.
We must also ensure the boundary conditions (Bookstein 1989) and get a linear least squares problem with linear equality constraints.

$$P_{OLS} \begin{cases} Min & J \\ \mathsf{h} & \\ s.t & \mathsf{P}^T\mathsf{W} = \mathbf{0}_{3\times1} \end{cases} \tag{5}$$

We use direct elimination method (Bjorck 1996) to reduce (by a QR decomposition) the constraint matrix P^T to upper triangular form. The resulting reduced *unconstrained* least squares problem is solved classically by pseudo-inversing.

3.3 Model selection

In this section we describe how to select a good subset of features as deformation centres. A lot of model selection criteria for balancing the residual and the degree of freedom of the model have been proposed in the literature. These criteria use an accuracy criterion and a penalty term which measures the complexity of the model. Most are based on statistical and information-theoretic criteria. Among them, the most widely used criterion is the geometric Akaike's criterion (AIC, (Akaike 1974)) and the minimum description length (MDL) criterion (here we present the gMDL term (Kanatani 2004)):

$$\mathrm{G_{AIC}} = J + 2k\sigma^2 \tag{6}$$

$$\mathrm{G_{gMDL}} = J - k\sigma^2 log(\sigma^2) \tag{7}$$

where k is the number of degrees of freedom (d.o.f.) of the transformation (proportional to l). The noise level σ^2 cannot be simply estimated e.g. from the estimated residuals. In our case since interpolation with null residuals can be reached by increasing the d.o.f., we will use a fixed noise level estimation based on the uncertainty of manual landmarks. We have successfully used these two criteria in our experiments. Their results are visually equivalent.

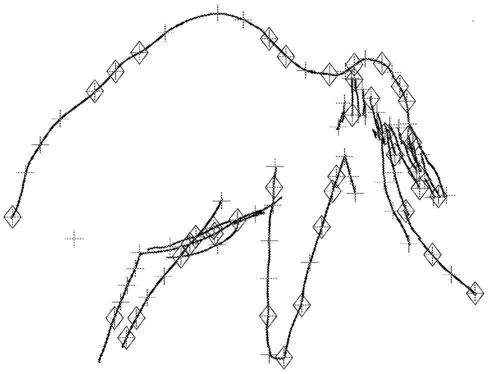


Figure 5: Manual landmarks (red crosses) and selected deformation centres (blue diamonds)

In order to minimize these new criteria, we used a backward stepwise selection principle. We start with the full model (each feature location is used as a deformation centre) and sequentially remove a centre. At the beginning the criterion J is null (no positive residual with an interpolation) and the complexity term is high. At each selection step we remove the centre that leads to the smallest increase of J criterion. The process termination is simple: we stop the backward stepwise selection as soon as our model selection criterion G reaches a local minima.
Figure 5 shows the selected subset of features for the mammoth data set (38 blue diamond markers selected among 91 red crosses). We can notice that the selected deformation centres are correctly spread over the mammoth curves.
At the end of the selection process we have an estimated radial basis mapping (RBM) that can be used to transform the expert drawing (see the second part of figure 4). At this stage the registered graphic resource matches the photograph precisely as shown in the rightmost part of figure 4.
Note that, even though they have been drawn by hand, the reprints based on the work of C. Barrière have proved extremely accurate.

4 SETUP DESCRIPTION

The ability to register precisely the photographic data with the expert drawings allowed us to design a complete visualization system which we describe in this section. The setup should be used by a non expert audience typically composed of visitors of the Gargas cave, including children, which imposes constrains on the usability of the system.

4.1 Interactive board

Tangible interfaces often prove very suitable in a context like the one of this work. The use of a known device for interacting with the system (here a simple stylus and a board) often shorten the learning stage and make users more at ease with the system.
Our system is composed of a tracking board, a video projec-

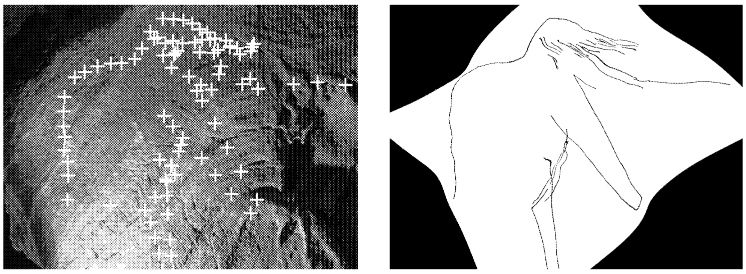


Figure 4: Registration process for the mammoth

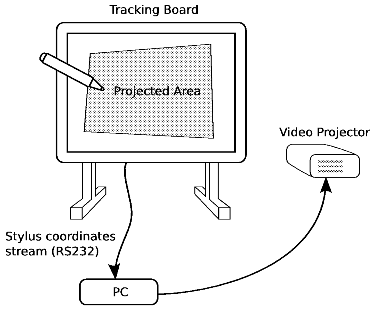


Figure 6: Overview of the different parts of our interactive system.

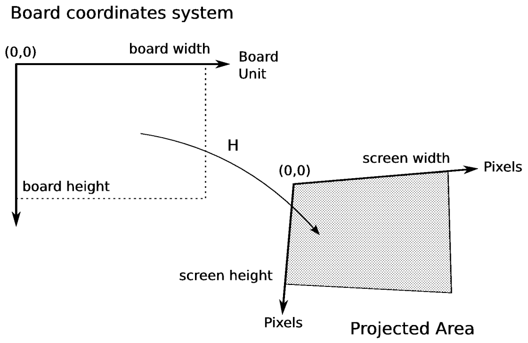


Figure 7: Transformation between board coordinates system and screen coordinates system.

tor, and a standard Linux-based PC. Those elements are very easy to setup together and adapt to different practical setup conditions, as opposed to other augmented and virtual reality systems that use more complex display and interaction devices.

The board is able to report the position of an electronic stylus hold close or onto the surface (see figure 6). The board is also able to distinguish between hovering the stylus at a few centimetres above the surface and pressing the surface with the tip of the stylus. The board transmits its information to the PC using a simple RS232 serial interface. The software uses the stream of positions and reacts accordingly, updating the image projected onto the board.

The projector provides a direct visual feedback onto the tracking board, so the position of the user input has to accurately match the projected image. This requires a calibration step described in the following section.

4.2 Calibration

A key problem to solve in our context is the reprojection of visual information in the precise location of the input. In order to do that we need to precisely map the tracked position of the stylus into the application screen space (being projected onto the board). A very common problem when using a video projector is that the shape of the reprojected area is subject to various deformations, due to misalignment between the projected image plane and the planar surface onto which it projects. It is well known that such deformations can be modeled by a homographic transformation that maps the 2d homogeneous space $\mathcal{P}_2$ to itself (Hartley and Zisserman 2004). In practice this transformation is represented by a 3×3 matrix H, so that a point $p_B = (x_B, y_B)$ in board coordinates is mapped to the point $p_S = (x_S, y_S)$ in screen coordinates this way:

$$\widetilde{p_S} = (\widetilde{x_S}, \widetilde{y_S}, \widetilde{w_S}) = H. \begin{bmatrix} x_B \\ y_B \\ 1 \end{bmatrix}$$

$$p_S = (x_S, y_S) = (\widetilde{x_S}/\widetilde{w_S}, \widetilde{y_S}/\widetilde{w_S})$$

The matrix representing a given homography in unique up to scale factor, so that it has 8 degrees of freedom. Computing this homography involves providing at least 4 correspondences between the source frame and the destination frame. In our application, the calibration procedure simply requires the operator to click the 4 corners of the reprojected area. The system is then calibrated as long as the relative position of the projector with respect to the board is not changed.

Note that we do not try to compensate for other types of deformation, like barrel and pincushion. Instead we rely on the correction provided by the video projector itself and assume an image free of such distortions.

4.3 Software design

The software part of the system is designed as a full screen application that allows the selection of a given augmented panel, composed of its high resolution photograph and a segmented and wrapped expert line drawing as described in section 3. The segmentation of this layer is done by copying the greyscale value of the scanned document to its alpha channel. This way only the lines are opaque and the delineation is naturally anti aliased.
We designed a software that combines in real-time the two layers based on user input (see figure 8 for a synthetic diagram of this process). The stream of stylus positions coming from the board is used to update a footprint mask which is intersected with the alpha channel of the augmentation image in order to reveal only the part drawn over by the user. The accuracy required and thus the level of difficulty can easily be adapted by changing the size of the virtual paintbrush used to fill in the footprint image. This principle has two advantages: first it provides immediate visual feedback to the user that feels like he draws directly onto the photograph. Second it is very fault tolerant and allows safe trial, which is especially suited to children.
The possibility to track alternatively different stylus (each one being assigned a unique ID) allowed the application to be used in a collaborative way. Indeed, the highlighted features can be drawn with different colours corresponding to different stylus/users (see lower right photograph in figure 8).

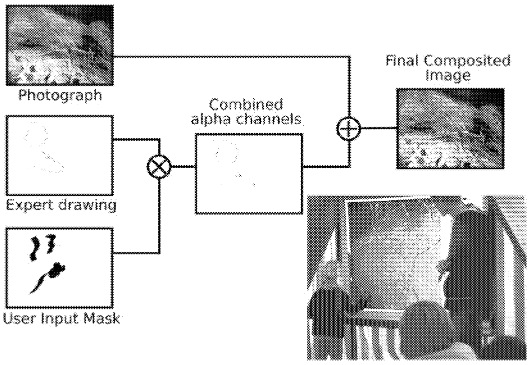


Figure 8: Realtime composition of the photograph and augmentations. Lower right corner: photograph shot during a public session.

The software was easily build thanks to the powerful open-source GUI toolkit Gtk+ (The GIMP Toolkit 1997-2007).

4.4 Public session

A public session took place in late September 2006 in Aventignan, Hautes-Pyrénées, France. Tens of visitors enjoyed using our system that also proved interesting for the watching attenders. The main limitation of our system appeared to come from the projection principle: it would be better to project the image from the back of a transmissive board in order to prevent users from casting their own shadow on the screen. However this problem did not impair the overall usability of the system.

5 CONCLUSION

We presented a complete and usable setup for introducing the public to prehistoric features in an interactive fashion. This was made possible by combining natural images of engraved walls with manual drawings made by experts. This experiment is an effective example of how augmented reality applications can enrich and improve one discovering experience.
The combination of a very accurate registration, the high resolution of the tracking board and photographs lead to a very satisfying interactive experience. The right lines were always revealed at the right location and without latency. We are pursuing this conclusive experiment by extending it to the other artistic treasures of the caves of Gargas, such as the famous stencilled hands.

REFERENCES

H. Akaike. A new look at the statistical model identification. *IEEE Trans. on Aut. Ctrl.*, 19(6), 1974.

Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34–47, 2001.

Claude Barrière. *L'art pariétal de la grotte de Gargas*, volume 1. Oxford Press, 1976.

A. Bjorck. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia (PA), 1996.

Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transaction on PAMI*, 11(6), 1989.

V. Charvillat and A. Bartoli. Feature-based estimation of radial basis mappings for non-rigid registration. In *Vision Modeling and Visualization*, pages 195–200, 2005.

Gargas. Gargas caves information website, 1906. URL www.garga.org.

Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, March 2004.

K. Kanatani. Uncertainty modeling and model selection for geometric inference. *IEEE Trans. on PAMI.*, 26(4), 2004.

The GIMP Toolkit. Gtk+ toolkit homepage, 1997-2007. URL www.gtk.org.

MEDIA DATA COLLECTION

THE USE OF STORYBOARDS IN AUDIO-VISUAL DATA COLLECTION

Z. Yang, X. Wang and L. J. M. Rothkrantz
Faculty of Electrical Engineering, Mathematics and Computer science,
Delft University of Technology
Mekelweg 4, 2628CD Delft, The Netherlands
E-mail: {Z.Yang, L.J.M.Rothkrantz}@ewi.tudelft.nl

KEYWORDS

Storyboards, Aggression detection, Multi-modal communication

ABSTRACT

Aggressive behavior in public places can cause great distress on the part of innocent bystanders. Left untreated, aggressive behavior may escalate and can cause physical aggression or destruction of property, leading to even more frustration and anguish. This paper is a result of data collection work for an ongoing project on aggression detection in trains. The goal in this paper is to gather a realistic audio-visual aggression dataset that can be used to test and evaluate future aggression detection algorithms. The dataset is gathered in a real train with four microphones and four cameras. Semi professional actors are hired to perform aggressive and non-aggressive scenarios.

INTRODUCTION

According to a report by the Dutch Ministry of Transport, Public Works and Water Management (Ministerie van Verkeer en Waterstaat), 20% of all train travellers in 2004 have been the victim of an incident (Ferwerda et al., 2005). The dutch railway company strives decrease this number and thus make public places more save.

The advent of modern technology has created opportunities in this area, as inexpensive sensors have gradually found their way into public transport (e.g. the trains in the Zoetermeer Stadslijn in the Netherlands are already equipped with cameras. The primary role of these cameras is to increase the feeling of security of the passengers and to have a deterring effect on people with bad intentions. Currently, the camera images have to be inspected manually after the incident has occurred and the damage is already done.

The goal of an ongoing project at the Man-Machine Interaction (MMI) group in Delft is to solve this problem by creating a system to automatically detect aggression in a train as it is happening or is about to happen. A first step towards this goal is the collection of a realistic dataset in a real train on which future aggression detection algorithms can be tested. Due to the scarcity of occurrences of aggressive situations in trains and the privacy issues involved when using these recordings, it is difficult/not allowed to use real data. Therefore, we hired semi-professional actors to perform aggressive situations in a train. This paper describes how we used storyboards to achieve this.

The resulting dataset consists of recordings of aggressive as well as non-aggressive scenarios with multiple cameras and multiple microphones. All the recordings were made inside a Benelux train that was provided by the dutch railway company (NS/ProRail). We also had an experienced train conductor to play the part and give advise. The train was standing still while the recordings where made (although the air conditioning system in the train was active and producing background noise).

The remainder of the paper is structured as follows. First we give an overview of the background and the related work in the area, then we present our approach and describe the design of our storyboards. Next we present the experiment setup and describe the data gathered during the experiment. We finish with a discussion and conclusions.

BACKGROUND AND RELATED WORK

With the availability of inexpensive sensors and the ever increasing processing power at our disposal, the number of surveillance and surveillance related research projects have increased. The most commonly used modalities for this purpose are video (Foresti et al., 2005; Cupillard et al., 2004; Javed et al., 2003; Buxton and Gong, 1995; Hosie et al., 1998), audio (Clavel et al., 2005; Härmä et al., 2005; Goldhor, 1993; Pradeep et al., 2006) or a combination of both(Lo et al., 2003; Beal et al., 2002). We observe that in complex surveillance environments, such as in public transport systems, the combination of multiple modalities is more common, e.g. PRISMATICA for railways (Velastin et al., 2002; Lo et al., 2003) and ADVISOR for metro stations (Cupillard et al., 2004).

The usual approach to the surveillance problem is to view the individual events (e.g arm motions, gestures) as related parts of a bigger scenario e.g. fighting, ticket checking. (Cupillard et al., 2004) define a scenario as a combination of states, events or sub scenarios. This means that in the representation of the scenario, the influence of the individual events on the outcome of the scenario is also included e.g. the occurrence of shouting might cause the ticket checking scenario to escalate.

The goal of our surveillance systems is to recognize the scenario based on the recognition of the individual events or features from the input data. More importantly, we want to recognise the position in a scenario (so that we can assess the situation and predict and decide on the appropriate action to take if necessary). Some researchers e.g. (Buxton, 2002; Härmä et al., 2005; Wojek et al., 2006) use the term activity recognition for this.

The basic idea is that for a given domain, specific scenarios (and their probabilities of occuring) are defined a priori. At runtime, the surveillance system tries to infer the consequences of the activity/scene recognized based on this apriori knowledge. Bayesian networks can be used for the inference (Velastin et al., 2002), but other approaches have also been proposed, including multilayered HMMs (Zhang et al., 2006; Wojek et al., 2006) and CHMM (Oliver et al., 2000).

For surveillance/activity recognition in normal environments (e.g. rooms, offices) data can be collected quite easily e.g. (Härmä et al., 2005) collected all interesting acoustic events in an office over a duration of more than two months. Similar research was conducted by (Stauffer and Grimson, 2000).

Our work differs from others by the fact that our system has to work under a more problematic setting (no stable lighting conditions). Moreover, the scenarios we are interested in (ones containing aggression) are somewhat unusual. So it is difficult to set up a controlled experiment for data collection. Under these circumstances most researchers adopt simulations or other scripted approach for data collection. In this paper we present the disadvantages of a scripted approach and focus on audio-visual data collection by means of un-scripted storyboards.

APPROACH

In order to test and to evaluate aggression detection algorithms, recordings of actual aggressive activity in the train is required. Due to the scarcity of these recordings and the privacy issues involved, it is not possible to use real data. Against this backdrop, we hired semi-professional actors to perform aggressive scenarios in a real train. The actors are given scenario descriptions to perform, we used multiple microphones and cameras to record their actions. Apart from the requirement that the setting has to be in a real train and some technical requirements concerning the audio and video recording devices, there are important additional requirements to be imposed on the scenario descriptions:

- Realism, it is important that all the scenarios can really and do really occur in reality.

- Flexibility, we view a recording of an aggressive scenario as just one instance of that scenario. As a result, it is un-desirable to impose too many restrictions in the scenario description (since that would limit the number of instances we can get for a scenario). A scenario description should enable an actor to understand what situation is expected/requested and then allow him the freedom of a personal interpretation.

- Re-usability, in the future, we want to be able to hire other actors to perform the same scenarios again (resulting in another instance of the scenario). This is comparable with emotion recognition for facial expressions where data collection is done by having several people perform the basic emotions.

Given the above restrictions, a scripted approach, in which all actions are specified in advance, has serious limitations on the flexibility of the scenarios. Our proposed storyboards do not have this limitation. In a storyboard the viewer is confronted with a sequence of sketches, each sketch represents a characteristic scene and all the sketches, put in the correct sequence, tell a story. According to the "Cognitive dissonance" theory developed in the field of social psychology (Festinger, 1957), the actors do not need fixed lines to memorize; the aggressive situation will emerge spontaneously as the actors will try to interpret the situation they have seen in the storyboard. One thing we have to make sure of is that the same storyboard will invoke the same interpretation in all actors. In a verifaction step, performed beforehand, a test group of ten students was used verify that each storyboard invokes the same story in the mind of the viewer. To fulfill the realism criteria, we asked a train conductor with 15 years of experience to verify the realism of the scenarios.

DESIGN OF STORYBOARDS

The storyboard model

A storyboard is a sequence of related sketches (pictures, drawing etc.) telling a story. The general structure of our storyboards contain five phases (figure 1).

- Context, in the first frame(s) of the storyboard a general context is presented. The situation, time and location is sketched.

- Problem definition, in the next frames key player(s) are introduced and an indication of the main theme/topic is given i.e. one or more people violating the rules of general accepted correct behavior.

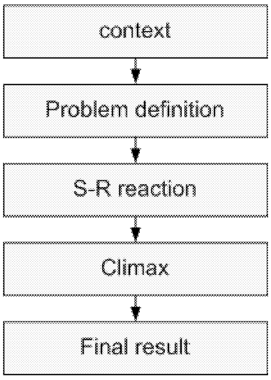


Figure 1: The phases in our storyboard model

- Stimulus-response reaction, the story may continue in increasing violence or decreasing violence. In the first case, the aggressor shows behavior of overt aggression. This can result in more aggression by the aggressor or aggression against the aggressor. We observe a spiral of increasing violence. Otherwise, violence can be neutralized or decreased by counteracts e.g. passengers who inform the aggressor that he is violating the rules. As a result the aggressor adapts his behavior.
- Climax, the S-R behavior results in a climax e.g an explosion of aggression, a peaceful arrangement, a successful repression of aggression, or some actors leave the scene.
- Final result, the last frames show the final result (happy or unhappy end).

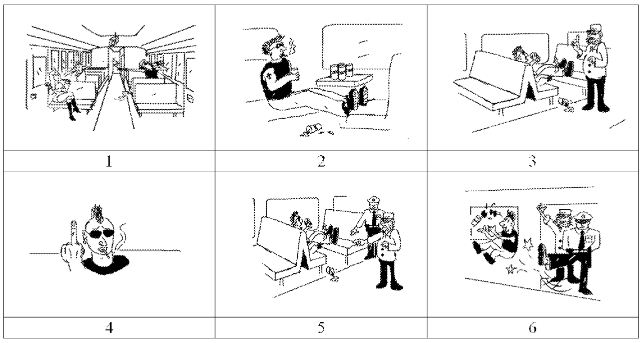


Figure 2: The five phases of our storyboard model: context (1), problem definition (2), stimulus-response reaction (3 and 4), climax (5) and final result (6).

The storyboard creation process

Ingeneral, a storyboard is assumed to tell a story. So, in designing a storyboard a sketch of a story has to be made. But stories can also be based on the fantasy of the storyteller. In case of designing storyboards about aggression in trains it is necessary that the stories are realistic, that is to say based on real life experiences. To ensure this, the process of creating a storyboard is split up into three steps: storyboard design, storyboard verification and storyboard formalisation.

Storyboard design

The process of designing story can be split up into several main steps.

1. The choice of the main theme. At first it has to be decided what will be the dominant theme in the context, e.g. is the story about hooligans, travelers with non adapted behavior, pickpockets, or travelers without a ticket.
2. Once the main theme is chosen the designer has to decide about the main aggressive acts or behavior. Options are hooligans demolish a train, start fighting with other travelers or the conductor.
3. The next step is to decide about a reasonable rising action. The behavior during the rising action should culminate in the aggressive act. So during the rising action there is a sequence of acts with increasing tension, threat of aggression.
4. Next the designer has to design one or more representative scenes for every step. A scene is specified by a sketch.
5. Once the designer has a rough outline of the story, he looks if the story is complete or over complete. He will have to add or delete sketches depending on this.
6. The final step is the refinement of the sketches and creation of the world model. A sketch should express the semantic content of a scene, the main actors should be presented in a prototype way. In figure 2 sketches 2 and 3 for example, we see a hooligan showing non adaptive behavior and a conductor giving a reprimand. Furthermore, the designer should create a world model by specifying the objects and their relationships within the storyboard.

Storyboard verification

Before we can use a storyboard we need to verify that the situation described in it can or did really occur in reality and that the storyboard (i.e. the sketches) will trigger the same image in the mind of an actor. To ensure that each storyboard will be interpreted identically by each actor, we tested each storyboard on a group of ten students. The students were individually asked to write down the story that comes up in their mind once they saw the storyboard. We only used the storyboards that induced the same (and correct) story in all the students. The description of the scenarios of the storyboards that resulted from this process is summarized in (table 1).

Storyboard formalisation

Concerning the relationship with the final goal of aggression detection, a literature survey has shown that the most common approach adopted by scientists is to view individual events (e.g arm motions, gestures) as related parts of a bigger scenario (e.g. fighting, ticket checking). We do not differ from this approach. However, we also include in the representation of the storyboard the influence of events on the outcome of the scenario. This way we have a mechanism to predict the outcome of a scenario based on the evnts witness so far. We use a formal language, first proposed by (Schank and Abelson, 1977) in the Conceptual Dependency theory (CD), to represent such scenarios. An example, showing the formalization of the ticket checking scenario, is shown in figure 3. CD essentially provides a formal way to specify the influence of events on the outcome of a scenario.

Table 1: The list of scenarios for which we created a storyboard

Nr	Description
1	A passenger enters the train, he starts shouting and making problems with other passengers.
2	A group of passengers enter a train yelling and shouting.
3	A beggar traverses the train asking for money.
4	A conductor is checking train tickets. Passengers politely show their ticket
5	Same as 4, but a passenger does not have a ticket and tries to escape.
6	Same as 5, but a passenger does not have a ticket and starts acting aggressive toward the conductor.
7	A passenger is talking through his mobile phone, he is harassed by other passengers
8	A passenger is shouting through his mobile phone. Another passenger kindly asks him to be more quiet. The passenger with the mobile phone obliges.
9	Same as 8, but the situation escalates.
10	A drunkard enters the train compartment.

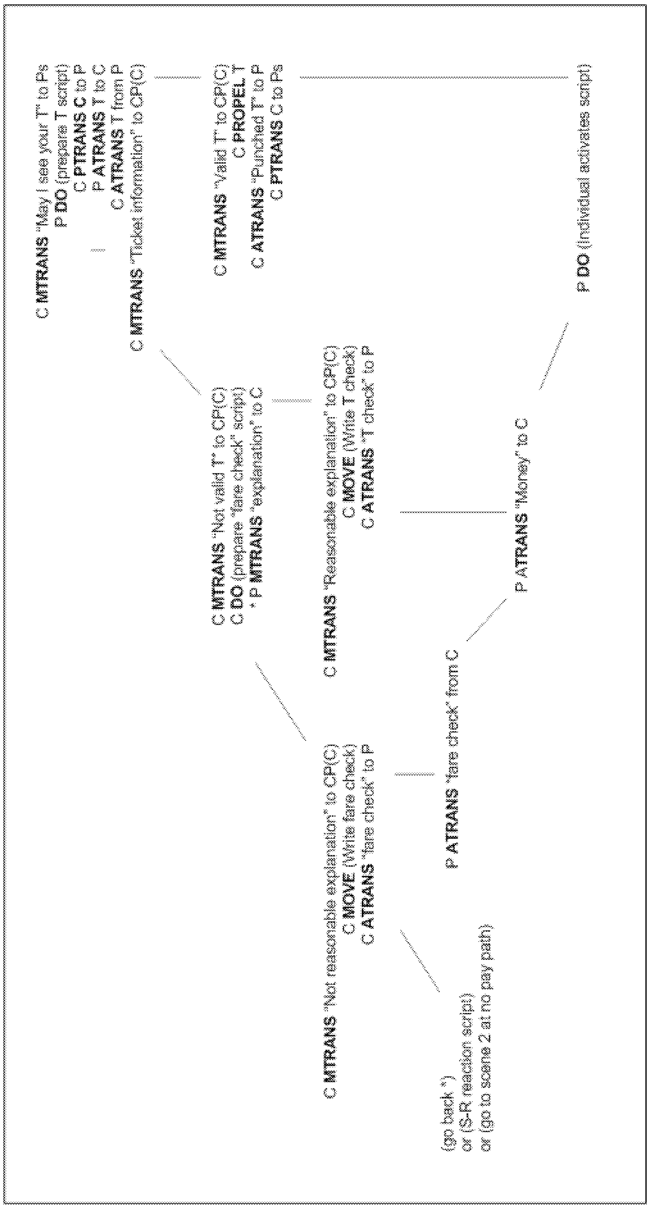

C **MTRANS** "May I see your T" to Ps
P **DO** (prepare T script)
C **PTRANS C** to P
P **ATRANS** T to C
C **ATRANS** T from P
C **MTRANS** "Ticket information" to CP(C)
C **MTRANS** "Valid T" to CP(C)
C **PROPEL** T
C **ATRANS** "Punched T" to P
C **PTRANS** C to Ps
C **MTRANS** "Not valid T" to CP(C)
C **DO** (prepare "fare check" script)
* P **MTRANS** "explanation" to C
C **MTRANS** "Reasonable explanation" to CP(C)
C **MOVE** (Write T check)
C **ATRANS** "T check" to P
C **MTRANS** "Not reasonable explanation" to CP(C)
C **MOVE** (Write fare check)
C **ATRANS** "fare check" to P
P **ATRANS** "fare check" from C
P **ATRANS** "Money" to C
P **DO** (individual activates script)
(go back *)
or (S-R reaction script)
or (go to scene 2 at no pay path)

Figure 3: An example of the ticket checking scenario formalized in CD

DATA COLLECTION

The aim of the data collecting experiment is to gather data that can be used in future aggression detection algorithms. The main problem however is that we do not know what features of the data these future aggression detection algorithms will use. However, we do suspect that the solution will lie in algorithms using audio-visual data fusion. We expect audio data to play a more important role in the future of aggression detection. In fact, microphones have already been installed in Benelux trains (international passenger trains travelling between Belgium, Luxemburg and the Netherlands) and a system for discrete event detection is already in use in these trains. Anticipating more extensive use of the audio data, our dataset consists of recordings with multiple cameras and multiple microphones of aggressive as well as non-aggressive scenarios. All the recordings were made inside a Benelux train that was provided by the dutch railway company (NS/ProRail). We also had an experienced train conductor to play the part and give advise. The recordings were made in a train that was standing still with only the air conditioning system producing background noise.

Experiment setup

The sensor setup used to capture the scenarios specified in the storyboards consists of four cameras and four microphones. The location of the sensors in the train compartment and their orientation is shown in figure 4. Most scenarios were performed in the middle of the train, where the two cameras in the middle have the largest overlap.

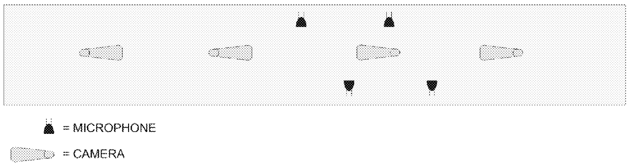


Figure 4: The locations of the sensors seen from a top view of the train compartment. All sensors are attached to the roof of the train.

Camera setup

For activity recognition from video data, occlusion of objects is a problem that needs to be dealt with. Especially in the confined space of the train compartment this is expected to occur frequently. We use multiple cameras mounted at different locations in the train to minimize the risk of occlusion. The cameras also have zooming, panning and tilting capabilities, but these settings were fixed during the recordings.

Microphone setup

Audio based systems for audio event detection are already employed in some trains. However, apart from being able to recognize the type of sound (e.g. screaming, gun shot), it is also important to locate the source of the sound. Several sound localisation algorithms using multiple microphones (Birchfield and Gangishetty, 2005; Brandstein and Silverman, 1997) have been proposed. Our setup includes four microphones providing four synchronised audio streams. In future research we will apply the aforementioned algorithms on the audio data.

Background noise can have a deteriorating effect on the performance of signal processing algorithms. Nevertheless we have to acknowledge its existence, therefore all the recordings were made against a background noise of the trains air conditioning system. Due to logistic reasons we couldn't record in a running train, so the background noise in the recordings is comparable with that of a train standing still in a train station.

Description of the data

The scenarios are recorded in non overlapping sequences and each sequence is marked with a label corresponding to the label of the storyboard. In the end, about one and a half hours of usable audio and video data was recorded. The data contains the scenarios of the storyboards (see table 1) as well as recordings of normal and spontaneous situations. All the data of the sensors is stored in separate streams (four audio stream and four video streams).

The four video cameras captured video at about 13 frames per second, at a resolution of 640x256 pixels. The encoding format is motion jpeg, making it possible to extract the raw jpeg image frames from the entire video. In addition, within each frame the GPS time (with 2 second accuracy) is stored. As the cameras 2 and 3 have the biggest overlapping view, most scenarios were performed in front of these cameras (see figure 5).

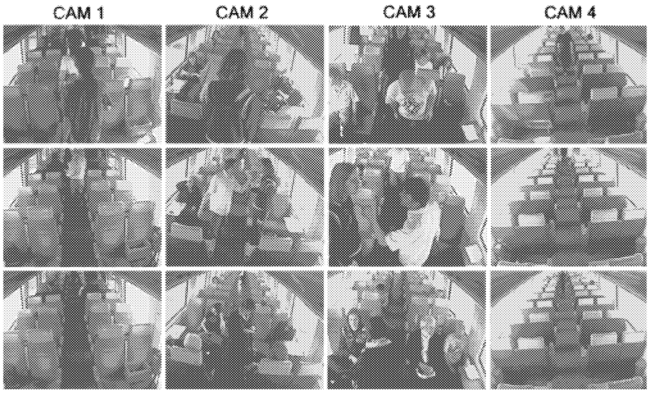


Figure 5: Three scenes as captured by the four cameras, most action was performed before camera 2 and 3 because they have the biggest overlap.

Each microphone captured sound generated by the actors performing the scenarios at a sample rate of 44100Hz with a 24 bit sample size. Each track is synchronized in hardware with sample accuracy. The audio data can be addressed in a single synchronised project consisting of the four streams of the four microphones, or as separate mono audio streams for each individual microphone (figure 6).

PRELIMINARY EXPERIMENTS

To study the aggression recognition capability in humans, we first let humans analyse the recorded videos. As expected, they were all able to identify the aggression in each situation immediately, easily and correctly. However, a movie contains many frames and it is hard and probably unnecessary to analyze every one of them. That is to say, there is a lot of redundancy, because we've seen that storyboards can convey the same information in just a few sketches. Storyboards have the advantage that the designer can add effects to suggest things not possible with raw video frames e.g. movement, gestures etc.

We used a video experiment to seek the number of key frames required from video recordings for a rela-

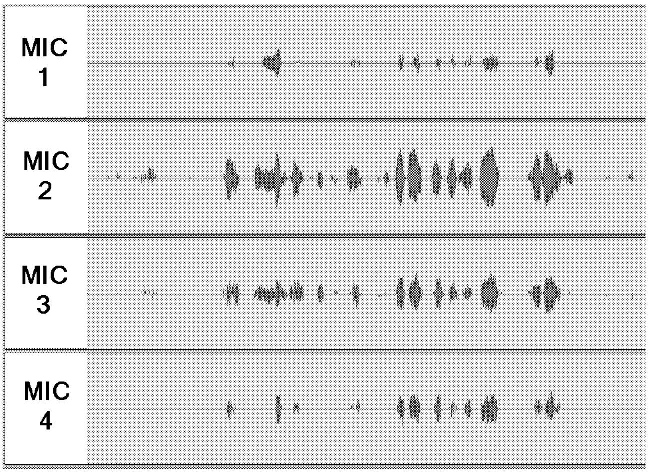


Figure 6: Four waveforms of a shouting scene recorded by the microphones. The waveforms are different in energy yet similar in form, indicating that sound source localisation may be possible with the audio data.

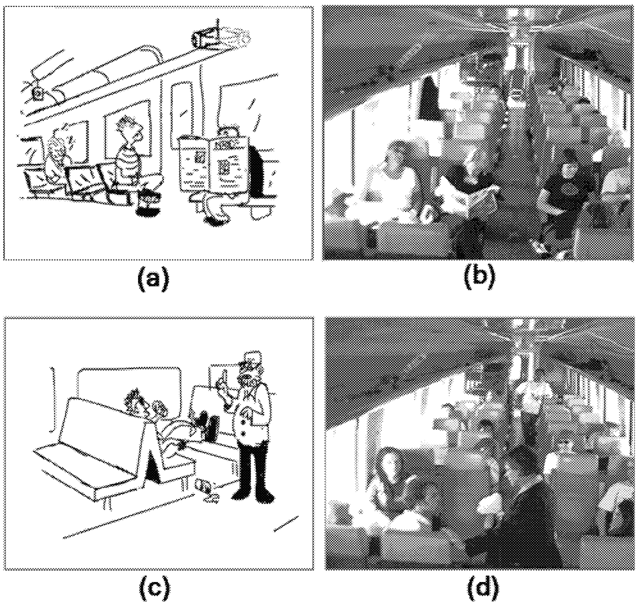


Figure 7: Sometimes a video frame resembles a sketch in the storyboard (a) and (b), sometimes the storyboard designer can add effects that suggest certain situations that are not possible to capture in video frames. (c) suggests that the conductor is giving the passenger a reprimand, not distinguishable in a single frame (d).

tively complete storytelling. From the videos of the 'no ticket' storyboard we selected sample frames to analyse. The storyboard is about a passenger without ticket, it takes around 3 minutes to check this persons ticket. The recording comes from one fixed camera and most frames are describing the interaction between conductor and this passenger.

If we choose the sampling rate of 1/7 frames, 1/15 frames and 1/30 frames the scenario is correctly recognized by our test subjects. With 1/60 frames and 1/120 frames we also get a recognition of aggressive behavior, but however, every story is different. If we look to 1/240 frames, 1/480 frames and 1/960 frames no aggression is detected. Table 2 shows the results.

Table 2: Results of the experiment to seek the number of key frames required from video recordings for a relatively complete storytelling

Sampling	Result
every 7, 15, 30 frames	full recognition.
every 60, 120 frames	recognition of abnormal behaviour, but different stories
every 120 frames or more	no recognition at all.

DISCUSSION AND CONCLUSIONS

To develop and evaluate aggression detection systems, large collections of training and test data are needed (preferably in the environment where the system will be used). While video material such as motion records are necessary for studying temporal dynamics of actions, audio data are also important for obtaining information on activity in the audio modality. In general, audio as well as video is needed for inferring the related meaning (e.g. in terms of scenarios).

The data we collected contains realistic recordings of aggressive as well as non aggressive situations, furthermore the dataset contains artifacts that can occur in practise such as occlusion, time of no actions, sudden explosions of activity etc. In addition, the recordings are in real time, giving an indication of the timing constraints involved e.g. on how long a system has to recognize a scenario or how fast a system has to react in order to be effective. Finally, when we let humans analyse the data, they were all able to identify the aggression in each situation immediately, easily and correctly. Therefore, we believe that our dataset could provide a basis for benchmarks for different efforts in the research on machine analysis of aggression or other topics.

Nevertheless, the dataset has several limitations. First, we instructed the actors to concentrate their activity near the cameras in the middle of the train, since these have the biggest overlapping coverage. In reality, of course, aggression could occur anywhere in a train. For the same reason we positioned the four microphones near the same area. We could have used more microphones to get a broader coverage. Second, we only have one instance of each aggression scenario we defined. Multiple instances of each scenario and different sets of actors are needed to complement the dataset.

As for the storyboards, we analysed the videos and tried to find single frames in it that correspond to each sketch in the storyboard. It turned out that they were very difficult and sometimes even impossible to find. Actors could understand the implied scenario after a quick view of a storyboard, hence storyboards seem to be an efficient way to capture the essence of the scenario. In the future we plan to use the same storyboards to record more instances of the scenarios. We also plan to have test people view the recorded data and annotate the features that triggered the recognition of the aggressive situation in their mind. This hopefully will increase our understanding of aggression detection in humans. All the data gathered in this paper and the corresponding storyboards are available on request.

ACKNOWLEDGEMENTS

This research was done as part of an ongoing project at the TUDelft funded by the MultimediaN project. We like to thank the NS/ProRail for providing us with a train to do recordings in and a train conductor to give advice.

REFERENCES

Beal, M. J., Attias, H., and Jojic, N. 2002. "Audio-Video Sensor Fusion with Probabilistic Graphical Models". In *Proceedings of the 7th European Conference on Computer Vision*, pp. 736–752, London, UK. Springer-Verlag.

Birchfield, S. T. and Gangishetty, R. 2005. "Acoustic Localization by Interaural Level Differences". In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, Vol. 4.

Brandstein, M. S. and Silverman, H. F. 1997. "A Robust Method for Speech Signal Time-Delay Estimation in Reverberant Rooms". In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97)*, Vol. 1, page 375.

Buxton, H. 2002. "Learning and Understanding Dynamic Scene Activity a Review". *Image and Vision Computing*, Vol. 21 pp. 125–136.

Buxton, H. and Gong, S. 1995. "Visual surveillance in a dynamic and uncertain world". *Artificial Intelligence*, Vol. 78 No. 1-2 pp. 431–459.

Clavel, C., Ehrette, T., and Richard, G. 2005. "Events Detection For an Audio-based Surveillance System". In *the IEEE International Conference on Multimedia and Expo (ICME 2005)*, pp. 1306– 1309.

Cupillard, F., Avanzi, A., Brmond, F., and Thonnat, M. 2004. "Video Understanding For Metro Surveillance". In *Proceedings of the IEEE International Conference on Networking, Sensing & Control*, Taipei, Taiwan.

Ferwerda, H., Verhagen, G., and de Bie, E. 2005. "Onderweg naar een veiliger openbaar vervoer 2004". Ministerie van Verkeer en Waterstaat, Adviesdienst Verkeer en Vervoer.

Festinger, L. 1957. *A theory of cognitive dissonance.* Stanford University Press.

Foresti, G., Micheloni, C., Snidaro, L., and Remagnino, P.and Ellis, T. 2005. "Active video-based surveillance system: the low-level image and video processing techniques needed for implementation". *IEEE Signal Processing Magazine*, Vol. 22 No. 2 pp. 25–37.

Goldhor, R. S. 1993. "Recognition of Environmental Sounds". In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP '93)*, Vol. 1, pp. 149–152.

Härmä, A., McKinney, M. F., and Skowronek, J. 2005. "Automatic Surveillance of the Acoustic Activity in our Living Environment". In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2005).*

Hosie, R., Venkatesh, S., and West, G. 1998. "Classifying and Detecting Group Behaviour from Visual Surveillance Data". In *Proceedings of the 14th International Conference on Pattern Recognition*, Vol. 1, pp. 602–604.

Javed, O., Rasheed, Z., Alatas, O., and Shah, M. 2003. "Knight, A Real-time Surveillance System for Multiple Overlapping and Non-overlapping Cameras". In *Proceedings of the International Conference on Multimedia and Expo (ICME 2003).*

Lo, B. P. L., Sun, J., and Velastin, S. A. 2003. "Fusing Visual and Audio Information in a Distributed Intelligent Surveillance System for Public Transport Systems". *ACTA Automatica Sinica: Special Issue on Visual Surveillance, Chinese Academy of Science*, Vol. 29 No. 3 pp. 393–407.

Oliver, N., Rosario, B., and Pentland, A. 2000. "A Bayesian Computer Vision System for Modeling Human Interactions". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22 pp. 831–843.

Pradeep, K. A., Namunu, C. M., and Mohan, S. K. 2006. "Audio Based Event Detection for Multimedia Surveillance". In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICSSP06)*, pp. 813–816.

Schank, R. and Abelson, R. 1977. *Scripts, Plans, Goals and Understanding.* Hillsdale, NJ: Erlbaum.

Stauffer, C. W. and Grimson, E. L. 2000. "Learning Patterns of Activity Using Real-Time Tracking". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22 pp. 747–757.

Velastin, S. A., Maria Alicia Vicencio-Silva, B. L., and Khoudour, L. 2002. "A Distributed Surveillance System For Improving Security In Public Transport Networks". *Special Issue on Remote Surveillance Measurement and Control*, Vol. 35 No. 8 pp. 209–13.

Wojek, C., Nickel, K., and Stiefelhagen, R. 2006. "Activity Recognition and Room-Level Tracking in an Office Environment". In *the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 25–30, Heidelberg, Germany.

Zhang, D., Gatica-Perez, D., Bengio, S., and McCowan, I. 2006. "Modeling Individual and Group Actions in Meetings With Layered HMMs". *IEEE Transactions on Multimedia*, Vol. 8 No. 3 pp. 509–520.

BUILDING A DATA CORPUS FOR AUDIO-VISUAL SPEECH RECOGNITION

Alin G. Chiţu and Leon J.M. Rothkrantz
Man-Machine Interaction Group
Delft University of Technology
Mekelweg 4, 2628CD Delft,
The Netherlands
E-mail: {A.G.Chitu,L.J.M.Rothkrantz}@ewi.tudelft.nl

KEYWORDS
Audio-visual data corpus, lipreading, audio-visual speech recognition.

ABSTRACT

Data corpora are an important part of any audio-visual research. However, the time and effort needed to build a good dataset are very large. Therefore, we argue that the researchers should follow some general guidelines when building a corpus that guarantees that the resulted datasets have common properties. This will give the opportunity to compare the results of different approaches of different research groups even without sharing the same data corpus. In this paper we will formulate the set of guidelines that should always be taken into account when developing an audio-visual data corpus for bi-modal speech recognition. During the process we compare samples from different existing datasets, and give solutions for solving the drawbacks that these datasets suffer. In the end we give a complete list with all the properties of some of the most known data corpora.

INTRODUCTION

Data corpora are an important part of any audio-visual speech recognition research. Having a good data corpus, (i.e. well designed, capturing both general and also particular aspects of a certain process) might be of great help for the researchers in this field as it could greatly influence the research results. However, partly because the field is still young, or partly because the time and resources it takes to record a multi-modal data corpus can be overwhelming, the number of existing multi-modal data corpus is small compared to the number of uni-modal datasets. In order to evaluate the results of different approaches for a certain problem the data corpora should be shared between researchers or otherwise there should be some exact guidelines for building a corpus that all datasets should comply with. In the case when a data corpus is build with the intension to be made public, a greater level of reusability is required. In all cases, the first and probably the most important step in building a data corpus is to carefully state the targeted application(s) of the system that will be trained using the dataset. Currently the main applications of an audio-visual dataset are: audio-visual speech recognition (TULIPS1, AVletters, AVOZES, CUAVE, VidTIMIT, DAVID, IBM LVCSR and DUTAVSC), speaker detection, identity verification (VALID, M2VTS, XM2VTS, VidTIMIT and DAVID), user affective state recognition and talking heads generation.

In the current paper we will focus on the issues related to the audio-visual datasets built having the stated target speech recognition. From the point of view of speech recognition the common limitations that an audio-visual dataset has are:

- The recordings contain only a small number of respondents. This greatly reduces the generality of the results, since it generally generates highly under-trained systems. Hence not all the possible sources of variances are captured. The bias of such datasets comes from the fact that they are unbalanced with respect to gender, race and age of the respondents. Usually the number of respondents is a two digit number, with very few exceptions that use some 200-300 respondents. Even is these situations a good practice is to carefully record the speaker's data, such as age, gender, race, dialect, etc.
- The pool of utterances is usually very limited. The datasets usually contain only isolated words or digits or even only the letters of the alphabet rather than continuous speech. This induces a poor coverage of the set of phonemes and visemes in the language. Therefore, if continuous speech is targeted then the prompts used should always contain phonetically rich words and sentences. A good idea will be to search for the words that efficiently cover the possible combinations of phonemes in the language. This will help keeping the respondent's effort in reasonable limits. Moreover, phonetically rich speech will also better represent the co-articulatory effect in the language.
- The quality of the recordings is often very poor. This usually holds for the video data. It can be argued that for specific applications, such as speech recognition while driving, using dedicated databases (for instance AVICAR database; see Lee 2004) might better represent the specifics of the speech in this situation, but however the use of such dataset will be entirely restricted. We will show in the next sections what the main pitfalls are, and give exact workaround solutions.
- The datasets are not publicly available. Many datasets that are reported in scientific papers are not open to the public. This makes impossible the verification of the results of the different methods, and forces the researchers to build their own dataset.

One of the first datasets used for lipreading was TULIPS1 (Movellan 1995). This database was assembled in 1995 and consists of very short recordings of 12 subjects uttering the first 4 digits in English. Another very small dataset is AVletters (Matthews 1998). Later, other datasets were compiled which are larger and have a greater degree of usability, for instance ValidDB (Fox 2005), AVOZES (Goecke and Millar 2004), CUAVE (Patterson et al. 2002), VidTIMIT (Sanderson and Paliwal 2004) and IBM LVCSR.

In the following sections of the paper we will underline the main issues of the existing datasets for speech recognition with respect to audio and video quality in section 2 and 3, and with respect to language completeness in section 4. The different datasets will be compared during the process. In the comparison we introduce our own dataset DUTAVSC specially built for audio-visual speech recognition. Details about the DUTAVSC corpus can be found in the paper (Wojdeł et al. 2002).

AUDIO QUALITY

The complexity of audio data recording is much smaller than the one of the video recordings. The required hardware was developed long before speech recognition research was born. Therefore all datasets store the audio signal with sufficient high accuracy, namely using a sample rate of 22kHz to 48kHz and a sample size of 16bits. For comparison the audio CDs use a 44kHz sample rate with the same sample size per channel. Therefore the quality of the audio data should not be considered from the point of view of storage accuracy but from the perspective of recording conditions. It is interesting to know what the level of signal to noise ratio (SNR) is allowed during recordings. There could be two approaches here. Firstly, the database can be built with a very narrow application domain in mind such as speech recognition in the car. Also different versions for each possible situation can be recorded (for instance the dataset BANCA (Bailly-Baillire et al. 2003) built for identity verification has three versions for each of the three environments: controlled, degraded and adverse). However, the result is either a too dedicated dataset, or implies a large amount of work for building the dataset. Secondly, the dataset can be recorded in controlled, noise free environment and later on, following the necessities, the noise can be added to the recordings. The specific noise can be simulated or recorded in the required conditions and later superimposed on the clear audio data. An example of such database is NOISEX-92 (Varga and Steeneken 1993). This dataset contains white noise, pink noise, speech babble, factory noise, car interior noise, etc. For our dataset we used the second approach, and used white and pink noise to simulate a noisy environment.

VIDEO QUALITY

In the case of video data recording there are a larger number of important factors that control the success of the resulted data corpus. Hence, not only the environment, but also the equipment used for recording and other settings is actively influencing the final result. The environment where the recordings are made is very important since it can determine the illumination of the scene, and the background of the speakers. A large majority of the datasets were recorded indoors in controlled environment. In these cases the speaker's background was usually mono-chrome so that by using a "color keying" technique the speaker can be placed in different locations inducing in this way some degree of visual noise. In the case of dedicated datasets, as was shown in the previous section, the video data is also characteristic to the environment present at the location where the system is used.

Contrary to the audio case, the equipment used when recording plays a major role. Hence, while Tulips1 and AVletters datasets were compiled at the resolution of 100x75pixels and 80x60pixels, respectively, the newer datasets use much higher resolutions. For instance AVOZES, CUAVE uses 720x576pixels resolution. The same improvement in quality is also observed in the way the color information is sampled. The days of grayscale, 8bits per pixel images are long over. All datasets today save the color information using three channels each having 8 bits size. This is very important because the discriminatory information is highly degraded by converting to grayscale. The frame rate used is usually conforming to one of the color encoding systems used in broadcast television systems. Hence, by the place where the dataset was compiled we can have 24fps, 25fps, 30fps or 29.97fps recordings. Another important quality related property of the device used for recordings is the performance under changing illumination conditions. It is well know that available camcorders perform poorly under low illumination conditions. To alleviate this problem most video recording devices apply algorithms that increase the image intensity, however in chrome image information detriment. Therefore a good illumination is always required when recording. The light should be cast by all means at least uniformly on the scene, not to generate shadow patterns.

Another decision that needs to be made is where to focus, how large should be the scene? Should we define a region of interest (ROI), for instance only show the face of the speaker or maybe only show the lower half of the face, or show more background? Most of the datasets show however a passport like image of the speaker. We argue that defining a small ROI has many advantages and should be considered. Of course a much reduced ROI puts very high constraints on the performances of the video camera used and it might be argued that this is not the case in real life where the resulted system will be used. Recording only the mouth area as is done in the Tulips1 data set is clearly a very tough goal to achieve. However, by using a face detection algorithm combined with a face tracking algorithm we could automatically focus and zoom in on the face of the speaker. A small ROI facilitates acquiring a much greater detailed view of the interesting parts, while keeping the resolution of the frames in regular limits. To exemplify this in figure 1 is shown the area of the mouth as it is retrieved in some of the available datasets. The mouth area is manually clipped such that the bounding box touches the lower and upper lips and the left and right corners of the mouth. The frames were chosen such that to show approximately the same viseme, since is not possible to show exactly the same viseme in each picture. The resulted images were scaled up to a common size. During scaling some distortions appeared due to the fact that the obtained bounding boxes had different aspect ratios. The most visible distortions appeared in the case of the sample from AVOZES dataset. The smallest area reserved to the mouth is found in this example in the VidTIMIT dataset. The contrast of the images in this dataset is also quite poor. In general all datasets that have a large ROI, reserve a very small number of pixels to the mouth area which is however the main source of information for lipreading. In the DUTAVSC dataset, which was compiled in our group, we recorded only the lower part of the face which makes the mouth a central object in the image.

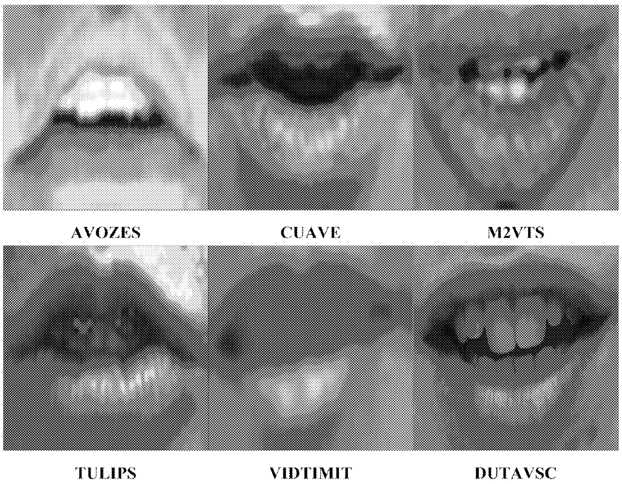


Figure 1: Quality of the ROI in Audio-Visual Data Corpora

Table 1 gives the sizes of the bounding boxes in all 6 samples. We see that the height in the case of DUTAVSC dataset is from 3.5 times to 5.5 times larger than in the case of the other datasets. The same pattern is seen when comparing the width of the bounding boxes, however a smaller difference from 2 times to 4.5 times larger is found.

Table 1: Sizes of the Bounding Boxes Surrounding the Mouth Area in 6 Different Datasets

Corpus	Width	Height
AVOZES	122	24
CUAVE	75	34
M2VTS	46	28
TULIPS1	76	37
VidTIMIT	53	25
DUTAVSC	225	133

During the recordings, attention should be paid to the way the respondents stand and move, especially when a small ROI is considered. If no automatic method is used for tracking the region of interest then the user should be very careful not to go out of the scene and not to move his head very much. Also during talking many speakers use their tongue to wet the lips. By doing so the mouth area will be covered making it impossible to recover correct information about what is being said. This will generate large amounts of noise in the resulted feature vectors. Random movement of the speaker head gives many problems to any method that attempts to extract movement information, for instance methods based on optical flow analysis. If this is the case, then the effect of head movement should be removed prior to feature extraction. The figure 2 shows some examples of broken clips recorded for the DUTAVSC dataset.

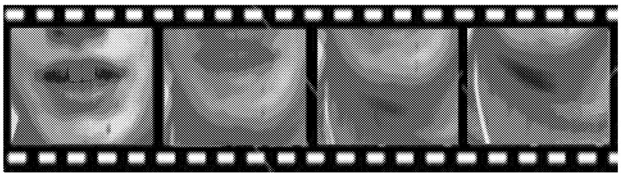


Figure 2: Faulty Clip Section from DUTAVSC Dataset

LANGUAGE QUALITY

As we said in the introduction, the quality of an audio-visual data corpus that has as target application speech recognition systems is measured by the degree of coverage of the phonemes and visemes of the targeted language. The number of phonemes differs from language to language as one may expect. For English there are 40 to 45 different phonemes depending on the dialect, for instance in Australian English there are 44 phonemes (Goecke 2005). In Dutch there are 42 different phonemes. In order to build a reliable system a good coverage of the phonemes is strongly required. For this purpose the utterances should contain phonetically rich words and sentences. The same goes for the visemes, which are the visual counterpart of phonemes. However the number of visemes is slightly smaller, for instance in English are 11 to 14 different visemes, while in Dutch 16. The words used should provide a good coverage of the combination of sounds, such as consonant vowel mixtures, so that all co-articulatory effects appear in a reasonable number of samples. Spelling samples should also be included in the dataset.

CONCLUSIONS

The quality of the data corpus used has a great impact on the results of the research initiative. For this reason, a good data corpus should be build following some strict rules that can guarantee the success of the final product. In this paper we emphasized the most important properties that a good data corpus should have, showed the main fallbacks of the current data corpora and give solutions that can alleviate the problems.

Table 2 lists the most well known data corpora to date and gives for each corpus the main characteristics with respect to their quality. The structure of the table is build such that to emphasize the quality of the datasets with respect to video, audio and language. In all cases only the gender of the speaker is recorded. We can see that our dataset scores quite reasonable.

Another aspect that was not covered until know by any of the data corpus is the frame rate at which a data set is recorded. All databases were recorder using low frame rates in conformity with the standard used at the location where the dataset was compiled. An interesting question is how is the frame rate of the video influencing the performances of the speech recognition system? When fusing the audio and video data for lipreading, a well know problem is the difference in the sample rate of the two data streams. Hence the audio stream is usually sampled at 100fps while the video stream will only provide 24-30fps. To solve this problem some up-sampling techniques need to be used. But what if the video data is recorded using a high speed camera so that the frame rate matches the audio frame rate? In order to tackle this question we plan to build such a database in the near future.

Table 2: Comparison Among Existing Corpora; Their Characteristics and Stated Purpose

Corpus	Language	Sessions	Respondents	Audio Quality	Video Quality	Language Quality	Stated purpose
TULIPS1	English	1	7male, 5female	11.1kHz, 8bits controlled audio	100x75, 8bit, 30fps mouth region	first 4 digits in English	small vocabulary isolated words recognition
AVletters	English	1	5male, 5female	22kHz, 16bits controlled audio	80x60, 8buts, 25fps mouth region	the English alphabet	spelling English alphabet
AVOZES	English	1	10male, 10female	48kHz, 16bits controlled audio	720x480, 24bits, 29.97fps entire face, stereo view	digits from ‘0’ to ‘9’ continuous speech application driven utterances	continuous speech recognition for Australian English
CUAVE	English	1	19male, 17female	44kHz, 16bits controlled audio	720x480, 24bits 29.970fps passport view	7,000 utterances connected and isolated digits	continuous speech recognition
Vid-TIMIT	English	3	24male, 19female	32kHz, 16bits controlled audio	512x384, 24bits, 25fps upper body	TIMIT corpus 10 sentences per person	automatic lipreading, face recognition
DAVID	English	12	132male, 126female (in 4 groups)	--	entire face, upper body, profile view multi corpora: controlled and degraded background, highlighted lips	vowel – consonants alternation, English digits	speech or person recognition
IBM LVCSR*	English	1	290 Unknown gender	22kHz, 16bits --	--	connected digits isolated words	audio-visual speech recognition
AVICAR	English	5	50male, 50female	48kHz, 16bits, 8channels 5 levels of noise car specific	4 cameras from different angles, passport view car environment	isolated digits, isolated letters, connected digits, TIMIT sentences	speech recognition in a car environment
DUTAVSC	Dutch	10-14	7male, 1female	48kHz, 16bits, controlled audio	384x288, 24bits, 25fps lower face view	spelling, connected digits, application driven utterances, POLYPHONE corpus**	audio-visual speech recognition, lipreading

* Not available to the public
** Data corpus for Dutch. Recordings are made over phone lines. More details can be found in (Damhuis et al. 1994)

REFERENCES

Bailly-Baillire, E.; Bengio, S.; Bimbot, F.; Hamouz, M.; Kittler, J.; Mariéthoz, J.; Matas, J.; Messer, K.; Popovici, V.; Porée, F.; Ruiz, B. and Thiran, J. 2003. "The BANCA Database and Evaluation Protocol" In *Proceedings of Audio and Video Based Biometric Person Authentication,* (Springer Berlin / Heidelberg, *2688*, pp. 625-638)

Damhuis, M.; Boogaart, T.; Veld, C. In't; Versteijlen, M.; Schelvis, W.; Bos, L. and Boves, L. 1994. "Creation and analysis of the Dutch polyphone corpus", In *ICSLP-1994,* (pp. 1803-1806)

Fox, N.A. 2005. "Audio and Video Based Person Identification", PhD Thesis at *Department of Electronic and Electrical Engineering Faculty of Engineering and Architecture University College Dublin*

Goecke, R. and Millar, J. 2004. "The Audio-Video Australian English Speech Data Corpus AVOZES" *Proceedings of the 8th International Conference on Spoken Language Processing* (ICSLP2004, vol. III, 2525-2528)

Goecke, R. 2005. "Current Trends in Joint Audio-Video Signal Processing: A Review" In *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications,* (August 28-31, pp. 70-73)

Lee, B.; Hasegawa-Johnson, M.; Goudeseune, C.; Kamdar, S.; Borys, S.; Liu, M. and Huang, T. 2004. "AVICAR: Audio-Visual Speech Corpus in a Car Environment" In *Proceedings of International Conference on Spoken Language Processing – INTERSPEECH2004*, (Jeju Island, Korea, October 4-8)

Matthews, I. 1998. "Features for Audio-Visual Speech Recognition" PhD thesis, School of Information Systems, University of East Anglia, October

Messer, K.; Matas, J. and Kittler, J. 1998. "Acquisition of a large database for biometric identity verification" In *BIOSIGNAL 98,* Vutium Press, (pp 70-72)

Movellan, J.R. 1995. "Visual Speech Recognition with Stochastic Networks" In *Advances in Neural Information Processing Systems, MIT Press*

Patterson, E.; Gurbuz, S.; Tufekci, Z. & Gowdy, J. 2002. "CUAVE: A New Audio-Visual Database for Multimodal Human-Computer Interface Research" In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*

Sanderson, C. and Paliwal K.K. 2004. "Identity Verification Using Speech and Face Information." In *Digital Signal Processing* (vol. 14 nr. 5 pp. 449-480)

Wojdeł, J.C.; Wiggers, P. and Rothkrantz, L.J.M. 2002. "An audio-visual corpus for multimodal speech recognition in Dutch language" In *Proceedings of the International Conference on Spoken Language Processing (ICSLP2002)* (Denver CO, USA, September, pp. 1917-1920)

Varga, A. and Steeneken, H. 1993. "Assessment for automatic speech recognition II: NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems." *Speech Communication*, (vol. 12, no. 3, pp. 247-251, July)

AUTHORS BIOGRAPHY

ALIN GAVRIL CHIȚU was born on November 8, 1978 in Bușteni, Romania. He graduated in 2001 at the Faculty of Mathematics and Computer Science at University of Bucharest, which is one of the top universities in Romania. In 2003 he received the MSc. degree in applied computer science at the same university. Starting September 2003 he joined the Risk and Environmental Master Program at Delft University of Technology, Delft, The Netherlands which he graduated with honors in August 2005. Since then he is pursuing his PhD degree in the Man-Machine Interaction Group, Mediamatics Department at Delft University of Technology under the supervision of Dr. Leon J.M. Rothkrantz. His main interest is in data fusion as the means to build robust and reliable systems, audio-visual speech recognition being one of the case studies. He is also interested in robust computer vision, machine learning and computer graphics.
Email: a.g.chitu@ewi.tudelft.nl
Web address: http://mmi.tudelft.nl/~alin

LEON J.M. ROTHKRANTZ received the MSc. degree in mathematics from the University of Utrecht, Utrecht, The Netherlands, in 1971, the Ph.D. degree in mathematics from the University of Amsterdam, Amsterdam, The Netherlands, in 1980, and the MSc. degree in psychology from the University of Leiden, Leiden, The Netherlands, in 1990. He is currently an Associate Professor with the Man-Machine Interaction Group, Mediamatics Department, Delft University of Technology, Delft, The Netherlands, since 1992. His current research focuses on a wide range of the related issues, including lip reading, speech recognition and synthesis, facial expression analysis and synthesis, multimodal information fusion, natural dialogue management, and human affective feedback recognition. The long-range goal of his research is the design and development of natural, context-aware, multimodal man–machine interfaces. Drs. Dr. Rothkrantz is a member of the Program Committee for EUROSIS.
Email: l.j.m.rothkrantz@ewi.tudelft.nl
Web address: http://mmi.tudelft.nl/~leon

UBIQUITOUS COMPUTING

A Proposal of a Sensor-handling Mechanism Using a P2P Agent Platform for Ubiquitous Environment

Yoshimasa Ishi[1], Yuuichi Teranishi[1], Kaname Harumoto[2], Shinji Shimojo[1]
[1] Cybermedia Center, Osaka University
5-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan
ishi.yoshimasa@ais.cmc.osaka-u.ac.jp
{teranisi,shimojo}@cmc.osaka-u.ac.jp
[2] Graduate School of Engineering, Osaka University
2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
harumoto@eng.osaka-u.ac.jp

ABSTRACT

To realize context-aware services which change their contents dynamically according to user's situations in the ubiquitous environment, making use of real-world information obtained from sensors widely spread is important issue. However, dynamic discovery and utilization of sensors that are established by voluntary users or organizations have not been considered enough. In this paper, we propose a novel P2P-based sensor handling mechanism based on P2P agent platform 'PIAX'. Our mechanism enables grass-root sensor cooperation and utilization. Moreover, we describe an appliance device called 'PIAX nano', which enables sensor providers to set up easily and join to the PIAX network instantly.

KEYWORDS

Ubiquitous computing, Sensors, Mobile Agent, Peer-to-Peer

I. INTRODUCTION

Recently, portable computing devices have become very small and powerful, giving their users access to a variety of services in their personalized forms. Moreover, technological advances of the mobile network enable such computing devices to connect to the network regardless of the locations.

In such so-called "ubiquitous environment", personalized services are required to be provided according to the user's fresh, real-time situation. Therefore, many researches have been done on effective method to achieve such services that follows real-time situation obtained by distributed sensors.

Former researches have been focusing on the sensor data handling method to collect distributed sensor information to the centralized server. Hence existing sensor utilization and service composition method assume such centralized data.

However, it requires enormous cost to set up sensors to cover all over the world. It is hard to construct such huge sensor network by one organization. We believe there is a possibility to construct high-level, intelligent services based on wide-area sensors at low-cost if each individual or organization set up their own sensor, provide information voluntary, and utilize them mutually. Our goal is to realize high-level, intelligent service by connecting such grass-root based sensors and cooperate with each other, selecting appropriate sensors according to their behaviors and availabilities. For example, a user can obtain a weather map by combining temperature sensors embedded in the air conditioner located at each home. Also a user can search nearest restaurants with real-time crowdedness information by utilizing person number sensors embedded in each restaurant. In such environment, assumptions to collect distributed sensor information to one centralized server are not realistic. It is very hard to keep scalability and extensibility of the service.

In this paper, we propose a novel P2P-based sensor handling mechanism based on a P2P agent platform 'PIAX', which we have been developing as a ubiquitous application framework. Our mechanism enables grass-root sensor cooperation described above, without a centralized server. An appliance device called 'PIAX nano', which can handle sensors with USB and serial interface to connect to the PIAX network without complicated settings, is also described in this paper.

II. SENSOR MECHANISM FOR UBIQUITOUS ENVIRONMENT

Requirement

To treat grass-root based sensors and their automatic cooperation, followings are required. We assume these sensors are already connected to the network. Hence we assume ad-hoc mobile network already exists for the mobile sensors.

- Transparent access
  From the application developer's point of view, sensors should be able to treat as just an active information source, e.g. a sensor object. Application should be able to connect and incorporate them easily. However, applications cannot assume sensors are always online. They easily disappear from network and appear again. Moreover, application cannot know what kinds of sensors are available on the network. Applications should be able to access to the sensors regardless of their availability and difference of the behavior. Hence transparent access to the sensors is required.

- Extensibility
  From the sensor resource provider's point view, sensor extensibility is required. Sensors should be free from establishment restrictions. Sensor providers should be able to add sensors at anytime. They can freely remove sensors, change their locations, and upgrade its behavior. New kind of sensors should be added by a sensor vendor without announcement.
- Easy set up
  We assume anyone can attend to the service as a sensor resource provider. For example, housewives, restaurant managers can establish a sensor at their home or restaurant. We cannot assume that resource provider is always a professional of computers, networks and related technologies. Hence sensors are required to be set up easily at low-cost.

Related Works

Sensor network researches mainly focused on a issues of the sensor network itself, e.g. message routing, self-organization and MAC layer control, and so on.[1] On the other hand, OSNAP[2], LON have been proposed as a common protocol to access a sensor network. However, it is difficult to use a sensor on these sensor networks because they don't prescribe how to discover sensors on a network. A service in Ubiquitous Home[3] is a example of context-aware service using sensors. In the paper [4], a mechanism for aiming at sharing sensing data by connecting sensors using JXTA is introduced. However these researches assume that the kinds of available sensors are known for the application developers. Therefore, they cannot support dynamic, grass-roots sensor establishments by users.

III. SENSOR-HANDLING MECHANISM ON PIAX

P2P Agent Platform 'PIAX'

PIAX is a Java based distributed agent platform which we have been developing. It is composed of two layers. Lower layer is P2P overlay network which connects between peers and make up PIAX network. Upper is mobile agent layer that realizes discovery and cooperation of services based on the relations between information and geographical position of peers.

PIAX provides scalability and flexibility to the applications by merging resource discovery functions of the P2P overlay network and flexible functionalities that mobile agents offer. Moreover, PIAX supports multi overlay network that merges two or more overlay networks properly. Currently it supports LL-net[5] for location-based search and DHT for index-based search.

Discovery Messaging, which is a unique function of PIAX, realizes to send a dynamic procedure call with a discovery message to the P2P network. By this functionality, applications can send process request to the mobile agents that satisfy the condition distributed on the P2P network. It enables dynamic distributed processing for ubiquitous environment.

About details, refer to paper[6] or Web site[7].

PIAX-based Sensor Handling

We propose a novel sensor handling mechanism on PIAX, by utilizing its *Discovery Messaging* functionality.

Kind of Peers

In our sensor handling mechanism, following kind of peers exist. All peers are connected to each other by the PIAX's P2P overlay network.

- Repository Peer
  Repository Peer has *Repository Agents* and *Driver Agents*. We assume this kind of peer is established by the vendors of the sensor devices.
- Sensor Peer
  Sensor Peer has one or more *Driver Agents* to manage sensor devices attached the peer. *Sensor Peer* provides sensing data from sensors to the other peers.*Sensor Peer* also monitors sensor status, i.e. attached and detached.
- Application Peer
  Application Peer has one or more *Application Agent* which receives sensing data from sensors.

Kind of Agents

In PIAX, each peer can hold one or more mobile agents. There are three kind of mobile agents in our sensor handling mechanism.

- Repository Agent
  Repository Agent manages *Driver Agents* on the *Repository Peer*. It registers *Driver Agents* to the DHT of the PIAX. *Repository Agent* also creates a clone of *Driver Agent* and dispatch it to the requester if a *Sensor Peer* requests a *Driver Agent*.
- Driver Agent
  Driver Agent hides detailed behavior of a sensor, as same way as a device driver in the OS does. It provides sensing data to the other agents. *Driver Agents* have common methods so that *Application Agents* can easily handle.
- Application Agent
  Application Agents provide services by using sensing data from sensor agent.

Operation Model

Driver Agent registration phase

The *Repository Agents* register its ID and corresponding *Driver Agents*' ID to the DHT when the peer joins to the PIAX network or there is an instruction from the other agents (fig.1 upper part). At this time, sensor ID, which is embedded in the sensor device, is used as a hash key of the DHT.

Driver Agent delivery phase

When a *Sensor Peer* detects attached sensor device or it joins to the PIAX network, it queries DHT to obtain corresponding *Driver Agent* with a sensor ID as a hash key. As a response, a *Sensor Peer* gets a corresponding *Driver Agent's* ID and a *Repository Agent's* ID which manages the *Driver Agent*. Then the *Sensor Peer* sends a query to the *Repository Agent* to obtain

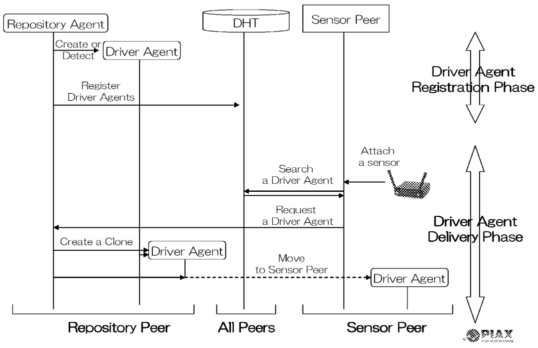


Fig. 1. Driver Agent register/delivery phase

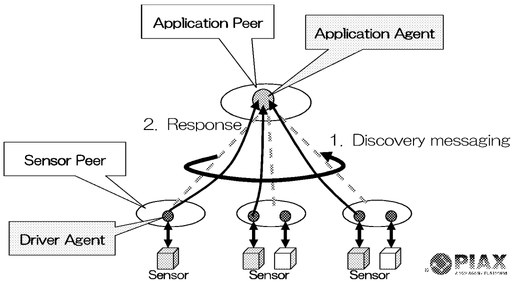


Fig. 2. Sensor utilizing phase

a corresponding *Driver Agent*. The *Repository Agent* which received the query creates a clone of the specified *Driver Agent* and moves it to the requester peer.

The *Driver Agent* which arrived at *Sensor Peer* initializes the sensor, and waits until a request from an application agent is received (fig.1 lower part).

This operation does not require any manual configuration.

Sensor utilizing phase

An *Application Agent* can dynamically obtain data from *Sensor Agents* by *Discovery Messaging* function of PIAX. At this time, applications do not worry about the availability of the sensors.

An application can continue sending messages to the discovered *Sensor Agents* after *Discovery Messaging* if it requires additional processes to communicate with discovered sensor (fig.2).

PIAX nano

'PIAX nano'(fig 3) is a small appliance we have been developing to enable sensor providers to connect sensors to the PIAX network easily. PIAX nano has USB ports and serial interfaces so that it can connect sensor devices. PIAX nano has following features.

- Easy IP network set up and sensor configurations through Web browser.
- Low power consumption due to ARM architecture.
- High performance Java processing by ARM native code generated by GCJ compiler.

PIAX nano itself is actually a Linux-box, which has optimized kernel for PIAX and sensor interfaces.

TABLE I
PIAX nano Spec (prototype)

CPU	EP9307 200MHz (Core ARM920T)
Memory	64MB SDRAM
Storage	128MB (Internal Flash Memory)
Interface	USB2.0 × 2, Serial Port × 2
Network	100Base-TX (PoE Support)
Size	W83.0 x H58.0 x D24.0 [mm]

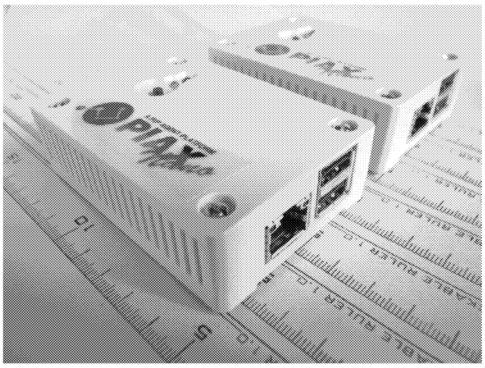


Fig. 3. PIAX nano (Prototype)

IV. CONCLUDING REMARKS

In this paper, we proposed a P2P-based sensor handling mechanism that enables grass-root sensor cooperation based on a P2P agent platform 'PIAX'. An appliance device called 'PIAX nano', which can handle sensors with USB and serial sensors easily, is also described.

We must consider a standardization of a sensing data format to exchange between agents and a profile format of a sensor devices to achieve zero-configuration of sensors.

We are planning to evaluate 'PIAX nano' through the practical experiment and improve it as a more effective, useful sensor handling device.

V. ACKNOWLEDGMENT

This work was supported in part by the Research and Development Program of the Ubiquitous Network Authentication and Agent (2006), of the Ministry of Internal Affairs and Communications (MIC), Japan.

References

[1] C Chong, S Kumar, B Hamilton. Sensor networks: evolution, opportunities, and challenges. *Proceedings of the IEEE*, 91(8):1247–1256, 2003.

[2] A Hasegawa, K Ishibashi, J Ichimura, et al. Open sensor network archtechture to integrate sensor networks. In *Proceedings of the 2006 IEICE General Conference*, volume BS-2-5, March 2006.

[3] T Fujii, H Ueda, M Mino. An implementation of the looking for something service in home ubiquitous environment. In *Proceedings of the 67th IPSJ General Conference*, volume 3, pages 3–369 – 3–370, March 2005.

[4] S Krco, D Cleary, D Parker. P2P Mobile Sensor Networks. *Proceedings of the 38th Hawaii International Conference on System Sciences-2005.*

[5] Y Kaneko, K Harumoto, S Fukumura, et al. A location-based peer-to-peer network in a ubiquitous environment. *IPSJ Transactions on Databases(TOD)*, 46:1–15, December 2005.

[6] M Yoshida, Y Teranishi, K Harumoto, et al. PIAX: A P2P platform for integration of multi-overlay and distributed agent mechanisms. In *IPSJ SIG Technical Repors*, volume 2006-DPS-128, pages 43–48, September 2006.

[7] PIAX.org. available at `http://www.piax.org/`.

PERSONALIZED ADAPTIVE PDA INTERFACE[1]

Siska Fitrianie
Leon J.M. Rothkrantz

Man-Machine-Interaction Group, Delft University of Technology
Mekelweg 4 2628CD Delft,
The Netherlands,
E-mail: {s.fitrianie, l.j.m.rothkrantz}@ewi.tudelft.nl

KEYWORDS
Adaptive (Graphical User) Interface, human-computer interaction, text prediction, text entry for PDA.

ABSTRACT

User demands for usability in mobile context due to the small size of personal data assistants (PDAs) challenge traditional input design. An on-screen keyboard that offers an easier and faster method of entering text with a pen on PDAs, has been developed. We have developed a method for adapting its predictive ability according to user's personal word usage, input context and syntax rules. Frequently used characters are presented to the users in different key sizes and color contrasts according to their relative probabilities to aid visual searching. For this purpose, an experiment has been conducted on which and how to use (user's) data source for faster prediction. In this experiment, we compared four dictionaries recorded from the British National Corpus, personal documents, chat logs and personal e-mails. The experimental results show ways to improve the performance of the word prediction and the language coverage of the word completion.

INTRODUCTION

The needs of being able to access information anytime and anywhere makes personal digital assistants (PDAs) more popular due to its portability and facility for wireless connection. The PDAs are now designed to be smaller and sleeker. They are advancing to a more powerful device and equipped with increasing numbers of features. Word processors, personal schedulers, e-mailing, language programming and other traditional desktop applications are increasingly available on this platform. However, PDA's text input is still a bottle-neck (Karlson et al. 2006).

Mobile activity situations often require multitasking. The requirements include unstable environment, eyes-free interaction, competition for attention resources and varying hand availability (Pascoe et al. 2000). In demanding situations, e.g. walking and talking, where the user's attention cannot be devoted fully on inputting, improvement in the input method performance is highly desired. Recent research has been done in developing speech recognition for text entry. However, speech recognition is not yet used for general purpose text input on mobile devices (MacKenzie and Soukoreff 2002). The reason is because the current technology still makes speech input less suitable for mobility (Bousquet-Vernhettes et al. 2003). Therefore, manual pen-based text entry remains one of the dominant forms of user interaction on PDAs. These devices accommodate single-handed interaction to offer users freeing a hand for holding the device or other mobile activity demands.

One of the challenges of a new keyboard design is the user requirement on ability to use it without the need for extensive practice (Bohan 1999). Handwriting is arguably the most intuitive input interaction method for PDAs. However, current handwriting recognition technology is still around 87%-93% accuracy (MacKenzie and Chang 1999). Lalomia (1994) reported that users are willing to accept a recognition error rate of only 3%. Although it can be improved to 97% after 3 hr of practice (Santos 1992), human's hand text entry speed is limited to 15 wpm (Card et al. 1983). Thus, the entry rates of handwriting can never reach those of touch typing - 20-40 wpm (MacKenzie and Soukereff 2002).

In contrast to physical keyboards, with on-screen touch keyboards the key layout has a major effect on the text entry performance (Isokoski 2004). This is because typing is strictly sequential. To type a character, we have to move the pen from one key to the next and during this time there can be no preparation for the following key. Thus, minimizing the distance to be traveled can greatly enhance text entry speed. Nevertheless, visual scan time is still necessary to distinguish an individual character from the group (Eriksen and Eriksen 1974). Familiarity with the location of the characters on the keyboard does appear to facilitate entry performance (MacKenzie et al. 1999). Entry performance can also be increased by adding visual cues to draw a user's attention to the next most probable character(s) in a word they are typing (Magnien et al. 2004). In such situation, certain characters should have a distinctive appearance that differs from others (Wolfe 1994). One of the ways is by expanding some keys' size that allows users to select larger target to improve target acquisition time (McGuffin and Balakrishnan 2002).

[1] The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024

Everyone has his/her own style of writing and communication, especially in personal writing, such as mail, SMS, personal note or diary. The style reflects on word choices and compositions in a sentence. An adaptive text entry system is able to provide prediction to a user based on its experience with this user and improve its ability based on the user's needs over time. The system collects traces of user linguistics compositions, constructs knowledge about the user from these traces through learning, and using this knowledge to alter its future interactions. In this way, the resulting text entry system is personalized to the individual user.

In this paper, we introduce the idea of an adaptive and personalized single-handed pen-based text entry on a PDA. We develop an n-gram based predictive feature that is able to propose next-character and next-word selections based on the user's personal way of formulating language, the context of the user's task and the English syntax. Using the results of this prediction, the user interface is able to display characters in different sizes and color contrasts according to their relative probabilities.

The structure of this paper is as follows. In the following section, we start with related work. We continue with describing our experiment in developing our system's dictionary. Further, our text prediction is presented. Then, our developed pen-based text entry model is described. Finally, we conclude the paper.

RELATED WORK

In practice the most popular pen-based keyboard design is still the QWERTY layout and its language-specific adaptations. It has been observed that this layout is not optimal for pen-based text entry because the distance between common adjacent characters is too far. Previous work in developing adapted keyboard layouts for handhelds and single-handed use has concentrated on alternative key configuration for improving entry speed, such as Metropolis (Zhai et al. 2000), ABC (MacKenzie et al. 1999), and OPTI (MacKenzie and Zhang 1999). Fitaly keyboard introduces two space bars and the characters arrangement so that common pairs of characters are often on neighboring keys (Langendorf 1988). An extensive study on pen-based text entry has been reported in (MacKenzie and Soukoreff 2002).

Typically, tapping-based text entry, in which the pen must be tapped for selecting characters, requires intense visual attention, virtually at every key tap, which prevents the user from focusing attention on text output (Zhai and Kristensson 2003). Gesture-based text entry methods interpret informal pen motions as character inputs, such as T-Cube (Venolia and Neiberg 1994) and Quikwriting (Perlin 1998). Another example is Cirrin (Mankoff and Abowd 1998), which arranges the characters inside the perimeter of an annulus (Figure 1). This circular layout means that when the user places his/her pen in the center of the Cirrin, the distance to each character is equal. The most commonly used digrams are nearest to each other, therefore distances traveled from character to character are usually shorter than a QWERTY-based on-screen keyboard. However, since there is not any "head-up" feature, a user must attend to the interface when entering text. A space is entered by lifting the pen. Punctuation and mode shifts are accomplished by using an auxiliary technique, such as keys operated by the nondominant hand.

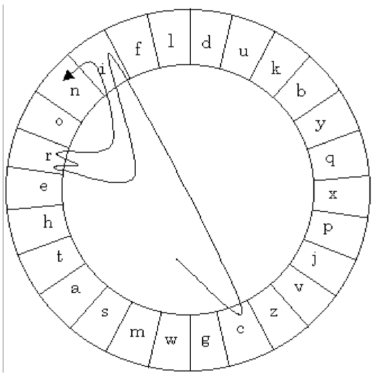


Figure 1: Standard Cirrin (Mankoff and Abouwd 1998)

Unexpected results appearead in a research of real-time expanding Cirrin's key size as the pen approach it (Figure 2 - Cechanowicz et al. 2006). It indicates a slower and more error prone user performance than the standard Cirrin. The problem is in finding an optimum threshold between two adjacent keys, so that the user does not make incorrect selection. Another reason is the position of "backspace" key being outside the Cirrin wheel, which is needed for faster error recovery.

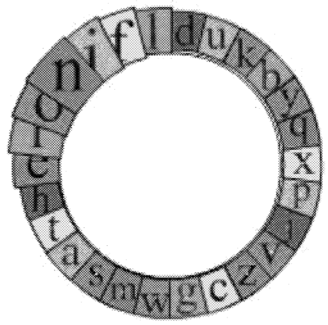


Figure 2: Expanding Cirrin (Cechanowicz et al. 2006)

SHARK is a hybrid method on the ATOMIC keyboard that augments tapping-based input with gesture-based input (Zhai and Kristensson 2003). The researchers reported that visually guided tapping is easier for novice users. Since simple tapping movement may feel tedious to repeat for prolonged use, gesture-based input is preferred by experts.

Some text input techniques have been developed with both movements minimizing and predictive features. T9 text entry works by comparing sequences of key presses to a stored database of possible words (Tegic Communication). Dasher uses prediction by partial matching, in which a set of previous symbols in the uncompressed symbol stream is used to predict the next symbol in the stream (Ward et al. 2000). It employs continuous input by dynamically arranging characters in multiple columns positioning the next most likely character near the user's pen input. The options are presented to the user in boxes sized according to their relative probabilities, to optimize the movement time. Because the character arrangement constantly changes, Dasher demands user's visual attention to dynamically react to the changing layout.

EXPERIMENT

An important aspect of our proposed text entry system is the word prediction, which is based on the user's personal way of writing. But, which source can be used so that our system can learn and trace the user's writing style? How useful are these sources for faster prediction? How to use this data source? To answer these questions, an experiment has been performed by comparing common English words use and personal use.

To collect data for analysis, we have prepared a set of words from four different sources: (1) most common English words from British National Corpus (BNC), (2) 5.5 Mb personal documents, such as words documents, spreadsheets, and schedulers, (3) 4.2 Mb personal chat logs (ZetaTalk 2001-2003) and (4) 7.2 Mb corporate e-mails (Corrada-Emmanuel). The author of the personal documents is a researcher in the field of multimodal communication. The chat logs contain a multitude discussion from philosophical topics like life aftertime or aliens presence to the science and government acknowledgement on aliens. The e-mails were taken from internal e-mails of the Enron corporation, an energy company in Houston, Texas.

As the first step, we collected all words from each dataset and counted their frequency. The BNC database has provided word frequency counts. We selected 5500 most frequent words from each personal dataset. These words appear at least 20 times in each dataset.

Which and How Useful are the Data Sources?

In this step, we compared the coverage of the common English words represented by BNC Database to all personal datasets. Table 1 shows that the BNC Database can cover in average 87% for each context and about 74% for the union of all personal datasets. Most words that are not covered by the BNC database from personal documents are abbreviations, names and specific terms, such as: "xml", "website", "lexicalized" and "synset" in the field of computer science. 78% of the words in e-mail datasets that are not covered by the BNC database are addresses and names of persons, products and organizations. Other 11% are specific terms, such as "teleconference", "worldnet" and "unsubscribe" in the field of communication network. Some of the words in chat logs that are not covered by the BNC database are popular terms in chatting or informal conversation, such as "lol" (laugh out loud), "okidok" or "yup" (OK), "thingie" (such thing), "heck" (hell) and emoticons, for example: ":)" for smile and ":))" for laughing. Others (91%) are names and internet addresses.

Table 1: The Coverage of BNC Database towards the Personal Datasets

Unigram	Number of words	BNC Database (166261 words)	A∨B∨C
A:Personal Docs	5500	4982 (90%)	49%
B:E-mails	5500	4740 (86%)	49%
C:Chat Logs	5500	4754 (86%)	49%
A∧B∧C	1685	1674 (99%)	15%
A∨B∨C	11168	9579 (85%)	

As a next step, we calculated the bigram frequency for each dataset and discarded those bigrams that contain words not covered by the BNC database. Table 2 shows that the BNC database has the lowest coverage for the personal document dataset. Although all words in each bigram are covered by the database, the compositions of them may not. Most of these bigrams are terminologies in a specific domain. For example: "human interaction", "usability testing", and "interface design" in the field human-computer interaction; "multimodal fission", "dialogue management" and "natural language" in the field multimodal system; and "emotion expressions", "facial recognition", and "muscle coordination" in the field nonverbal communication. They are considered as the most frequent bigrams (at least 29 times).

Most bigrams in the e-mail dataset that are not covered by the BNC database are terminologies in corporate domain, such as: "financially bankrupt", "employee transition", "expense report" and "retirement plans". Small amount bigrams are in the field of communication, such as "intended recipient", "conference call", and "video connection". The chat logs also contain bigrams in a specific domain that are not covered by the BNC database, such as: "planet x", "pole shift", and "star children". Small amount bigrams are about science, such as "gravity particle", "volcanic ash" and "orbital path".

Table 2: The Coverage of BNC Database towards the Personal Datasets

Bigram	Number of bigrams	BNC Database (726000 bigrams)	A∨B∨C
A:Personal Docs	54829	33994 (62%)	56%
B:E-mails	10505	7016 (83%)	11%
C:Chat Logs	36801	29809 (81%)	37%
A∧B∧C	2426	2348 (96%)	2.4%
A∨B∨C	89275	68742 (77%)	

Moreover, although the coverage of the BNC database to the convergence of the personal datasets is quite high (99% for unigrams and 96% for bigrams), these datasets themselves share a small amount of the corpus (15% words and 2.4% bigrams). One of the reasons could be that these datasets are not retrieved from the same source (nor produced by the same person). Another reason could be that each dataset is taken from a specific context. Thus, these findings show that there is a strong correlation between user personal word usage and the context of the user task.

How to Use the Data Sources?

In this step, we built a hierarchical hash-table for each dataset. This hash-table simulates user character entries to serve as a prefix before a completion of a word without any prediction (see an example in Figure 3). The end of a hierarchy shows that there is no longer possible word for the next prefix input. The different columns show that some character inputs are necessary for completing the word, for example for the word "thereby" a user needs to input "t", "h", "e", "r", "e", and "b" to distinguish this word with "there".

Prefix(es):

h	e	s/r/o	e/m/r	b/e/o	m/i

Hash-table:

to					
	the				
	their				
		thesaurus			
		these			
		thesis			
		there			
				Thereby	
		thermo			
					thermometer
		theory			
				Theoretic	
					theorist
	then				

Figure 3: A Part of a Hierarchical Hash-table for The First Character "t" (Schematic View – Read From Left to Right)

Using hash-tables, we analyzed how many appropriate number of character entries are necessary before a user can select a completion. Figure 4 shows the coverage of each dataset. According to the graph, a user has in average a 3.6% chance of being able to enter the word she/he desires in just one character entry. It also shows an almost similar coverage of 5500 most frequent words in all datasets for every prefix.

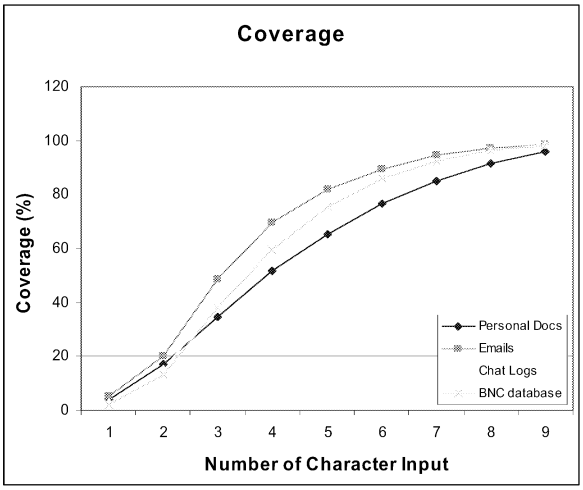


Figure 4: Coverage of 5500 Most Frequent Words from Four Datasets

Assume the completion is using all words in datasets: (1) BNC database contains 166261 words, (2) personal document dataset contains 19121words, (3) chat log dataset contains 15432 words and (4) e-mail dataset contains 13046 words.

It proves that the performance of the completion is degraded due to the inclusion of lower frequency words (Figure 5). The completion will be more effective using a relatively small dictionary containing the highest frequency words in the English language based on normal word usage. This implies to the previous finding, which shows that the personal datasets share only a small number of the corpus. A set of context-based dictionaries (for each user's context) would be more efficient for the completion than one large dictionary that contains all possible corpora.

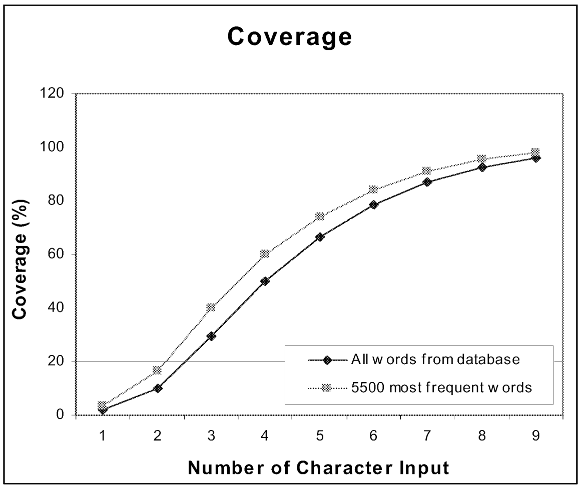


Figure 5: Average Coverage of All Words Versus 5500 Most Frequent Words from Four Datasets

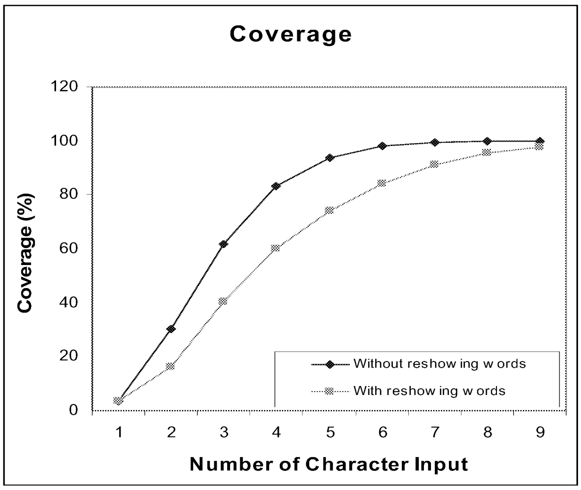


Figure 6: Average Coverage of the 5500 Most Frequent Words from Four Datasets with Reshowing Words and without Reshowing Words

Figure 6 shows if the completion is not reshowing the same word completions once these words have been shown for a given word being entered. For example, when "ther" is written, "there" is one possible completion. If "e" is inputted next, a better option is to show a different word completion, for example "thereby". By this way, those empty cells, for example from "there" to "thereby" and from "thermo" to "thermometer", are disappeared. This option will reduce the number of inputs to select a desired word, since users sometimes miss the initial appearance of the word they intended and enter more characters than necessary. This finding is coherence with Wobbrock and Myers (2006).

COMPUTATION OF CONDITIONAL PROBABILITIES

Our developed text entry system has a *word prediction*, which consists of several components (see Figure 7). The prediction result is then presented to the user. Each component in the developed word prediction is explained below.

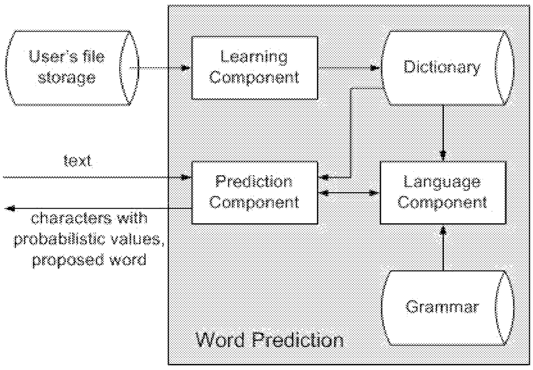


Figure 7: Schematic View of Our Developed Word Prediction

Dictionary

Our word prediction system has two main dictionaries, such as: (1) a common dictionary and (2) a user-personal dictionary, which consists of sub-dictionaries for every user's context. The current implementation defines three contexts, such as: (a) writing a document, (b) writing an e-mail and (c) chatting. Both dictionaries consist of a unigram list, a trigram list and a bigram list. They include information about part-of-speech tags and frequencies of each element. The common dictionary has been extracted from the BNC database. It provides the same frequency for all users at the beginning. The user-personal dictionary is empty at the beginning. During interaction, the system will change and adapt both dictionaries.

Learning Component

When the text entry system is used at the first time, the *learning component* parses all personal documents in the user's storage. The user may specify folders and files that can be extracted by this component. Otherwise, by default, it will extract first all personal word processor documents (including spreadsheets and schedulers) and e-mails (including its address book). This process fills the personal dictionary and updates the common dictionary.

The learning component updates the dictionaries by two ways. Firstly, it extracts the user's inputs during interaction. Finally, this component extracts the dictionary from the user's storage frequently. The user may schedule this process.

Prediction Component

The *prediction component* operates by generating three lists of suggestions for possible words after the first character is inputted, such as: (a) from the common dictionary, (b) from the personal dictionary and (c) based on the context of user's task. If the input is the character of the first word in a sentence, this component will return all words that start with the same set of characters.

After the first word is inputted, the next possible words are predicted using a statistical approach that was derived from a probabilistic language model. The probability of a sentence is estimated with the use of Bayes rule as the product of conditional probabilities:

$$P(s) = P(w_1, w_2, ..., w_n) = \prod_{i=1}^{n} P(w_i \mid h_i) \qquad (1)$$

where h_i is the relevant history when predicting a word w_i. To predict the most likely word, a global estimation of the sentence probability is derived which is computed by estimating the probability of each word given its local context (history). Our prediction component uses estimating conditional probabilities of trigrams type features. The probabilities obtained from uni-, bi- and trigrams are weighted together using standard linear interpolation formula. The system will calculate the prediction on all three dictionaries.

The results of the prediction are ranked based on their probability. The information about the part of speech tag given a word in both suggestion lists is also included, since a word form may be ambiguous and adhere to more that one part-of-speech. These lists are filtered to have all words that start with the same set of characters as the user's input.

Language Component

Besides for improving the input speed by personalizing the word prediction, our developed text entry system aims to improve the quality of syntax. Most available word-predictions have been developed based on n-gram frequencies, which often suggest syntactically implausible or excluding more-plausible but lower probability from its suggestion list. This can confuse users by inappropriate suggestions. Therefore, the overall motivation for the *language component* is to enhance the accuracy of the prediction suggestions. This component does not by itself generate any prediction suggestions but filter the suggestions produced by the n-gram model so that the grammatically correct word forms will be presented to the user prior to any ungrammatical ones.

Input to this component is three ranked lists of the most probable word forms according to the n-gram model with their part-of-speech. The language component checks all suggestion words based on its tense and morphology rule. The current implementation is able to check and change the form of a verb (tense), a noun (pluralism) and an adverb using WordNet (Fellbaum 1998) in three steps: (1) stemming all words, (2) creating all forms for each word, and (3) checking in the WordNet whether each new form is a correct form. Since a word form may be ambiguous and adhere to more forms, all word forms are added to the suggestion lists with the same probability.

The language component parses the sentence fragment entered so far. The part-of-speech tag model requires information about the possible part-of-speech tags of each word in the user's sentence. For this purpose, we used the QTAG POS Tagger (Tufis and Mason 1998), which is a (n-grams) probabilistic tagger using a dictionary of (tagged) words and a matrix of tag sequences with corresponding probabilities. The output of this tagger is the part-of-speech of each word (for example noun, verb, and adjective) in a sentence. Our developed language component assigns a value to each word in the suggestion lists whether it is confirmed by grammatical, ungrammatical or out of scope of the grammar. Based on those values, the ungrammatical suggestions are discarded from the lists. Future work needs to be done to update the POS tagger, therefore, it includes the user-personal corpora into its dictionary.

Since only one suggestion will be presented to the user, this component will choose the highest probability word from the context-based dictionary preceded the personal and common dictionary. The suggestion from personal dictionary will be chosen preceded the common dictionary, if the context-based suggestion list is empty or the probability is lower than a threshold. Future research still needs to be done in defining optimum threshold of a suggestion's probability value.

PEN-BASED TEXT ENTRY MODEL

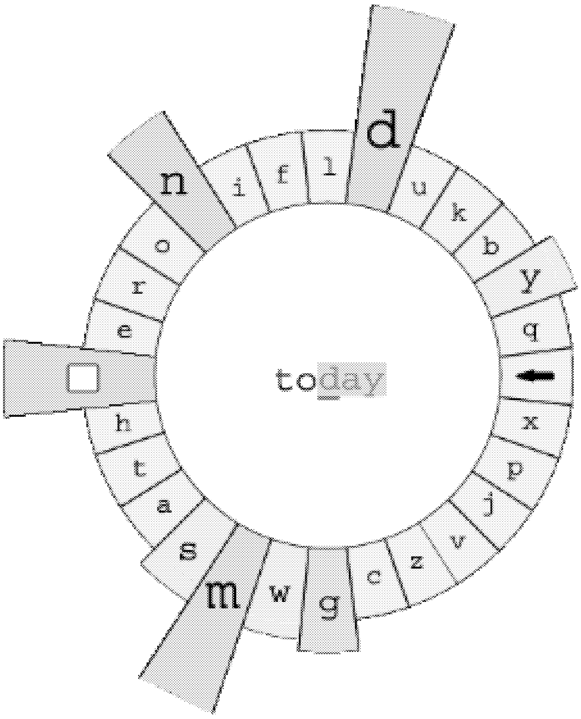


Figure 8: Personalized Adaptive Cirrin. The Previous Input Contains "go shopping"

Figure 8 shows our developed pen-based text entry on a PDA. The interface gives visual cues, such as different key sizes and color contrasts, for the next-character and next-word selections, without changing the character layout. A standard on-screen keyboard does not fit with this specification, but on a keyboard whose characters are arranged on the circumference of a polygon or a circle or in two parallel columns, it is possible to expand the keys' size. Therefore, we adopt the Cirrin device's (Mankoff and Abowd 1998) method to display all characters in a circular way. Our design differs from the original Cirrin in four aspects: (1) geometry, (2) character set, (3) input style and (4) word completion.

Geometry

Similar to the original Cirrin, we use a single column circular layout, which creates a ring of characters. The middle of the ring is an input area, where selected characters of a single word are displayed. The current implementation of our text entry system has a flowing text area, where a user can compose a message. To support direct perception of the user, every time a character is selected on the input area, it will be displayed in the text area too.

The visual cue on a key gives information about the likelihood of the next character selection. The current implementation uses 200% expansion and the most contrast color for the most likelihood characters. The key of lower probability characters is expanded and colored based on its proportion to the highest probability character.

Character Set

We use the original Cirrin's character set (26 English characters) and layout, which was based on a scoring function to calculate the most used adjacent characters. The difference is two additional characters: space and backspace. Although, like punctuation and return, they can still be entered using any common character-level input technique, these common characters are added into the ring to support a quick error recovery (Cechanowicz et al. 2006). In the event of an erroneous completion, the user can make a backspace stroke or press the backspace key, undoing the selection and restoring the completion as it appeared before. This makes completions quickly undoable. An additional matrix 6 x 5 is placed on the right side of the circle for numbers, shift, return, control, period, punctuations and comma.

Input Style

Unlike the original Cirrin, which allows only a gesture-based input, our developed text entry system allows both tapping-based and gesture-based input and combination of them. The transition of both inputs works as follows. When entering a word, the user may begin with the tapping mode and continue with the gesture mode. By this way, the new selections will be appended to the previous selections. In the gesture mode, when the user stops dragging and lifts the pen from the screen, a space will be added at the end of the input word. The user may continue inputting the next characters for the next word. When a space is selected after a word, this word will be flushed to the text area. Selecting a backspace on a space will result the word back in the input area.

Word Completion

As the user enters each keystroke, our developed text entry system displays the most likely completions of the partially typed word on the input area. It indicates which characters of the word are not yet selected. As the user continues to enter characters, the system updates the suggestion accordingly. The special feature of our word completion is that it only shows a suggestion word completion once after this suggestion is turned down by selecting the next character. If the intended word is displayed, the user simply can select it with a single tap on the input area. The system will flush this word to the text area and add a space next to the new word.

CONCLUSION

Learning from previous research on developing pen-based text entry for PDAs, we have developed a personalized and adaptive text entry system. Our developed on-screen keyboard offers a fast input and allows users to input less tedious, less visually demanding and fast error recovery by four ways: (1) visual cue for next-character prediction, (2) next word completion, (3) combining both tapping-based input and gesture-based input and (4) adding space and backspace into the circle. Inspired by Cirrin (Mankoff and Abowd 1998), the characters are arranged in a circular ring. In this research, we aim at exploring a method for adapting the text entry system according to user's personal word usage the context of user's task to reduce the time necessary to search for a desired key.

An experiment has been performed by comparing the most common English words taken from the BNC database with personal datasets, such as personal documents, e-mails and chat logs. Although the BNC database covers most of the personal corpus, the experimental results showed that the intersection of the personal datasets is small. Moreover, the word completion showed better performance using a relatively small dictionary containing the highest frequency words based on normal word usage. This indicates that, besides personal word usage, the ability to improve effective text entry and typing rate may also dependent on the context of the user task. The current implementation of our developed text entry system has a personal dictionary that consists of user context-based sub-dictionaries.

Besides saving time and energy in inputting the number of characters for completing a desired word, the proposed text entry system can also assist the users in the composition of well-formed text. For this purpose, our developed word prediction uses both syntactical and n-grams probabilistic approaches to predict next possible words. In displaying the prediction result, the system takes an assumption that a suggested word is rejected after the user selects the next character. By this way, the user can have a better language coverage since each suggestion word is shown only once.

The primary results show that the developed personalized adaptive approach offers a usable text entry device to investigate. To understand all issues involved and the full potential this our text entry system, especially in mobile situation and how people experience this, requires a great deal more research and intensive evaluations in the future. Currently, we improve the developed system by providing supports for better user-system interactions.

REFERENCES

Bohan M., Phipps C.A., Chaparro A. and Halcomb C. 1999. A Psychophysical Comparison of Two Stylus-Driven Soft Keyboards. *Proc. of Graphics Interface*, 92-97.

Bousquet-Vernhettes C., Privat R. and Vigouroux N. 2003. Error Handling in Spoken Dialogue Systems: Toward Corrective Dialogue, *Proc. of ISCA*, USA.

British National Corpus, Unigrams and Bigrams, Retrieved on January 5, 2007, from http://natcorp.ox.ac.uk

Card S. K., Moran T. P. and Newell A. 1983. *The Psychology of Human–Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cechanowicz J., Dawson S., Victor M. and Subramanian S. 2006. Stylus Based Text Input Using Expanding CIRRIN. *Proc. of AVI*. ACM Press, New York, NY, 163-166.

Corrada-Emmanuel A. (n.d.). *Enron E-mail Dataset Research*. Retrieved in January 5, 2007, from http://ciir.cs.umass.edu/~corrada/enron/

Eriksen B.A. and Eriksen C.W. 1974. Effects of Noise Letters Upon the Identification of a Target Letter in a Non-Search Task. *Perception and Psychophysics*, 16, 143-149.

Fellbaum C. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press.

Isokoski P. 2004. *Manual Text Entry: Experiments, Models, and Systems* (Ph.D. thesis), Report A-2004-3, Department of Computer Sciences, University of Tampere, Finland. http://www.cs.uta.fi/~poika/vk/isokoski_thesis_complete.pdf

Karlson A., Bederson B. and Contreras-Vidal J. 2006. Understanding Single Handed Use of Handheld Devices. Lumsden Jo (Ed.), *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*, in press.

LaLomia M.J. 1994. User Acceptance of Handwritten Recognition Accuracy. *Proc. of ACM CHI*, Boston, MA, USA, 2:107.

Langendorf D.J. 1988. Textware Solution's Fitaly Keyboard V1.0 Easing the Burden of Keyboard Input. *WinCELair Review*.

MacKenzie I.S. and Chang L. 1999. A Performance Comparison of Two Handwriting Recognizers. *Interacting with Computers*, 11(3), 283 - 297.

MacKenzie I.S. and Soukoreff R.W. 2002. Text Entry for Mobile Computing: Models and Methods, Theory and Practice. *Human-Computer Interaction*, 17: 147-198.

MacKenzie I.S. and Zhang, S.X. 1999. The Design and Evaluation of a High-Performance Soft Keyboard. *Proc. of ACM CHI*, 25-31. New York: ACM.

MacKenzie I.S., Zhang S.X. and Soukoreff R.W. 1999. Text Entry using Soft Keyboards. *Behaviour and Information Technology*, 18, 235-244.

Magnien L., Bouraoui J.L. and Vigouroux N. 2004. Mobile Text Input with Soft Keyboards: Optimization by Means of Visual Clues, *Proc. of Mobile HCI*, Springer-Verlag , 337-341.

Mankoff J. and Abowd G. D. 1998. Cirrin: A Word-Level Unistroke Keyboard for Pen Input. ACM UIST'98, 213-214.

McGuffin M. and Balakrishnan R. 2002. Acquisition of Expanding Targets. *Proc. of ACM CHI*, 57-64.

Pascoe J., Ryan N. and Mores D. 2000. Using While Moving: HCI Issues in Fieldwork Environment. *Transaction on Computer Human Interaction*, 7(3).

Perlin K. 1998 Quikwriting: Continuous Stylus-Based Text Entry. *Proc. of ACM UIST*, 215-216. New York: ACM.

Santos P.J., Baltzer A.J., Badre A.N., Henneman R.L. and Miller M.S. 1992. On Handwriting Recognition System Performance: Some Experimental Results. *Proc. of the Human Factors Society*, CA:Human Factors and Ergonomics Society.

Tegic Communication. 1998. T9. http://www.t9.com/faq.html.

Tufis D. and Mason O. 1998. Tagging Romanian Texts: a Case Study for QTAG, a Language Independent Probabilistic Tagger, *Proc of LREC*, Spain, 589-596.

Venolia D. and Neiberg, F. 1994. T-Cube: A Fast, Self-Disclosing Pen-Based Alphabet. *Proc. of ACM CHI*, 265-270. New York: ACM.

Ward D.A., Blackwell A. and MacKay D. 2000. Dasher – a Data Entry Interface Using Continuous Gesture and Language Models, *Proc. of ACM UIST*, 129-136.

Wolfe J.M. 1994. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1(2). 202–238.

Wobbrock J.O., Myers B.A. and Chau D.H. 2006. In-stroke Word Completion. *Proc. of ACM UIST*. Switzerland. New York: ACM Press, 333-336.

Zhai S., Hunter M., and Smith B.A. 2000. The Metropolis Keyboard - An Exploration of Quantitative Techniques for Virtual Keyboard Design. *Proc. of ACM UIST*, 119-128. New York: ACM.

Zhai S. and Kristensson P.-O. 2003. Shorthand Writing on Stylus Keyboard, *Proc. of ACM CHI*, 97-104.

ZetaTalk, Chat Logs – ZetaTalk Live Dec 2001-May 2003, Retrieved in January 5, 2007, from http://ww.zetatalk3.com/index/zetalogs.htm.

BIOGRAPHY

SISKA FITRIANIE was born in Bandung, Indonesia and went to Delft University of Technology, the Netherlands, where she studied Technical Informatics and obtained her master degree in 2002. After doing her two years post-graduate programme at Eindhoven University of Technology, she involves in the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, since 2004 as her PhD project at Delft University of Technology. Her project aims at designing and developing models of a computer-human interaction system.
E-mail: s.fitrianie@ewi.tudelft.nl
Webaddress: http://mmi.tudelft.nl/~siska

LEON J.M ROTHKRANTZ received the M.Sc. degree in mathematics from the University of Utrecht, Utrecht, The Netherlands, in 1971, the Ph.D. degree in mathematics from the University of Amsterdam, Amsterdam, The Netherlands, in 1980, and the M.Sc. degree in psychology from the University of Leiden, Leiden, The Netherlands, in 1990. He is currently an Associate Professor with the Man-Machine-Interaction Group, Mediamatics Department, Delft University of Technology, Delft, The Netherlands, since 1992. His current research focuses on a wide range of the related issues, including lip reading, speech recognition and synthesis, facial expression analysis and synthesis, multimodal information fusion, natural dialogue management, and human affective feedback recognition. The long-range goal of his research is the design and development of natural, context-aware, multimodal man–machine interfaces. Drs. Dr. Rothkrantz is a member of the Program Committee for EUROSIS.
E-mail: l.j.m.rothkrantz@ewi.tudelft.nl
Webaddress: http://mmi.tudelft.nl/~leon

MATHEMATICS IN EVERYDAY LIFE BETWEEN ART AND SCIENCE

A. Cascone
G. Durazzo
V. Stile
Department of Information Engineering and Applied Mathematics
University of Salerno,
via Ponte don Melillo, 4084 Fisciano (SA), Italy
E-mail: {cascone, durazzo, stile}@diima.unisa.it

KEYWORDS
Mathematics, Arts, Scientific spread

ABSTRACT

With this essay we want to rediscover and analyze the main "artistic aspects" that we can find in mathematical science in order to present an engaging science to young students. The real intention is to awaken again the interest for the scientific science, turning on the attention and the possible hidden passions by means of the support of unexpected relationships between Mathematics and any other kind of arts: from architecture to music, from painting to dancing.

In the first sections we outline the situation, in the succeeding ones we trace out the characteristics of the experimentation, explaining the decisions.

INTRODUCTION

In recent years, it is enough to attend a conversation during a high school exam to become aware of how much the students do not have an appropriate scientific education.

Their essays, thought as interdisciplinary thesis, are exclusively based on, for the most part, humanistic subjects. In rare case they mention, in a banal way, some subjects as Physics, and Natural Sciences. Sometimes the connection with and among these subjects seems to be a real forcing, and Mathematics is considered a detached subject. It is evident that the students look at Mathematics, and its formalism, like something distant from the reality, as well as from other subjects, with which it has constrained interconnections.

In the present paper, we aim to promote, divulge, popularize a Mathematics education, above all among high school students. They represent our main target, seeing that there is recently a wide gap between the labour demand for scientific graduates and the available labour supply. Therefore, it seems necessary to orientate the students towards scientific subjects. With this proposal, we would like to make the young students involved with Mathematics, showing the existing links between Mathematics and Architecture, Painting, Music, Dancing, and Cinema. We would like to arouse students' interest in scientific research and experimentation, eliminate any preconception about Mathematics, and show how Mathematics, usually seen as a hard and problematic subject, could be explained by very simple means and a clear language. In this way, Mathematics could be approachable for the "no experts" too.

Our scope is to communicate with young people through Mathematics, and show them what benefits they can get observing the world by means of Mathematics. We hope to create and boost their scientific education so that they can understand the existing connections between the different subjects and topics and the Mathematics.

SCIENTIFIC SPREAD. MATHEMATICS

A society where scientific education is widespread could be a more rational and human society. In order to have a good scientific dissemination, it is essential to be able to communicate the most significant aspects of Mathematics to all the potential people we can talk to, without misrepresentation, distortion or trivializations.

To that end, we plan to recognize contents and methods suitable for this kind of communication, and investigate, for example, the opportunities offered by multimedia tools, and the relationship between Mathematics and Arts.

It is common practice that students don't like Mathematics, indeed they look at it with awe. Most of this aversion comes from a short-sighted image of Mathematics that humble the creative aspect of this subject, preferring a rote learning and a repetitive approach. Mathematics is new and old, its roots are ancient, but its structure and its contents evolve more and more rapidly. Mathematics is an art and a science. It is everywhere in the "real" world, it is not only a mean to understand this world, but it is itself an universe, and it should be appreciated for what it is, its beauty and its unique and special nature.

Specifically, in this contest, we intend to attract the interest of young students for mathematical disciplines through the analysis of the existing links between them and the various forms of art that we can find in everyday life.

Our project plans for an experimentation in high schools oriented to promote enterprises showing new and original educational solutions in mathematical teaching. The organization of periodic meetings could represent the chance to display, through examples that can be noticed in everyday life, Mathematics as science and the ways in which it is connected with the arts and viceversa.

In these meetings, some aspects of Mathematics – as a tool for the art – have been deeply evaluated.

The foremost subjects investigated are: Mathematics and Architecture, Mathematics and Painting, Mathematics and Music, Mathematic and Dancing, Mathematics and Cinema.

With the support of professional personnel, we wanted to achieve a precise goal: make people aware of how much important Mathematics is in the elaboration and realization of whatever project. Let think about the designing of a street or the more complex planning of a bridge or a dam. Even in these cases the role of Mathematics is unavoidable.

In order to involve people with scientific and mathematical education, as well as people with different background, it is essential to choose the right contents, tools and languages close to their needs. The purpose of our research is to convey the idea that Mathematics is not an unemotional, distant and detached discipline, as the most part of people use to think about it. Instead, it is pervaded with beauty, elegance, emotions, and, above all, it is tightly connected to the meaningful and intense issues of the human life.

EXPERIMENTATION PROPOSALS

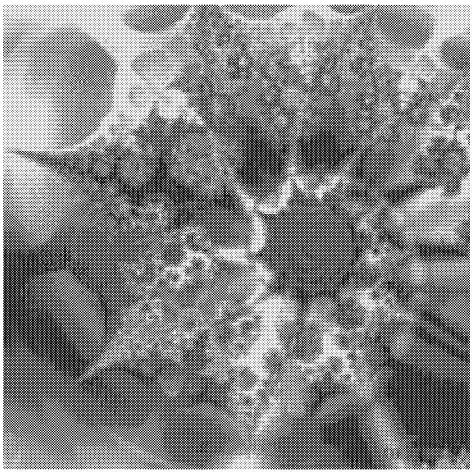
The topics discussed in these meetings helped to illustrate and clarify the interconnections between Mathematics and every kind of Art (Music, Painting, Architecture, and so on). The specific characteristics of Mathematics were emphasized, as well as the peculiarities of the different didactic levels.

Our priority and main goal is to increase the interest in culture, knowledge, research, technological and cultural innovation, inventiveness, planning imagination, so distinctive characteristics of Mathematics. Besides, we want to stimulate the attention and the interest of young people.

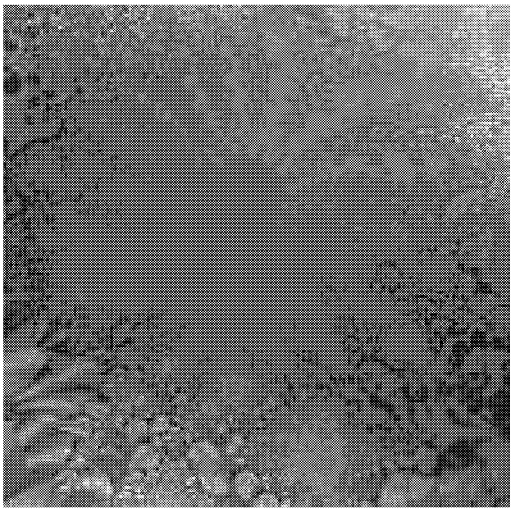
The wonderful lines you can notice in a fractal are well-know, and they are very similar to the lines discoverable in nature. Nature itself could be regarded as a huge fractal.

In 1903, in “The Principles of Mathematics”, Bertrand Russell wrote: *Mathematics, rightly viewed, possesses not only truth, but supreme beauty - a beauty cold and austere, like that of sculpture*. In “A Mathematician’s Apology” (1940), the English mathematician Godfrey Harold Hardy asserted that: *the mathematician’s patterns, like the painter’s or the poet’s must be beautiful; the ideas like the colors and the words, must fit together in a harmonious way.*

It may be very hard to define *mathematical beauty, but that is just as true of beauty of any kind – we may not know quite what we mean by a beautiful poem, but that does not prevent us from recognizing one when we read it.*



Figures 1: Example of a fractal.



Figures 2: Example of a fractal.

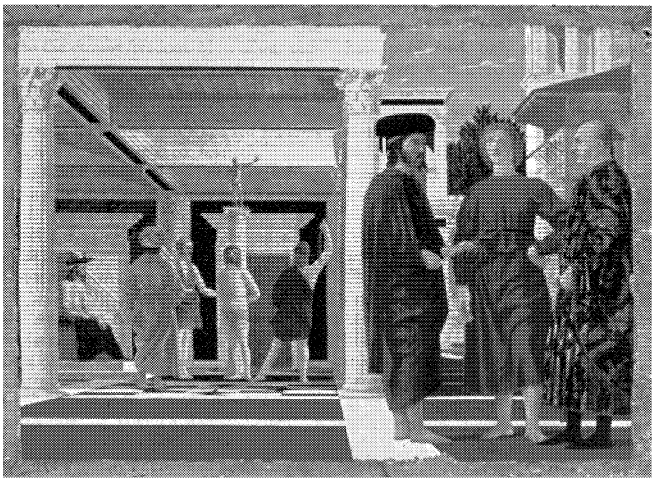
We intend to convey the passion for the beauty of Mathematics. Therefore, we would like to present the hidden correlations between Mathematics and other kinds of art.

Could be very interesting to present the correlation between Mathematics and Architecture through the example of the use of geometric figures in paving design. Showing the ways in which people at different times, in different places have used geometric patterns as embellishment could be a drive for young people to take an interest in Mathematics and Geometry. One of the main goals we would like to achieve is to be able to intrigue the students, presenting the universality of the geometric patterns that we can find always and everywhere.

Figures 3: Wakefield Cathedral Precinct. Paving by Tess Jaray.

It is not so hard to detect a lot of other connections with different subjects as Music, Dancing, Painting, and so on. For example, let’s think about the rhythm, the counterpoint and the symmetry in Music. Consider the perspective in Painting, and the simple rules at the roots of Dancing, a discipline that could seem chaotic and instead is strict and harmonic.

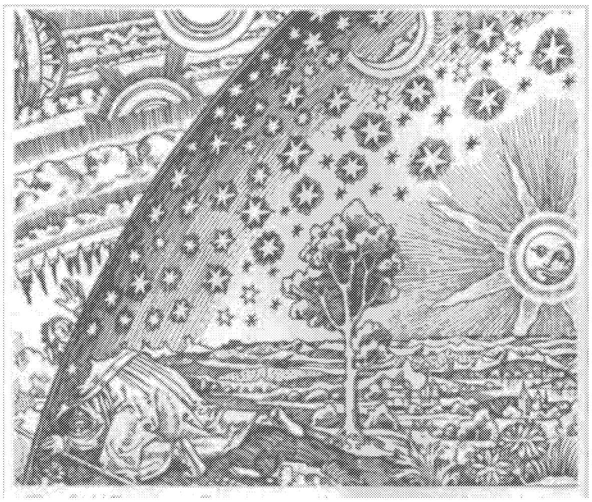
Figures 4: American Academy of Ballet.



Figures 5: Piero della Francesca, The Flagellation.

Let's reflect upon the close and elusive ties between Mathematics and Poetry.
Stanislaw Lem in "The Cyberiad" wrote:

In Riemann, Hilbert or in Banach space
Let superscripts and subscripts go their ways.
Our asymptotes no longer out of phase,
We shall encounter, counting, face to face.



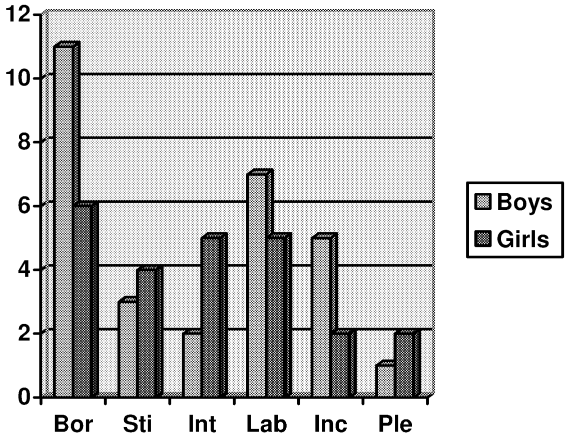
Figures 6: Sixteenth century woodcut representing a man pushing beyond the boundaries of the celestial sphere.

Certainly it is not easy to catch immediately its essence, but it is possible to think about the poetic interpretation of the infinity, or about the stimulating definitions of an asymptote as the platonic love or as a life devoted to holiness…

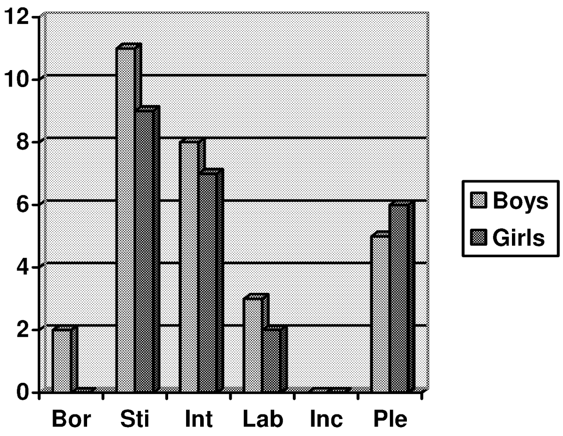
We tested our aims at a Secondary School, where we wanted to realize and spread some events as exhibitions, seminars, workshops, CD-ROMs, multimedia tools, in order to value their impact on the school world.

RESULTS

At the end of this series of meetings we obtained some important results: a renewed, vivacious, and dynamic interest in Mathematics and in Sciences at large, thanks to some attractive, creative ways of presenting these disciplines, and to the enhancement of those aspects in Mathematics that are connected to the Art in general.
We gathered a certain number of filled questionnaires (see the Appendix) from a survey in a High School in Southern Italy. The taken sample was made of 54 students – 29 males, and 25 females – aged between 16 and 17. From their feedbacks we can deduce that there is a different way, between males and females, to consider and understand scientific subjects. While girls showed a previous interest in Mathematics apart from our experimentation proposals (as exhibits, seminars and multimedia means), boys instead are more attract by practical approach and the several links between Mathematics and the Art in general revealed through our experimentations. In general, it is important to observe a widespread improvement in teaching-learning processes independently from the gender.
We show some feedbacks in the following histograms. On the whole we can say, finally, that almost everyone changed one's own opinion about teaching and learning Mathematics and most of them propose the adoption of new methodologies in teaching and learning scientific disciplines as Mathematics.



Figures 7: Histogram about Mathematics liking before meetings.



Figures 8: Histogram about Mathematics' liking after meetings.

This new approach, that harmoniously connect Mathematics with different artistic disciplines, want to be a strong incitement for young people to direct their attention to topics they think to be boring, unattractive, and monotonous.

Using arts – as Music, Dancing or Painting – so close to the "Youth world" we hope to urge young students to develop, improve and spread a mathematical culture. It could be, at the same time, a way to boost and promote scientific research, and the right solution to the scarce number of people with a scientific degree, in spite of the present increasing demand for this kind of graduates.

CONCLUSION

Investigating about mathematics' "*popularity rating*" allowed us to invite young students to devote themselves to mathematics and science. This training project was based on the surprising discovery of the existing connections between Mathematics and arts. In such a way we wanted to arouse students' curiosity and interest. Students involved in this experience could become aware of mathematics' key role. They discovered an unexpected beauty hidden in artistic disciplines.

APPENDIX

In what follows, we show the feedback form submitted to students.

They filled it in an anonymous manner.

We collected a 54 filled questionnaires from a survey in a High School in Southern Italy. The taken sample was made of 54 students – 29 males, and 25 females – aged between 16 and 17.

Table 1: Feedback form

	Male □	Female □	
	Age	…………………	
1)	Favourite Subjects	Literature	□
		Mathematics	□
		Physics	□
		Biology	□
		Philosophy	□
		History	□
		Foreign Language	□
		Others …………………	□
2)	Favourite Sports	Soccer	□
		Basket	□
		Football	□
		Volley	□
		Cycling	□
		Riding	□
		Swimming	□
		Tennis	□
		Others …………………	□
3)	Hobbies & Leisure time	Music	□
		Dancing	□
		Painting	□
		Cinema	□
		Poetry	□
		Others …………………	□
4)	What did you think about Mathematical Subject before these meetings? It was …	Boring	□
		Stimulating	□
		Interesting	□
		Laborious	□
		Incomprehensible	□
		Pleasant	□
5)	What do you think now about Mathematical Subject after these meetings? It is …	Boring	□
		Stimulating	□
		Interesting	□
		Laborious	□
		Incomprehensible	□
		Pleasant	□
6)	Did you already know all these links existing between Mathematics and Arts?	Yes	□
		No	□
7)	Would you recommend a similar course to someone?	Yes	□
		No	□
8)	Please, write here your comments, suggestion, proposals …	…………………………	
		…………………………	
		…………………………	

REFERENCES

Bruner J. (1966): Toward a Theory of Instruction. Cambridge, MA, Belknap Press/Harvard University Press.

Emmer M. (1994): Thevisualmind: Art and Mathematics, *The MIT Press* (Cambridge Mass).

Emmer M. (2005):Mathland: The Role of Mathematics in Virtual Architecture, *Nexus Network Journal*, , vol. 7 no. 2 - http://www.nexusjournal.com/Emmer.html

Hardy G. H.: (1940): A Mathematician's Apology, Cambridge University Press; ed. it., ID., *Apologia di un matematico*, Bari, De Donato ed., 1969

Kitchener R. (1986): Piaget's theory of knowledge. New Haven: Yale University Press.

Lem S. (1995): Cyberiade, URANIA – Mondadori.

Russell B. (1903): The Principles of Mathematics, The University Press.

BIOGRAPHY

ANNUNZIATA CASCONE was born in Castellammare di Stabia, Italy. She graduated cum laude in Mathematics in 2004. She is a actually a PhD student in Mathematics at the University of Salerno. Her scientific interests are about Computer Aided Learning, Queueing Theory and Fluid-Dynamic Models for traffic flows on road and telecommunication networks.

GERARDO DURAZZO was born in Castellammare di Stabia, Italy. He graduated cum laude in Mathematics in 1992. After a long period as teacher at High School, actually he is a PhD student in Mathematics at the University of Salerno. His scientific interests are about Computer Aided Learning.

VALENTINA STILE was born in Castellammare di Stabia, Italy. She graduated cum laude in Sociology in 2002. She is actually a Researcher in Knowledge Sociology at the University of Salerno. Her scientific interests are about Computer Aided Learning, Distance Learning.

EXPERT ADVICE AND REGRET FOR SERIAL RECOMMENDERS

Anton Eliëns
Faculty of Sciences
VU University Amsterdam
email: `eliens@cs.vu.nl`

Yiwen Wang
Dept. of Math. and Comp. Sc.
Eindhoven University of Technology
email: `y.wang@tue.nl`

KEYWORDS
recommender systems, expert advice, decision theory, personalization, guided tours, digital dossier, cultural heritage

ABSTRACT

In this paper we propose a tentative framework (R3) for adapting a sequence of predictions (guided tour) generated by what we call a *serial recommender*. The R3 framework (*rate, recommend, regret*) is applied to the construction of personalized guided tours, based on expert advice, in the domain of cultural heritage, in particular *digital dossiers* about contemporary art. Guided tours are in first instance obtained by tracking expert users. Our proposal is based on a variant of decision theory, that uses a regret function to measure the difference between a proposed decision and a finite collection of expert decisions. In our framework, personalization may then be seen as a minimization problem over a weighting scheme, expressing the relative importance of experts of which tours are available. Our aim in this paper is to arrive at a formalization of the recommendation of sequences (guided tours) that allows for adaptation to individual user preferences by a revision of the weight attached to a particular advice based on user feedback.

INTRODUCTION

Leaving all responsibility for interaction to the user is usually not a good choice, in particular when an information system contains complex, highly interrelated information. Despite the wealth of recommendation systems, it still seems to be an open problem how to generate a related collection of recommendations, that is an organized sequence of recommended items that me be used as a guided tour, for example an overview of artworks and related information from a museum collection.

In Eliens et al. (2006b), Wang et al. (2006), van Riel et al. (2006) we describe the *3D digital dossier* format, in which we presented the information of respectively the Dutch-Serbian artist Marina Abramovic[1] and the Australian artist Jeffrey Shaw[2], contemporary artists with a variety of work, ranging from video to art installations. The *digital dossier* supports navigation using a concept graph and allows for presenting media-rich material, including 3D models of artwork installations. The digital dossiers have been implemented using X3D/VRML[3] to allow for deployment on the web.

Recently we have explored *guided tours* in digital dossiers, van Riel et al. (2006), which actually automate user interaction, by mimicking user actions through events generated by a script. Although this provides an easy way to create guided tours, this does not solve the problem of what to select as elements in the guided tour, or how to personalize these tours in an intelligent manner.

In this paper, we discuss techniques from decision theory as a means to aid the construction of guided tours by consulting an advice function based on tracking the navigation behavior of expert users. We will also indicate how a similar advice function can be used for personalizing tours in cooperation with a recommender system for artworks, by altering the weight given to particular properties.

More in general, our aim is to arrive at a formalization of the mechanics underlying the recommendation of sequences (guided tours) that allows for adaptation to individual user preferences by a revision of the weight attached to a particular advice based on user feedback. Moreover we will give an indication how to generalize our approach to include the refinement of content-based ratings from which sequences are generated, by adapting weight attached to specific attributes of items featured in the guided tour. We opt for the phrase *serial recommender*, to stress on the one hand that the recommendation concerns sequences and not individual items, and on the other hand what one may call the compulsive nature of the recommendations, due to the fact that they are originally generated by experts. Mind that our approach has been primarily motivated by the need to support guided tours in digital dossiers. As we discuss in more detail in the paper, digital dossiers, and in particular the concept graph as a navigation paradigm, adhere to specific constraints that do not apply in general. As a consequence, it might be hard to

[1] www.few.vu.nl/~dossier05

[2] www.few.vu.nl/~casus05

[3] www.web3d.org

generalize the approach to other domains where guided tours are useful. However, by including ratings based on content and an appropriate distance function between recommended items, it seems that the R3 framework introduced here is applicable to a wider class of (serial) recommenders.

structure The structure of this paper is as follows. First we will give a brief overview of recommdender systems, after which we will give a short introduction to decision theory. Then we will describe the *abramovic* dossier, and discuss how techniques from decision theory can be applied to the construction of guided tours in digital dossiers, followed by a discussion of how to realize expert advice functions in digital dossiers. We will then illustrate how to apply decision theory for the personalization of tours in a more conventional cultural heritage application, sketch a formal model for (serial) recommender systems, introduce a distance function for item recommendations, and indicate how to deal with user feedback discrepancy. Finally, we will give our conclusions and indicate directions for future research.

RECOMMENDER SYSTEMS – BRIEF OVERVIEW

There is a great wealth of recommender systems, and a daunting number of techniques for producing recommendations, based on content, user behavior or social groups. See the AAAI 2004 Tutorial[4] on recommender systems and techniques for an (extensive) overview.
In Van Setten (2005) a distinction is made between the following types of prediction techniques:

- social-based – dependent on (group) rating of item(s)
- information-based – dependent on features of item(s)
- hybrid methods – combining predictors

Social-based prediction techniques include collaborative filtering (CF), item-item filtering, popularity measures, etcetera. Information-based prediction techniques include information filtering, case-based reasoning and attribute or feature comparison. Finally, as hybridization techniques, Van Setten (2005) distinguishes between weighted combination, switching, mixed application and meta-approaches such as feature combination and cascaded application.
The approach we present in this paper, the R3 framework, has aspects of social-based as well as information-based methods and may be characterized as hybrid since it uses a weighting scheme to select between experts for advice.
For clarity, it is worthwhile to delineate briefly what we understand by the phrases *rate, recommend, regret*, and how the R3 framework fits within the wider scope of recommendation techniques:

- *rating* – a value representing a user's interest
- *recommendation* – item(s) that might be of interest to the user
- *regret* – a function to measure the accuracy of recommendations

In our approach, we (initially) proceed from the assumption that a rating is already present, and more in particular a rating that implies a sequential order on the presentation of a (limited) number of items. Later, however, we will explore how to relax this assumption and apply the R3 framework to sequences that are generated on the basis of content-based user preferences, to allow for an incremental adaptation of recommendations.

MATHEMATICAL PRELIMINARIES – DECISION THEORY

Before discussing how to realize guided tours in digital dossiers using user tracking and expert advice, we will give a very brief introduction to decision theory, more in particular a variant of decision theory introduced in Cesa-Bianchi and Lugosi (2006), that provides a mathematical foundation for our approach.
In classical prediction theory a prediction is a sequence of elements $x_1, x_2, \ldots$ that results from a stationary stochastic process. The risk of the prediction is taken to be the expected value of the accumulated *loss* function, measuring the discrepancey between predicted values and actual outcomes. Cesa-Bianchi and Lugosi (2006) introduce a variant of prediction theory in which no assumption is made with respect to the nature of the source of predictions. Instead, the *forecaster* is considered to be an entity that gives a prediction for an element based on *advice* of one or more *experts*. These experts might be actual sequences stored in a database. The deviation of the forecaster with the actual outcome is measured using a *regret* function, and the prediction task may hence be formulated as minimimizing the *regret* function by choosing the best expert for advice for each element of a prediction sequence.
For example, for the prediction of a bitstring of length *n*, the forecaster is a vector of *n* expert indices, that give advice for the bitvalue, *0* or *1*, in that position. In the general case, in which we have no information on the error rate of the experts' advice, we may use a weighting factor $0 \leqslant \beta_i \leqslant 1$ for each expert *i*, to indicate the credibility of the experts' advice. After each prediction, obtained by taking the majority decision of the experts, according to the weighting scheme, we may verify which experts fail to give the right advice, and decrease their weight, thus eliminating the influence of their advice in the long run.

[4] www.dfki.de/∼jameson/aaai04-tutorial

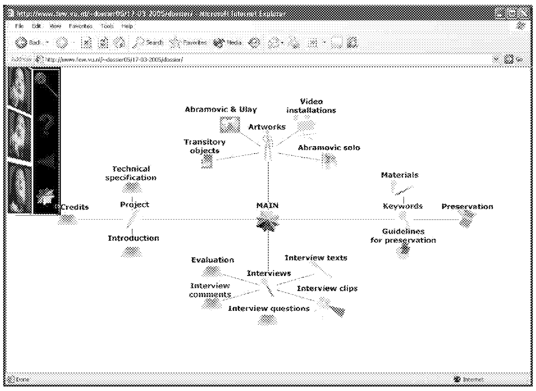


Fig. 1: Concept graph

THE *ABRAMOVIC* DOSSIER

As a user interface for navigating the *abramovic* dossier, we created a concept graph, fig. 1, that represents arbitrary information structures in a hierarchical way. The concept graph allows the user to detect relations and search for information. Unlike the 3D cone tree, Robertson and MacKinlay (1991), where the complete hierarchical structure is presented, only a subset of the hierarchy is shown - three levels deep.
Presentation is an essential part of the digital dossier but is separated from navigation. The digital dossier contains different presentation facilities for 2D and 3D content. For 2D media content we need to be able to present video, images or textual information. This is implemented as a presentation gadget with three windows, fig 2. In each of the three windows the user can view either text, image or video content. The windows are positioned in such a way that the user can inspect the information simultaneously. In our experience, three views can be presented at the same time without much visual distortion.

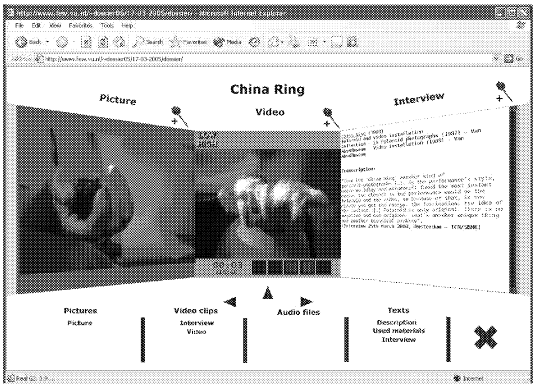


Fig. 2: Content gadget

usage scenario: When starting the dossier, it loads the concept graph that is used to navigate through the available information. In the center of the concept graph, a shining star is shown to illustrate the root of the information hierarchy, which is used as the start object. When clicked, a star structure spreads and child objects appear surrounding the center star object.
Clicking on the *Interviews* node gives an overview of all interview fragments, then going back clicking on the information node *Artworks* and then on *China Ring* will bring the node for *China Ring* into focus. When clicking on the center node *China Ring*, a content presentation environment appears. which has three windows to present different types of information, grouped into the categories text, pictures and video. If desired, the user can focus on any window by using a zoom function. When the presentation of media content is finished, clicking on the close button will result in going back to the concept graph. Alternatively, the home function of the tool bar may be used to return directly to where we started: the original shining star.
An important feature of our digital dossier is the possibility to include 3D models of artwork installation. For example, in fig. 3, the installation *Terra della dea Madre* is shown, which allows for interactive manipulation, such as rotation and positioning as a means to experiment with exhibition parameters in (virtual) space.

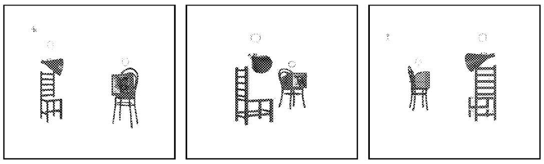


Fig. 3: Reconstruction of *Terra della dea Madre.*

GUIDED TOURS IN DIGITAL DOSSIERS

In digital dossiers, we explored the use of guided tours as a means to present the information in a story-like way, relieving the user of the often cumbersome task to interact, van Riel et al. (2006b). Guided tours, in the digital dossier, may take one of the following forms:

- automated (viewpoint) navigation in virtual space,
- an animation explaining, for example, the construction of an artwork, or
- the (narrative) presentation of a sequence of concept nodes.

In practice, a guided tour may be constructed as a combination of these elements, interweaving, for example, the explanation of concepts, or biographic material of the artist, with the demonstration of the positioning of an artwork in an exhibition space.
A pre-condition for the construction of guided tours based on user tracking is that navigation consists of a small number of discrete steps. This excludes the construction of arbitrary guided tours in virtual space, since it is not immediately obvious how navigation in virtual space may be properly discretized. In this

case, as we will discuss later, a guided tour may be constructed using a programmed agent showing the user around.

For navigation in the concept graph, as well as for the activation of the media presentation gadget, the discretization pre-condition holds, and a guided tour may be composed from a finite number of discrete steps, reflecting the choice of the user for a particular node or interaction with the presentation gadget.

For example, in the *abramovic* dossier, the user has the option to go from the *Main* node to either *Artworks*, *Video Installations* or *Interviews*, and from there on further to any of the items under the chosen category. Tracking the actual sequences of choices of a user would suffice to create a guided tour, simply by re-playing all steps.

To obtain more interesting tours, we may track the navigation behavior of several experts for a particular task, for example retrieving information about the installation *Terra degli della Madre*. In case the experts disagree on a particular step in the tour, we may take the majority decision, and possibly correct this by adjusting the weight for one or more experts. When we have a database of tours from a number of experts, we may offer the user a choice of tours, and even allow to give priority to one or more of his/her favorite experts, again simply by adjusting the weighting scheme.

As a technical requirement, it must be possible to normalize interaction sequences, to eliminate the influence of short-cuts, and to allow for comparison between a collection of recordings. For the actual playback, as a guided tour, a decision mechanism is needed that finds the advice at each decision point, from each expert, to select the best step, according to a decision rule that takes the weighting scheme into account.

PERSONALIZATION BY EXPERT RATING

In a more mathematical way, we may state that for each node n we have a successor function $S(n)$, that lists the collection of nodes connected with n, which we may write as $S(n) = n_1, ..., n_k$, where the suffix $i \leq k$ is an arbitrary integer index over the successor nodes. To take a history of navigation into account, we let $\overline{p}$ be a string of integers, representing the choices made, encoding the navigation path. So, for a node $n_{\overline{p}}$, with history $\overline{p}$, the collection of successor nodes is $S_{\overline{p}}(n) = n_{\overline{p}1}, ..., n_{\overline{p}k}$.

Now assume that we have a weight function w, that assigns to each expert e_i a weight $0 \leq \beta_i \leq 1$, indicating the relevance of expert i. Then for a particular node n we may assume to have an advice $\alpha_i = x$, with weight β_i and x in $S(n)$. If an expert has no advice for this node, we may simply assume its weight to be 0. For a collection of experts, the final advice will be $\alpha(n) = \alpha_i(n)$ with weight β_i and $w(e_i) > w(e_j)$ for $i \neq j$. If no such advice $\alpha_i(n)$ exists, we may query the user to decide which expert has preference, and adapt the weights for the experts accordingly. This procedure can be easily generalized to nodes $n_{\overline{p}}$ with history $\overline{p}$.

To cope with possible shortcuts, for example when a choice is made for a node at three levels deep, we must normalize the path, by inserting the intermediate node, in order to allow for comparison between experts.

Now assume that we have expert navigation paths with cycles, for example $n_{\overline{p}} \rightarrow n_{\overline{p}1} \rightarrow n_{\overline{p}13}$, where actually $n_{\overline{p}} = n_{\overline{p}13}$, which happens when we return to the original node. In general such cycles should be eliminated, unless they can be regarded as an essential subtour. However, in this case, they could also be offered explicitly as a subtour, if they have length $\geqslant 4$.

When offering guided tours for which several variants exist, we may allow the user to simply assign weights to each of the experts from which we have a tour, or allow for incrementally adjusting the weight of the experts, as feedback on the actual tour presented.

INTELLIGENT GUIDANCE – REALIZATION

Our aim is to arrive at a general framework for artist's digital dossiers, that provide intelligent guidance to both the expert user, responsible for the future re-installation of the work(s), and the interested layman, that wishes to get acquainted with a particular work or collection of works. In general, there are two techniques that we can apply to provide such guidance:

- filtering the information space according to the user's perspective, and
- intelligent agents, that (pro) actively aid the user in searching the information space.

Filtering the information space may be used to restrict the concept graph that defines the navigation structure, by stating assumptions with respect to the relevance of particular categories from a user's perspective.

Intelligent agents is an approach stemming from artificial intelligence which allows for providing guidance in a variety of ways, possibly even in an embodied form using a face or humanoid figure to give suggestions to the user on what interactions to perform, an approach that we will discuss later on.

For selecting the items to be presented in a guided tour, the most obvious way is to pre-define a sequence based on user profiles. Very likely this can be done in a more flexible way in a rule-based manner, applied to a template tour. More interesting, however, is to generate guided tours dynamically based on tracking actual user interaction of (expert) users, using techniques from prediction theory, as explained in the previous sections.

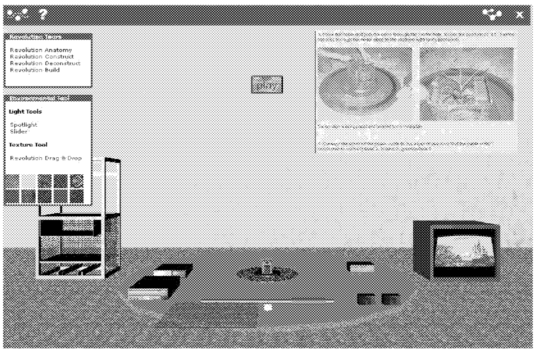


Fig. 4: Construction tool

A special case of a guided tour is the tool environment constructed for the *Revolution* installation of Jeffrey Shaw, which allows for experimenting with the (de-) construction of the installation, fig. 4, and exhibition parameters, fig. 5.

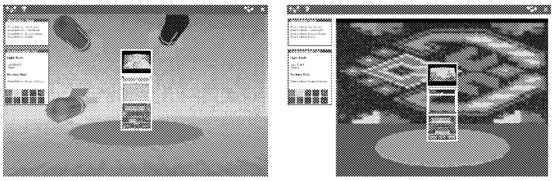


Fig. 5: (a) Light (b) Material

Tracking interaction with such 3D models is, given the limitations imposed by the tool environment, relatively simple, and can be used for creating a repository of navigation sequences. More difficult, however, is to find proper normalizations for these interactions, and so in this case we may possibly have to rely on expert weighting only.

agent technology In Hoorn et al. (2004) we have investigated the use of embodied agents in a digital dossier for the artist Marinus Boezem, fig. 6. To allow for a discrete mode of navigation we have used a map, displaying the interesting parts of the atelier, which contains locations where relevant information can be obtained, such as a filmprojector, for displaying interviews, a cabinet that contains biographical material and textual descriptions of the artworks, and an exhibition environment that displays (3D models of the) artworks. To construct a guided tour, we deployed a humanoid agent that shows the user around.

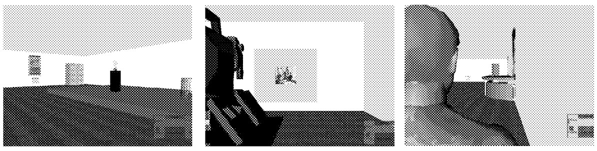


Fig 6. Overview Interview Agent

In a user evaluation test we found that humanoid agents where instrumental in providing information about the re-installation of artworks, but interestingly also that believability was positively affected by the degree of realism of the agent, Van Vugt et al. (2006) . However, in creating guided tours for the current generation of digital dossiers, using concept graphs for navigation instead of a spatial metaphor, we will not use humanoid agents. Our agent technology, however, can be used in a fruitful way.

In the I-GUARD[5] project (Intelligent Guidance in Archives and Dossiers). we investigate how to realize *advice functions*, implemented using agent technology, Eliens et al. (2002), based on actual navigation paths obtained by tracking expert users, that offer the user at any navigation point a choice of continuations and/or a selection of guided tours, focussing on a topic of interest.

INCREMENTAL ADAPTATION OF RECOMMENDATIONS

In the CHIP[6] project (Cultural Heritage Information Personalization), the aim is to develop a recommender system that generates a collection of artworks in accordance with the users' preferences based on the rating of a small sample of artworks. The properties on which the recommendation is based include *period*, *artist*, and *genre*. The recommender system will also be used to generate guided tours, where apart from the already mentioned properties the *location* (the proximity in the actual museum) will be taken into account.

Using a weighting scheme on the properties, that is a difference metric on the properties, a graph can be created, giving a prioritized accessibilty relation between each artwork and a collection of related artworks. By changing the weight for one of the properties, for example *location*, in case the tour is generated for the actual museum, the priority ordering may be changed, resulting in a different tour.

In contrast to the successor function for nodes in de concept graph of the digital dossier, we may assume to have a weighted successor function $S_w(n) = (n_1, \omega_1), \ldots, (n_k, \omega_k)$, with $\omega_i = w(n_i)$ the weight defined by the relevance of the node n_i, with respect to the attributes involved.

In a similar way as for the digital dossier, user tracking may be deployed to incrementally change the weight of the arcs of the graph, reflecting the actual preference of the user when deviating from an existing guided tour. In the remainder of this paper we will give the outline of a recommender model supporting the incremental adaptation of preferences by user feedback.

[5] www.cs.vu.nl/∼eliens/i-guard.html

[6] www.chip-project.org

SERIAL RECOMMENDER MODEL

Admittedly not the best way to do research, although common practice, we found a good starting point for modelling recommender systems, by googling on *serial recommender*, in a paper from Microsoft Research on privacy in distributed recommender systems, Oard et al. (2006). The model introduced in Oard et al. (2006), distinguishes between:

$U = user$
$I = item$
$B = behavior$
$R = recommendation$
$F = feature$

and allows for characterizing observations (from which implicit ratings can be derived) and recommendations, as follows:

- observations – $U \times I \times B$
- recommendations – $U \times I$

In a centralized approach the mapping $U \times I \times B \rightarrow U \times I$ provides recommendations from observations, either directly by applying the $U \times I \rightarrow I \times I$ mapping, or indirectly by the mapping $U \times I \rightarrow U \times U \rightarrow I \times I$, which uses an intermediate matrix (or product space) $U \times U$ indicating the (preference) relation between users or user-groups. Taken as a matrix, we may fill the entries with distance or weight values. Otherwise, when we use product spaces, we need to provide an additional mapping to the range of $[0, 1]$, where distance can be taken as the dual of weight, that is $d = 1 - w$.
In a decentralized approach, Oard et al. (2006) argue that it is better to use the actual features of the items, and proceed from a mapping $I \times F \rightarrow U \times I \times R$. Updating preferences is then a matter of applying a $I \times B \rightarrow I \times F$ mapping, by analyzing which features are considered important.
For example, observing that a user spends a particular amount of time and gives a rating r, we may apply this rating to all features of the item, which will indirectly influence the rating of items with similar features.

$B =$ [time = 20sec, rating = r]
$F =$ [artist = rembrandt, topic = portrait]
$R =$ [artist(rembrandt) = r, topic(portrait) = r]

Oard et al. (2006) observe that B and R need not to be standardized, however F must be a common or shared feature space to allow for the generalization of the rating of particular items to similar items.
With reference to the CHIP project, mentioned in the previous section, we may model a collection of artworks by (partially) enumerating their properties, as indicated below:

$A = [\ p_1, p_2, ...\ ]$
where $p_k = [\ f_1 = v_1,\ f_2 = v_2,\ \ldots\ ]$

with as an example

$A_{nightwatch} =$ [artist=rembrandt, topic=group]
$A_{guernica} =$ [artist=picasso, topic=group]

Then we can see how preferences may be shared among users, by taking into account the (preference) value adhered to artworks or individual properties, as illustrated in fig. 7.

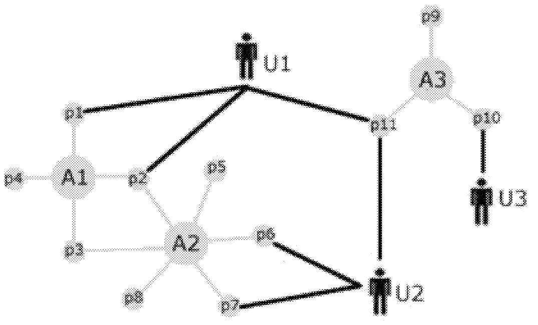


Fig 7. Users, artworks and properties

As a note, to avoid misunderstanding, Picasso's Guernica is not part of the collection of the Rijksmuseum, and does as such not figure in the CHIP studies. The example is taken, however, to clarify some properties of metrics on art collections, to be discussed in the next section.

CONTENT METRICS

To measure similarity, in information retrieval commonly a distance measure is used. In mathematical terms a distance function $d : X \rightarrow [0, 1]$ is distance measure if:

$d(x, y) = d(y, x)$
$d(x, y) \leqslant d(x, z) + d(z, y)$
$d(x, x) = 0$

From an abstract perspective, measuring the distance between artworks, grouped according to some preference criterium, may give insight in along which dimesnion the grouping is done, or in other words what attributes have preference over others. When we consider the artworks

$a_1 =$ [artist = rembrandt, topic = self-portrait]
$a_2 =$ [artist = rembrandt, name = nightwatch]
$a_3 =$ [artist = picasso, topic = self-portrait]
$a_4 =$ [artist = picasso, name = guernica]

we may, in an abstract fashion, deduce that if $d(a_1, a_2)<d(a_1, a_3)$ then $r(topic)<r(artist)$, however if $d(a_1, a_3)<d(a_1, a_2)$ the reverse is true, that is then $r(artist)<r(topic)$. Somehow, it seems unlikely that a_2 and a_4 will be grouped together, since even though their topic may considered to be related, the aesthetic impact of these works is quite different, where *selfportrets* as a genre practiced over the centuries indeed seem to form a 'logical' category. Note that we may also express this as $w(artist)<w(topic)$ if we choose to apply weights to existing ratings, and then use the observation that if $d(a_1, a_3)<d(a_1, a_2)$ then $w(artist)<w(topic)$ to generate a guided tour in which a_3 precedes a_2.

For serial recommenders, that provide the user with a sequence of items $\ldots, s_{n-1}, s_n, \ldots$, and for s_n possibly alternatives $a_1, a_2, \ldots$, we may adapt the (implied) preference of the user, when the user chooses to select alternative a_k instead of accepting s_n as provided by the recommender, to adjust the weight of the items involved, or features thereof, by taking into account an additional constraint on the distance measure. Differently put, when we denote by $s_{n-1} \mapsto s_n/[a_1, a_2, \ldots]$ the presentation of item s_n with as possible alternatives $a_1, a_2, \ldots$, we know that $d(s_{n-1}, a_k)<d(s_{n-1}, s_n)$ for some k, if the user chooses for a_k In other words, from observation B_n we can deduce R_n:

$$\begin{aligned} B_n &= [\ \text{time} = \text{20sec, forward} = a_k\ ] \\ F_n &= [\ \text{artist} = \text{rembrandt, topic} = \text{portrait}\ ] \\ R_n &= [\ d(s_n, a_k)<d(s_n, s_{n+1})\ ] \end{aligned}$$

leaving, at this moment, the feature vector F_n unaffected. Together, the collection of recommendations, or more properly revisions R_i over a sequence S, can be solved as a system of linear equations to adapt or revise the (original) ratings. Hence, we might be tempted to speak of the *R4* framework, *rate, recommend, regret, revise*. However, we prefer to take into account the cyclic/incremental nature of recommending, which allows us to identify revision with rating.

MEASURES FOR FEEDBACK DISCREPANCY

So far, we have not indicated how to process user feedback, given during the presentation of a guided tour, which in the simple case merely consists of selecting a possible alternative. Before looking in more detail at how to process user feedback, let us consider the dimensions involved in the rating of items, determining the eventual recommendation of these or similar items. In outline, the dimensions involved in rating are:

- positive vs negative
- individual vs community/collaborative
- feature-based vs item-based

Surprisingly, in Wang et al. (2007) we found that negative ratings of artworks had no predictive value for an explicit rating of (preferences for) the categories and properties of artworks. Leaving the dimension *individual vs community/collaborative* aside, since this falls outside of the scope of this paper, we face the question of how to revise feature ratings on the basis of preferences stated for items, which occurs (implicitly) when the user selects an alternative for an item presented in a guided tour, from a finite collection of alternatives.

A very straightforward way is to ask explicitly what properties influence the decision. More precisely, we may ask the user why a particular alternative is selected, and let the user indicate what s/he likes about the selected alternative and dislikes about the item presented by the recommender. It is our expectation, which must however yet be verified, that negative preferences do have an impact on the explicit characterization of the (positive and negative) preferences for general artwork categories and properties, since presenting a guided tour, as an organized collection of items, is in some sense more directly related to user goals (or educational targets) than the presentation of an unorganized collection of individual items. Cf. Van Setten (2005).

So let's look at $s_{n-1} \mapsto s_n/[a_1, a_2, \ldots]$ expressing alternative selection options $a_1, a_2, \ldots$ at s_n in sequence $S = \ldots, s_{n-1}, s_n$. We may distinguish between the following interpretations, or revisions:

- neutral interpretation – use $d(s_n, a_k)<d(s_n, s_{n+1})$
- positive interpretation – increase $w(\mathit{feature}(a_k))$
- negative interpretation – decrease $w(\mathit{feature}(s_{n+1}))$

How to actually deal with the revision of weights for individual features is, again, beyond the scope of this paper. We refer however to Eliens (2000), where we used feature vectors to find (dis)similarity between musical fragments, and to Schmidt et al. (1999), on which our previous work was based, where a feature grammar is introduced that characterizes an object or item as a hierarchical structure, that may be used to access and manipulate the component-attributes of an item.

CONCLUSIONS

In this paper we have shown how to adapt guided tours based on tracking expert users by modifying the weights attached to the experts that contributed to the construction of this tour. The application of these techniques requires that choices are discrete and hence do not apply to arbitrary navigation in virtual environments, unless we find proper ways to encode such navigation as a small finite collection of discrete steps. Also in the discrete case, however, we must be able to normalize navigation paths, in order to compare and weigh the contribution of the experts involved.

We have generalized our approach to a wider class of serial recommenders, and indicated how to apply the revision of ratings in an incremental fashion to adapt an existing tour to personal preferences, reflecting the actual navigation behavior of users.
As future work, we wish to investigate how we can use both positive and negative user feedback to revise and refine ratings for the actual features involved. Additionally, we would like to study features not directly related to artworks, but for example to group norms or personal likes and dislikes.

ACKNOWLEDGEMENT(S) We thank the reviewer who valued our original submission, which was meant to be a short paper, so positively that we were invited to write an extended paper. We took the challenge and hopefully not to the disappointment of our anonymous reviewer.

AUTHOR BIOGRAPHY

ANTON ELIENS is coordinator of the multimedia curriculum at the computer science department of the Faculty of Sciences of the Vrije Universiteit Amsterdam, He has written numerous papers and published books on distributed logic programming and object oriented software engineering.

YIWEN WANG is Ph.D. researcher in the CHIP project from the Technische Universiteit Eindhoven. She does her research at the Rijksmuseum Amsterdam, together with the other members of the CHIP team.

REFERENCES

Aroyo, L., Rutledge, L., Brussee, R., de Bra, P., *Gorgels, P., Stash N., Veenstra, M. (2005), Personalized Presentation and Navigation of Cultural Heritage Content*, Multimedia and Expo, ICME 2005. IEEE International Conference, 2005

Cesa-Bianchi N. and Lucosi G. (2006), *Prediction, Learning, and Games*, Cambridge University Press

Eliens A. (2000), *Principles of Object-Oriented Software Development*, Addison-Wesley Longman, 2nd edn.

Eliens A., Huang Z., and Visser C. (2002), *A platform for Embodied Conversational Agents based on Distributed Logic Programming*, In: *Proc. AAMAS Workshop – Embodied conversational agents - letś specify and evaluate them!*, Bologna 17/7/2002

Eliens A., van Riel C., Wang Y. (2006), *Navigating media-rich information spaces using concept graphs – the abramovic dossier*, In *Proc. InSciT2006*, 25-28 Oct. 2006, Merida, Spain

Hoorn J., Eliens A., Huang Z., van Vugt H.C., Konijn E.A., Visser C.T. (2004). *Agents with character: Evaluation of empathic agents in digital dossiers*, In: *Proc. AAMAS 2004*, Emphatic Agents, New York 19 July - 23 July, 2004

Oard D.W., Leuski A. and Stubblebine S. (2006), *Protecting the Privacy of Observable Behavior in Distributed Recommender Systems*, In: *Proc. SIGIR 2003*

Riel C. van, Eliens A., Wang Y. (2006), *Exploration and guidance in media-rich information spaces: the implementation and realization of guided tours in digital dossiers*, In: *Proc. InSciT2006*, 25-28 Oct. 2006, Merida, Spain

Riel C. van, Wang Y. and Eliens A. (2006b), *Concept map as visual interface in 3D Digital Dossiers: implementation and realization of the Music Dossier*, In: *Proc. CMC2006*, Costa Rica, Sept 5-8 2006

Robertson G.G. and MacKinlay J.D. (1991), *Cone trees: animated 3D visualizations of hierarchical information*, In: *Proc. of the SIGCHI 1991*, pp. 189-194

Schmidt, A.R. Windhouwer M.A., Kersten M.L, (1999). *Indexing real-world data using semi-structured documents*, CWI Report INS-R9902

Subrahmanian V.S. (1998), *Principles of Multimedia Databases*, Morgan Kaufmann

Van Setten M. (2005), *Supporting People in Finding Information – Hybrid recommender Systems and Goal-based Structuring*, Ph.D. Thesis, Telematica Institute Netherlands

Van Vugt, H. C., Konijn, E. A., Hoorn, J. F., Keur, I., & Eliens, A. (2006). *Realism is not all! User Engagement with Task-Related Interface Characters*, Interacting with Computers, 2006

Wang Y., Eliens A., van Riel C. (2006), *Content-oriented presentation and personalized interface of cultural heritage in digital dossiers*, In: *Proc. InSciT2006*, 25-28 Oct. 2006, Merida, Spain

Wang Y., Aroyo L., Stash N., Rutledge L. (2007), *Interactive User Modeling for Personalized Access to Museum Collections – The Rijksmuseum Case Study*, accepted for *11th Conf. on User Modeling, UM 2007*, June 25-29, Corfu, Greece

A Hybrid Multi Agent System Architecture for Distributed Supervision of Chronic Patients in the eHealth Setting

Olivier A. Blanson Henkemans[1,2]
Stefano Bonacina[3]
Nicola Cappiello[3]
Charles A.P.G. van der Mast[1]
Mark. A. Neerincx[1,2]
Francesco Pinciroli[3]

[1]Delft University of Technology
Mekelweg 4
2628 CD Delft
E-mail:
{O.A.BlansonHenkemans|
C.A.P.G.vanderMast} @TUDelft.nl

[2]TNO
Kampweg 5
3769 GZ Soesterberg
E-mail:
Mark.Neerincx@TNO.nl

[3]Politecnico di Milano
Piazza Leonardo da Vinci, 32
20133 Milano
E-mail:
{Nicola.Cappielo|
Francesco.Pinciroli|
Stefano.Bonacina} @Polimi.it

ABSTRACT

eHealth, the use of Information and Communication Technology (ICT) in the health sector, enhances today's health care environment. In particular, the use of Multi Agent System (MAS) technology, an aspect of ICT, can further contribute to the improvement of health care. Exceptional integration of this technology requires a Cognitive Engineering (CE) approach. Implying that the design and implementation of the architecture is done incrementally and using multiple distributed agents, which are facilitated by easy data entry, management and verification by all involved actors. Consequently, we designed a hybrid agent architecture, consisting of multiple distributed agents and a Virtual Personal Assistant that supervises patients' self-care, with chronic illness, and tested it in a laboratory setting according to the Cognitive Engineering approach. Results showed that the architecture designed meets with our requirements and the functionalities are comprehensible by the involved users. We expect that the incremental character of the developed hybrid architecture enables further development and could be applicable for the supervision of a variety of chronic diseases in the eHealth setting.

KEYWORDS

eHealth, Multi Agents System, Virtual Personal Assistants, Hybrid Agent Architecture, Cognitive Engineering,

INTRODUCTION

eHealth, the application of Information and Communication Technologies (ICT) in the health sector, is radically changing today's health care (Curry et al., 2002). Due to the combination of decreasing health care costs, exponentially increasing network connection speeds, and the suitability of eHealth to support patients' decisions-making and "supervised autonomy", the application of eHealth has the capacity to considerably increase the availability of self-care options (Leventhal et al., 2004). Self-care is defined by Bhuyan as activities individuals, families, and communities undertake with the intention of enhancing health, preventing disease, limiting illness, and restoring health, which in turn can improve a patient's lifestyle, medical adherence, and future health outcome (Bhuyan, 2004). An example of eHealth is the use of ICT in diabetes care. Rule-based reasoning is combined with Multi Agent Systems (MAS) technology (Sycara, 2001; Wooldridge, 2002) using mathematical models of blood glucose regulation for the identification of problems and treatment generation. Subsequently, treatment can be prescribed, including insulin therapy, diet and physical exercise (Blanson Henkemans, 2006; Haan et al., 2005).

For people with a chronic disease, eHealth solutions using Multi Agent System (MAS) technology can provide the following functionalities, concerning the personalized assistance for multiple distributed actors, e.g., patients, physicians, and other medical specialists (Xiao et al., 2006; Tonino et al., 2002; Lindenberg et al. 2003; European Commission, 2003). These solutions include improving relationships between clients and caregivers, detecting adverse trends in health proactively, stimulating a patient's motivation and bring about behavioral change. The latter is required for effective self-care, patient's quality of life, information sharing with the involved medical specialists, and productive and low-cost self-care.

To provide the above mentioned functionalities, we designed an architecture, consisting of Multi Agent System (MAS) technology. The design took place in the framework of the SuperAssist project, a collaboration between Delft University of Technology, TNO and LUMC (Haan et al., 2005). To facilitate computer-supported task performance by increasing insight into the cognitive factors, such as reasoning, quantitative knowledge, and short-term memory (Carroll, 1993) of human-computer interaction, we applied a Cognitive Engineering (CE) approach (Neerincx & Lindenberg, 2007). First, we wanted an architecture that enables designing and implementing incrementally, implying that the designed MAS architecture should be suitable for further expansion, concerning distributed actors, both human and machine, tools, e.g., medical devices, and data sources, e.g., electronic patient records, and should be adaptable to

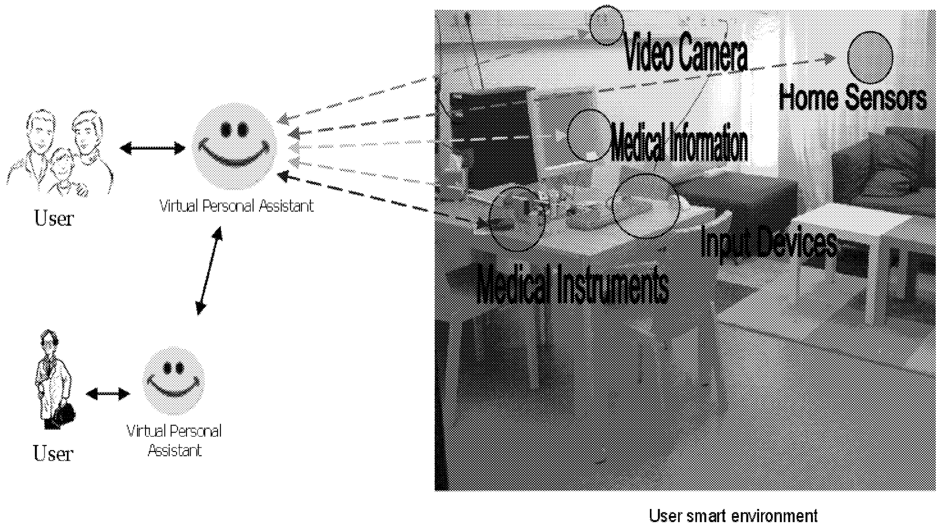


Figure 1: Multi-Agent System (MAS) working in a smart environment and interact with the users through their Virtual Personal Assistants (VPA)

diverse situations, which too are also dynamic with time. Consequently, the intelligence of the system should be added step by step. Second, we wanted multiple distributed agents that could act both as independent intelligent actors in the network managing the data and also as assistants for the involved human actors, e.g., patients, medical specialists, and technical specialists, by supporting them in their complex task environment. Third, we wanted a system that enables easy data manipulation, including data entry, management and review, because all actors involved should be able to perform each of these actions.

Several projects (e.g., Bellazzi et al. 2003; Camarina-Maots et al., 2003; DbMotion, 2005) offer well-founded agent architectures, but also have a number of shortcomings, in terms of the requirements mentioned above (Table 1). In short, the existing architectures apply a one-side centralized or decentralized approach. This implies that the data is either collected from the actors in a network, converted to a single format, and stored in one database that serves all its participants, or that the actors maintain ownership of their data, which then has to be retrieved by other actors.

To tackle these disadvantages, we designed a hybrid architecture, consisting of a decentralized system with a central node. This node can manage data that the actors decide to synchronize, e.g., medical test results and the medical specialists' calendar. The same node can contain a web server, a communication server and other central systems that need a centralized structure. The hybrid architecture enables the use of multiple distributed agents that both act as independent intelligent actors in the network that manage the data and as assistants to the users. Furthermore, this hybrid approach makes the architecture suitable for further expansion and adaptation.

In this paper, we will discuss the hybrid architecture designed to address the shortcomings mentioned. In addition, we will report the results of a qualitative experiment conducted to test the hybrid architecture's functional capacity and whether or not it fulfils the requirements of the Cognitive Engineering (CE) approach, concerning designing and implementing the architecture incrementally, using multiple distributed agents, and easy data entry, management and checking by all the involved actors. Finally, we will discuss the implications of the results.

Table 1: Disadvantages of centralized and decentralized agent architectures

Centralized architecture	Decentralized architecture
Constrained flexibility in distributed database management	Difficult to integrate and interoperate different platforms
High dependency on difficult accessible internal network	Distributed data sources and agents are required to be online for data synchronization
All the computation must take place centrally	Indistinctness of who is managing the data
	Prone to data redundancy

2. A HYBRID ARCHITECTURE IN AN FOR THE SUPERIVISON OF DIABETES PATIENTS

The main goal was to design a hybrid Multi Agent System architecture suitable for the supervision of diabetes patients in an eHealth setting according the Cognitive Engineering approach. In our architecture, agents are working in the background providing ambient intelligence to the users who reside in the smart environment. A Virtual Personal Assistant (VPA) acts as mediator between the users and agents active in the smart environment (Lindenberg et al. 2003; Maes, 1998; Grill et al., 2005). In the intelligent environment, the agents communicate with each other and receive information through sensors placed in the environment. The multiple agents acquire information through sensors, their behavior or their communication. The VPA shares data with the agents and interacts with the user. To test our architecture, we

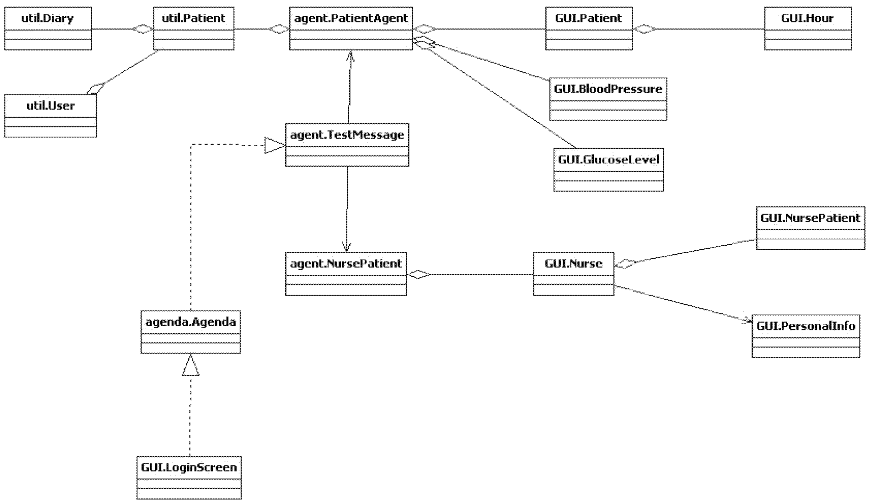


Figure 3: UML case diagram, diagram Virtual Personal Assistant (VPA) software

implemented a hybrid architecture, consisting of Multi-Agent System (MAS) technology and Virtual Personal Assistants (VPA) (Figure 1). Both the patient and the medical specialist have a VPA enabling them to interact with the Multi-Agent System. The Multi Agent System receives data derived from the patient's electronic patient record, electronic diabetes diary, and domestic medical instruments, e.g., glucometer. Based on this data, the patient's Virtual Personal Assistant interacts with the patient, e.g., about his or her health status and diseases. In addition, it updates the medical specialist's VPA, which in turn informs the medical specialist of the patient situation and supports managing basic medical data.

All the data that appears in the Diary is also stored in a database. The database is initialized with MySQL and managed by the java file DatabaseMediator.java of the library agendaDb.db (Figure 2). In the Unified Modeling Language (UML) case diagram (Figure 3), we represent the main tasks that every actor can perform with the system. There are four categories: objects needed for the actual communication with the database (agendaDb.db), objects needed to store data (agendaDb.util), objects needed for the graphical ser interface (agendaDb.GUI), and objects needed for the implementation of the agents (agendaDb.agent).

The software project is structured in five packages according to the general use of the classes contained. The agendaDb.agenda contains only one class system, which verifies if the user is in the database, and according with his group initiates a TestMessage object.

2.1 Agent Communication and Database

To enable optimal functionality, different agents run simultaneously on the patient and medical specialist's computers and communicate with each other. We have implemented two agents using the JADE platform, i.e., the *PatientAgent* on the patient's computer, and the *NurseAgent* on the medical specialist's computer. The agent communication takes place by sending messages from both sides. In order to recognize the different kind of messages, the behavior of the agents changes according to the prefix of the message. If the message starts with "start", the chat becomes active; if it starts with "chat", the chat window will show the message sent by the medical specialist. The third type of prefix is "response" followed by the name of the measurement, i.e. blood pressure or glucose level. Also, the *PatientAgent* checks how the new measurements influence the status of the patient.

2.2 Virtual Personal Assistant Activities

The Virtual Personal Assistants enable the involved human actors to interact with the different agents in the environment. This is realized through the performance of three main activities: data entry, formulate policies and make recommendations.

The medical specialist enters data regarding patient information, e.g., demographic data, medical history, and clinical diabetes information. The patient keeps track of self-care tasks in a personal electronic diabetes diary, including current mood, exercises performed, meals consumed, medication taken, and blood glucose measurement results. The data are entered by the patient and the medical specialist, through their Virtual Personal Assistants (VPAs) into a MySQL database.

The Virtual Personal Assistant gives short- and long-term suggestions, or so-called policies. This is based on the data in the electronic diabetes diary and electronic patient record. The VPA predominantly determines its prescribed policies after monitoring the results of the glucose test results, due to its importance in diabetes care.

The connection of the VPAs with the database is done with the Java Database Connectivity (JDBC) platform and the Application Program Interface (API), a programming interface allowing external access to the database manipulation and update commands. It allows the integration of SQL calls into a general programming environment by providing library routines, which interfaces with the database.

In theory, all actors have their database and the database is synchronized with a central node. In our architecture, the database, containing the electronic patient record and the electronic diabetes diary, is located on the patient's side. Thus when the patient is online, the database and current patient data is accessible by the medical specialist. When the patient is offline, the medical specialist has access to the patient's data retrieved the last time the patient was online and the databases were synchronized.

2.3 User Interface

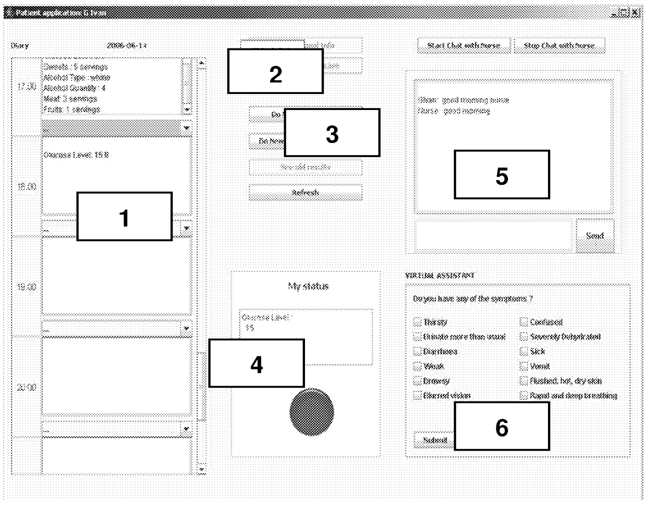
Both the patient and medical specialist have an interface in which their Virtual Personal Assistants are integrated. In this section we will shortly discuss the interfaces. The patient's interface consists of six fields (Figure 4):

1. An electronic diabetes diary in which the patient logs the performed self-care tasks;
2. Access to the Electronic Patient Record (EPR);
3. Entering medical test results and viewing old medical results;
4. A "traffic light" indicating current health status based on diary and last measurement results (green: healthy, orange: be aware, red: alert!);
5. A chat service to communicate with the remote medical specialist, and;
6. A frame to communicate with the Virtual Personal Assistant.

The medical specialist also has an interface used for remote monitoring of the patients. This interface can be divided in three main parts (Figure 5):

1. List of the patients in the medical specialist's folder with the current health status represented by the accompanying traffic light.
2. Access to Patient Data Management, and;
3. A chat service to communicate with the patient.

- In the UML case diagram (Figure 3), the interface classes (GUI.x) are represented by:
- *GUI.Patient*: The main window of the interface from the patient side;
- *GUI.Nurse*: The main window of the interface from the medical specialist side;
- *GUI.BloodPressure* and *GUI.GlucoseLevel* are linked to *agent.PatientAgent* through the ActionEvent functions. When the user pushes the button '*Measurement done: update agenda*' the Patient Agent is announced to start the communication with the NurseAgent;
- *Util package*: Contains the info that is later stored in the database or retrieved from the database. Also in the *util.Diary* class, the thresholds for the measurements are verified and according to the new status the *GUI.Patient* shows the correct frames of the Virtual Personal Assistant.



Sweets : 5 servings
Alcohol Type : white
Alcohol Quantity : 4
Meat 3 servings
Fruits 1 servings

Glucose Level: 18.8

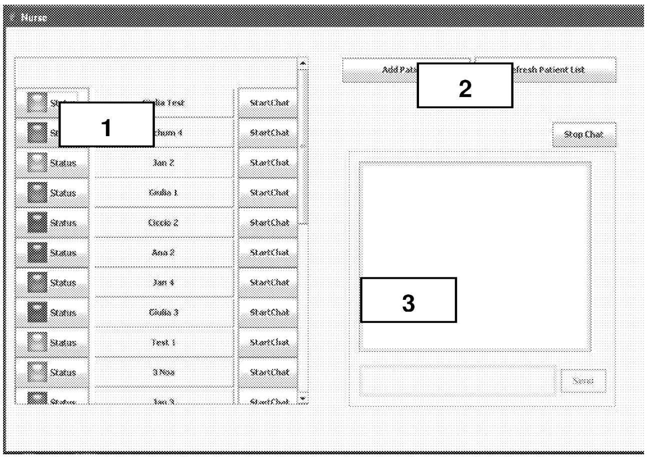
[illegible]: good morning nurse
Nurse: good morning

Glucose Level:
15

Do you have any of the symptoms ?

☐ Thirsty ☐ Confused
☐ Urinate more than usual ☐ Severely Dehydrated
☐ Diarrhoea ☐ Sick
☐ Weak ☐ Vomit
☐ Drowsy ☐ Flushed, hot, dry skin
☐ Blurred vision ☐ Rapid and deep breathing

Figure 4: Patient Interface, containing 1. Electronic Diabetes Diary, 2. EPR access, 3. Retrieving tests results, 4. Traffic light, 5. Chat service, and 6. VPA communication frame.



Status	Giulia Test	StartChat
Status	...hum 4	StartChat
Status	Jan 2	StartChat
Status	Giulia 1	StartChat
Status	Cicolo 2	StartChat
Status	Ana 2	StartChat
Status	Jan 4	StartChat
Status	Giulia 3	StartChat
Status	Test 1	StartChat
Status	3 Noa	StartChat
Status	Jan 3	StartChat

Figure 5: Medical Specialist Interface, containing 1. Patient List, 2. Access to Patient Data Management, and 3. Chat service.

3. METHOD: A QUALITATIVE EXPERIMENT

To test the developed hybrid architecture, consisting of Multi Agent System technology and the Virtual Personal Assistants, according the Cognitive Engineering requirements, we conducted a qualitative experiment. We invited eleven participants (nine students and two older adults) at Delft University of Technology (DUT) and TNO's Experience labs to evaluate the system. These labs enable testing prototypes of new technologies by offering a comfortable domestic atmosphere and encourage natural behavior in an experimental setting, while contextualizing the experience (Blanson Henkemans et al., 2007).

Because we wanted to evaluate the system before actually applying it in the field, we tested the system in the lab setting and also asked the participants to play the role of the patient according to predefined scenarios. A scenario is a description that contains actors, background information on the actors, their assumptions about their environment, actors' goals or objectives, and sequences of actions and events (Rosson & Carroll, 2001). Following the scenario-based design method, we used these scenarios as a general representation throughout the entire system lifecycle (analysis, design, prototyping and evaluation).

During the experiment, we observed the participant remotely with both an overview camera and a close-up camera. In addition, we had access to the medical specialist interface and a replica of the patient's interface. Finally, we logged the communication between the two agents. Consequently, we received a good overview of the tasks performed by the participant, the communication between the agents, and the functionality of the two Virtual Personal Assistants. Finally, we surveyed the participants' opinion on the VPA's usability.

4. RESULTS

When both Virtual Personal Assistants were online, they could correctly add and retrieve data to and from the database, containing the electronic patient record (EPR) and the electronic diabetes dairy. In the case of a health critical situation or when the patient sent a message to the medical specialist, the two VPAs could communicate with each other directly. The interaction and communication processes are illustrated in Figure 5. In practice, the patient logged self-care tasks in the electronic diary and based on the data, the patient's VPA made inferences about the patient's health status and gave feedback. The new data, based on the interaction between the patient and its VPA, was then sent to the database and retrievable by the medical specialist's VPA.

During the experiment, we observed that the data were accessible and manageable by the authorized actors. The patient and its Virtual Personal Assistant could edit the electronic diabetes dairy and view the electronic patient record. The medical specialist could add new patients to the database, edit the electronic patient record, and view the patient's diabetes diary and current health status. The medical specialist's VPA could view the patient electronic diabetes diary, electronic patient record and current health status.

When the patient's VPA was offline, it could still give comment on the diary entry and subsequently add the new data to the database. When the patient's VPA was online again, the medical specialist's VPA could synchronize the new data.

The patient's VPA reacted accurately and quickly to the newly added data of the patient electronic diary independent of the type of task or the health situation of the patient. The data was correctly added to the database and retrieved by the medical specialist's VPA. Also, the communication between the two VPAs was instantaneous. According to the usability survey, on a scale of 1 through 7 (7 being very usable) participants rated the usability of the assistant a 6.5.

5. DISCUSSION

The results depict that the designed hybrid architecture meets with our SuperAssist project requirements, concerning designing and implementing the architecture incrementally, using multiple distributed agents, and easy data entry, management and checking by all the involved actors. These functionalities work adequately and are comprehensible by the involved users through a usable Virtual Personal Assistant.

However, the system needs improvement on several aspects. One aspect we need to improve is the agents' intelligence and awareness of processes. Currently, the Virtual Personal Assistants only monitor newly added data and do not take into consideration existing data. For example, with diabetes patients, but also with other chronic patients, e.g., people with asthma, the course of their health is as important as their current health state. Another aspect is the necessity to elaborate on privacy and security issues, such as risks of corporate espionage, consumer/personal privacy validation, and location privacy [Xiao et al., 2006]. If we want to deploy the hybrid architecture in real practice, we will have to further study the problems and restrictions these issues pose. Finally, we suggest adding a technical specialist to the architecture. The technical specialist, also equipped with a Virtual Personal Assistant, could help maintain and troubleshoot technical problems with both the system itself and the domestic medical instruments used by the patient.

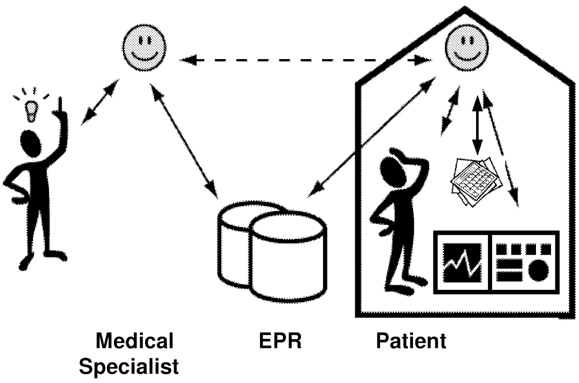


Figure 5: Virtual Personal Assistants for the supervision of patient's self-care

In conclusion, we expect that the expansive character of the developed hybrid architecture facilitates implementing these improvements. In addition, it could be applicable for the supervision of a variety of chronic diseases in the eHealth setting.

ACKNOWLEDGMENTS

We would like to thank Oana Dragomyr for her essential work on the system. The SuperAssist project is partially funded by SenterNovem IOP-MMI, a program of the Dutch Ministry of Economic Affairs.

REFERENCES

Bellazzi R., Carson E. R., Cobelli C., Hernando E., Gomez E. K., & Nabih-Kamel-Boulos M. (2001). Merging telemedicine with knowledge management: the M2DM project. In: IEEE editor. *Proceedings of 23rd Annual EMBS International Conference of the IEEE*, Istanbul, Turkey, 4, pp. 4117-4120.

Bhuyan, K.K. (2004). Health promotion through self-care and community participation: Elements of a proposed programme in the developing countries. *BMC Public Health*, 4(11).

Blanson Henkemans, O.A., Neerincx, M.A., Lindenberg, J., & Mast, C.A.P.G van der (2006). SuperAssist: Supervision of Patient Self-Care and Medical Adherence. *IEA 2006 Conference*, July 10 - 14 , 2006, Maastricht, The Netherlands.

Blanson Henkemans, O.A. Caine, K.E., Rogers, W.A., Fisk, A.D. Neerincx, M.A., & Ruyter, B. de. (2007). Medical Monitoring for Independent Living: User-Centered Smart Home

Technologies for Older Adults. Accepted for the *Med-e-Tel 2007*, April 18 - 20, Luxemburg, Luxemburg.

Camarinha-Matos L.M., Octavio Castolo L., & Rosas J. (2003) A multi-agent based platform for virtual communities in elderly care. Emerging Technologies and Factory Automation; ETFA' 2003. *Proceedings of 9th IEEE International Conference on Emerging Technologies and Factory Automation*; Lisbon (P), September 16-19; 3, pp. 421-428.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*, New York: Cambridge University Press, 1993.

Curry, R. G., Tinoco, M. T., & Wardle (2002). *The Use of Information and Communication (ICT) to Support Independent Living for Older and Disabled People*. Retrieved January, 2007 from http://www.rehabtool.com/forum/discussions/ictuk02.pdf.

DbMotion (2005). *The dbMotion Solution. A White Paper to access medical information*. Retrieved February, 2007 from http://www.ness.com/NR/rdonlyres/BCC63AB7-8C6D-402E-A923-7ED8497937FE/4757/WhitePaperFinalDec05.pdf

European Commission (2003). *The Health Status of the European Union: Narrowing the Health Gap. Office for Official Publications of the European Communities*. Luxembourg. Retrieved February, 2007 from http://ec.europa.eu/health/ph_information/documents/health_status_en.pdf

Grill, T. Khalil Ibrahim, I, & Kotsis, G. (2005). Agent interaction in ambient intelligent environments, *Proceedings of 13th DPS workshop*, November 29.-December 2, 2005, Okinawa - Japan.

De Haan G. de, Blanson Henkemans O.A., Ahluwalia A. (2005). Personal assistants for healthcare treatment at home. In: Marmaris N, Kontogiannis T, Nathanael D, editors. *Proceedings of European Association of Cognitive Ergonomics (EACE) 2005,* September 29-October 1, 2005, Crete (GR). pp. 217-223.

Leventhal H., Halm E., Horowitz C., Leventhal E. & Ozakinci G. (2004). Living with chronic illness: A contextualized, self-regulation approach. In: Sutton S, Baum A, Johnston M, editors. *The SAGE Handbook of Health Psychology*. London (UK): Sage Publications, 2004. p. 159-194.

Lindenberg J., Nagata S., & Neerincx M.A.. (2003). Personal assistant for on Line services: addressing human factors, human centered computing: cognitive, social and ergonomic aspects. *Proceedings of Human Computer Interaction International Conference*, 2003, Mahwah, NJ: Lawrence Erlbaum.

Maes P. (1998). Agents that Reduce Work and Information Overload. In: Maybury MT, Wahlster W: *Readings in Intelligent User Interfaces*, Morgan Kaufmann, 1998, pp.525-536.

Neerincx, M.A. & Lindenberg, J. (in press). Situated cognitive engineering for complex task environments. In: Schraagen, J.M. (Ed.), *Natural Decision Making & Macrocognition*. (will appear in 2007). Ashley.

Rosson M.B. & Carroll J.M. (2001) *Usability Engineering: Scenario-Based Development of Human Computer Interaction*. San Francisco, CA: Morgan Kaufmann.

Sycara K.P. (2001). Multiagent system: a modern approach to distributed artificial intelligence. *AI Magazine*, 22(2), pp. 105-108.

Tonino J.F.M., Bos A., Weerdt M.M., & Witteveen C. (2002). Replanning by Revision in Collective Agent-Based Systems. *Artificial Intelligence Journal,* 2002, 142(2), pp. 121-145.

Wooldridge, M. (2002). *An Introduction to MultiAgent Systems*, John Wiley & Sons Ltd, 2002

Xiao, Y., Shen, X., Sun, B., & Lin, C. (2006). Security and privacy in RFID and applications in telemedicine. *IEEE Communications Magazine*. 2006, 44(4), pp. 64-72.

BIOGRAPHY

OLIVIER BLANSON HENKEMANS is PhD student in Human-Machine Interaction at Delft University of Technology and is also affiliated to TNO Human Factors. The focus of his research is on personal computer assistant supervision of older diabetes patients' self-care and context aware and adaptive user-assistant interaction.

MARK NEERINCX is head of the Intelligent Interface group at TNO Human Factors, and professor in Man-Machine Interaction at the Delft University of Technology. He has extensive experience in applied and fundamental research. He has been involved in the organization of conferences, workshops and tutorials to discuss and disseminate human factors knowledge.

CHARLES VAN DER MAST has a PhD in Computer Science from Delft University of Technology. He is associate professor at the Man-Machine Interaction group at Delft University of Technology. He teaches courses on Multimodal Interfaces and Virtual Reality, Developing Highly Interactive Systems, Multimedia, Educational Software and Intelligent User Experience Engineering. He co-developed the Bachelor/Master curriculum in Media & Knowledge Engineering. His interests include using various media to improve teaching, VR therapy for phobia treatment and agent support in complex systems.

NICOLA CAPPIELLO has a Master degree in Telecommunication Engineering. He is currently working at UL International as Handler Engineer. During his master thesis, he worked at Delft University of Technology on telemedicine architecture and on and evaluated a virtual personal assistant for telediabetology applications. He graduated at Politecnico di Milano and did a one year internship at Lund University.

VEHICLE BASED MULTIMEDIA

Position Enhancement Technique Using GPS-GSM Model for Vehicle Location Systems

Jamal S. Rahhal[1] and Dia I. Abu-Al-Nadi

Electrical Engineering Dept.
The University of Jordan
Amman – Jordan
***E-MAIL:** rahhal@ju.edu.jo*

1.1. KEYWORDS:

GPS, GSM, GIS, AVL.

1.2. ABSTRACT

Global Positioning System (GPS) is used for vehicle location systems. When it detects less than four satellites, it will fail to determine the location of the vehicle. This happens mainly in crowded areas such as downtown areas where the buildings block the line of sight between the receiver and some of the satellites. When this happens the vehicle requires a more reliable system to depend on, this will be the GSM network. The proposed system makes use of the coverage of the GSM to determine the differential distance between the last known good location given by the GPS system and the current position of the vehicle. A prediction filter and a mathematical model is used to find this differential distance. This method does not require an upgrade for the GSM equipment. Experimental results showed that we could obtain the location in the shadowed areas to about 30 meters of error. This accuracy depends on the GSM environment and the differential nature of the reading.

1.3. INTRODUCTION

Global Positioning System (GPS) is a very popular system and is widely used in most modern land based transport, such as cars and trucks. Current systems have integrated relatively new technologies such as GPS, Global System for Mobile (GSM) and Geographic Information Systems (GIS). One important application for this integration is the Automatic Vehicle Location (AVL). Most of the previous research in the location determination is done using only the GPS system. The GSM network was used as a communication channel to transfer the information to a fixed server that is used to analyze the data using GIS and sends feedback to the vehicle. Better GPS readings is obtained using assistance data including differential GPS data and GSM assisted GPS where the basic idea is to establish a GPS reference network. This reference network consists of receivers that have clear views of the sky and can operate continuously. This reference network is also connected with the GSM network. At the request of a Mobile Station (MS) or network-based application, the assistance data from the reference network is transmitted to the MS to increase performance of the GPS receiver onboard. When the position is calculated at the network, it is called mobile-assisted. When the position is calculated at the handset, it is called mobile-based. Additional assisted data, such as differential GPS corrections, approximate handset location or cell base station location, and others can be transmitted to improve the location accuracy [El-Rabbany A. 2002] [Hofman *et. al.*1997] [Guorong H. and Weihong C. 2000]

The mobile-assisted solution shifts the majority of the traditional GPS receiver functions to the network. It requires an upgrade for both hardware and software of the GSM system. The network transmits a very short assistance message to the mobile station (MS), consisting of time, visible satellite list, satellite signal Doppler, and code phase search window. These assistance data are valid for a few minutes. It returns from the MS the range data processed by the GPS receiver. Then the network estimates the position of the MS in a more accurate way [Retscher G. and Mok E. 2001] [Yilin Z. 1997] [Gregory T. 1996] [Weill L. 1997].

The GPS calculates its position (*x, y, z*) by solving the range equations given by:

$$\begin{aligned}
d_1 &= \sqrt{(x-x_1)^2+(y-y_1)^2+(z-z_1)^2} = c(T_1-T_0)\\
d_2 &= \sqrt{(x-x_2)^2+(y-y_2)^2+(z-z_2)^2} = c(T_2-T_0)\\
d_3 &= \sqrt{(x-x_3)^2+(y-y_3)^2+(z-z_3)^2} = c(T_3-T_0)\\
d_4 &= \sqrt{(x-x_4)^2+(y-y_4)^2+(z-z_4)^2} = c(T_4-T_0)
\end{aligned} \qquad \mathbf{(1)}$$

where (x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3), and (x_4, y_4, z_4) are the known satellite positions, d_1, d_2, d_3 and d_4 are the measured ranges, c is the speed of light, T_1, T_2, T_3 and T_4 are the known satellite clock bias from GPS time, and T_0 is the unknown receiver clock offset from GPS time. The satellite clock bias terms are derived by the receiver from the satellite navigation message.

[1] This work was supported by the University of Jordan Deanship of Academic Research.

Errors encountered in the reading of the satellite clock bias terms will be reflected in errors in the measured location.

If the data to be collected is on Main Street of a large city, the receiver is likely to be surrounded by tall buildings that restrict satellite visibility resulting in poor location calculation since the only satellites that the receiver can see will be nearly straight up. Also, the structures all around the receiver act as nearly perfect multi-path reflectors and the receiver's multi-path rejection capability may actually be overloaded. These are very difficult problems to overcome, particularly in dense urban areas with many tall buildings. The enhancements were done on the methods of GPS readings but blockage and multi-path reflections problems remain unsolved. This is due to the fact that no way to the receiver to find a line of sight with the satellite in these crowded areas and the reflections are an external effects that the receiver can do nothing about [Rahhal J. 2001] [Sypniewski J. 2000] [Nusser S.*et.al.* 2003] [Briggs D. *et.al.* 2003].

1.4. System Description

In normal situations the GPS should detect four or more satellites to determine the location of the mobile vehicle. But in crowded areas such as down town areas where the buildings block the line of sight between the receiver and some of the satellites, the GPS fails to determine the location and loss of information occurs. Also, multipath reflections will introduce errors in the GPS readings and the measured location will be mistaken as shown in Figure 1. The GPS receivers usually need time to obtain the first point (after loss of sight), this time period will result in loss of some points when the GPS in motion. When this happen the vehicle requires a more reliable system to depend on. This will be the GSM network, such that, we will make use of the coverage of the GSM to determine the differential distance between the last known good location given by the GPS system and the current position of the vehicle. This method does not require an upgrade for the GSM equipment.

The system will integrate the measurements obtained by the GPS and GSM systems to accurately locate the mobile vehicle as shown in Figure 2. This integration will solve the problem of unreliable information received by the GPS receiver due to buildings and tree shadowing. The system will use a mathematical model to enhance the readings of the GPS and switch to differential mode when blockage happens and use the GSM system to locate the vehicle location. When reflections occur, the system will apply anon linear median filter to the readings of the GPS to reduce or eliminate the reflections effect.

A GIS system will be also integrated with the location measurement system to fix logical errors. For example, if the system determines the location, the GIS system will check to see if this point is located on a street or not. If not, a logical error is triggered and an algorithm to fix that error will be invoked such that it will find the best logical location to the vehicle to suggest.

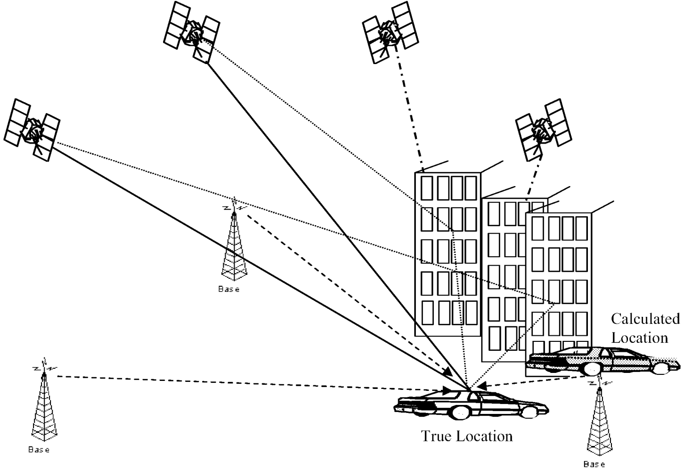


Figure 1: Downtown Situation Showing Blockage and Multi-path Effects.

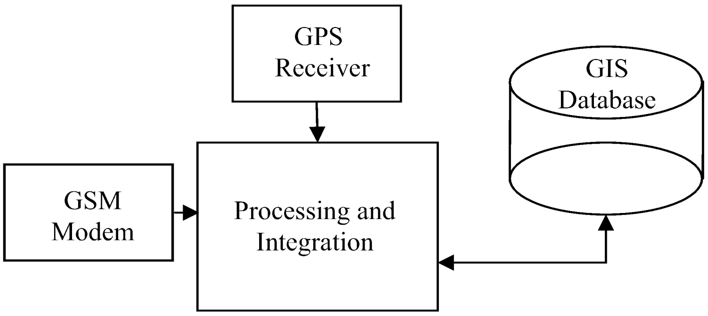


Figure 2: Block Diagram of Integrated GSM-GPS Receiver.

Prior to GIS processing, the GPS reading will be processed to reduce the multipath effect. There are two location enhancement techniques, the first one is predicting the current location and the second one is the use of path correction technique.

Prediction techniques are based on combining both the readings from GPS and GSM network. The final predicted location L_f at time n is given by:

$$L_f(n) = \mu L_{GPS}(n) + (1-\mu) L_{GSM}(n) \qquad \textbf{(2)}$$

where μ is the multipath probability that can be obtained using experimental data for each area type, $L_{GSM}(n)$ is the location found by using GSM data. Applying a linear prediction filter, the predicted GPS reading $L_{GPS}(n)$ is given by:

$$L_{GPS}(n) = P_n L_r(n) + (1-P_n)\sum_{i=1}^{I} \alpha_i L_{GPS}(n-i) \qquad \textbf{(3)}$$

where $L_r(n)$ is the current GPS reading, α_i's are the prediction filter coefficients, I is the prediction filter length, P_n is the normalized bias factor at time n and it is given by:

$$P_n = \frac{M}{6} \qquad M = \begin{cases} 0 & N_s < 3 \\ N_s & 3 \le N_s \le 6 \\ 6 & N_s > 6 \end{cases} \qquad \textbf{(4)}$$

where N_s is the number of satellites seen by the GPS receiver. The prediction coefficients α_i's can be obtained for each region by taking I actual current readings $L_r(n)$ and solving the following equation:

$$\min_{\forall \alpha} \left| L_r(n) - \left[P_n\, L_r(n) + (1-P_n)\sum_{i=1}^{I} \alpha_i L_{GPS}(n-i) \right] \right|^2 \qquad \textbf{(5)}$$

then solving for α_i's we get:

$$\alpha = L^{-1}\, L_r \qquad \textbf{(6)}$$

where:

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_I \end{bmatrix}, \; L_r = \begin{bmatrix} L_r(n) \\ L_r(n-1) \\ \vdots \\ L_r(n-I+1) \end{bmatrix},$$

$$L = \begin{bmatrix} L_{GPS}(n-1) & L_{GPS}(n-2) & \cdots & L_{GPS}(n-I) \\ L_{GPS}(n-2) & L_{GPS}(n-3) & \cdots & L_{GPS}(n-I-1) \\ \vdots & \vdots & \ddots & \vdots \\ L_{GPS}(n-I) & L_{GPS}(n-I-1) & \cdots & L_{GPS}(n-2I+1) \end{bmatrix}$$

The location obtained by the GSM $L_{GSM}(n)$ is calculated by triangulation using the received power level from the surrounding base stations and the timing advance. The triangulation is done by assuming a propagation model for the GSM radio frequency, which is given by [Rahhal J. 2001] [Sypniewski J. 2000]:

$$P_r = K_o + 10\gamma\, Log(r) \;\; dB \qquad \textbf{(7)}$$

where P_r is the received power, K_o is constant, γ is the propagation factor and r is the distance from the base station. The timing advance t_{TA} in microseconds will produce an estimate for the distance from the home base station r_{TA} as:

$$r_{TA} = 550\, t_{TA} + 275 \; m \qquad \textbf{(8)}$$

This will produce a ring with radius r_{TA} and a width of 550 meters. Solving for each r from the surrounding base stations will produce several circles. These circles and the ring produced by the timing advance equation should be intersected in a region as shown in Figure 3. If no intersection occurs then change K_o until all circles are intersected.

The calculated location from GSM $L_{GSM}(n)$ is the middle point of the intersection region between all circles and the timing advance ring.

After calculating the final predicted values of the position and saving them into an array we apply a median filter to remove the odd points such that it will first look for odd points that are not in a smooth position in the path. Then replacing that point by the median of the prior two points and two points after as shown in Figure 4.

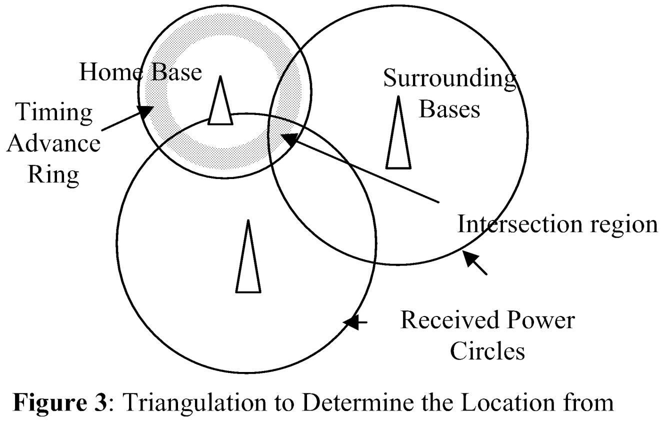


Figure 3: Triangulation to Determine the Location from GSM Data.

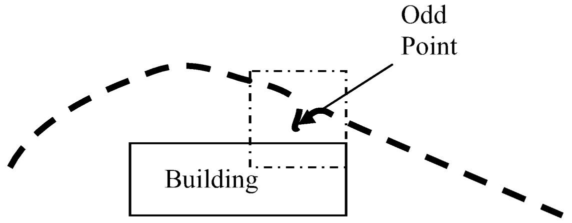


Figure 4: Odd Points along the Predicted Path.

The median filter sorts the Latitudes and Longitudes of the four points and produces the value that comes in the middle. Other smoothing techniques could be employed.

1.5. System Analysis

The performance of the devised system depends on the choice of the prediction filter as well as the fitness of the propagation model to the actual propagation environment. We start by analyzing the performance of the median filter that is used to reduce the effects of odd points. These points produced when reflections from surrounding buildings dominate the direct rays from satellites at the GPS receiver. As shown in Figure 5 the GPS reading might be in error at some discrete points while passing in a street surrounded by buildings.

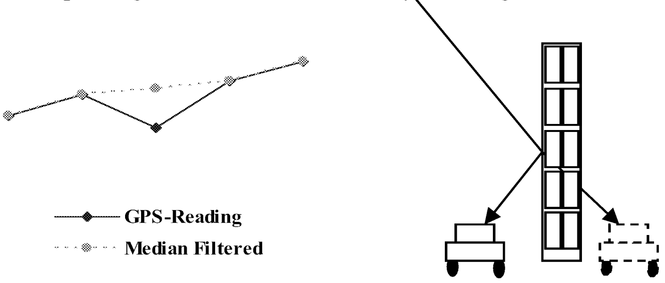


Figure 5: Reflection Effect and the Path Shape.

Odd points can be detected when the absolute value of the slope of the path from left and from right is greater than a certain threshold δ_{th} (street width deviation). This can be written as:

$$\left|\frac{\Delta y}{\Delta x}\right|^{+}\left|\frac{\Delta y}{\Delta x}\right|^{-} \geq \delta_{th} \qquad \textbf{(9)}$$

The threshold δ_{th} can be calculated from the street width W_s as:

$$\left|\frac{W_s}{\Delta x}\right|^{2} = \delta_{th} \qquad \textbf{(10)}$$

and Δx is the step length along the street.

1.6. Results and Conclusions:

An experimental test drive was conducted in a street as shown in Figure 6. We assume the multipath probability to be $\mu = 0.6$ and the propagation factor $\gamma = 2.73$ in the test area [Rahhal J. 2001].

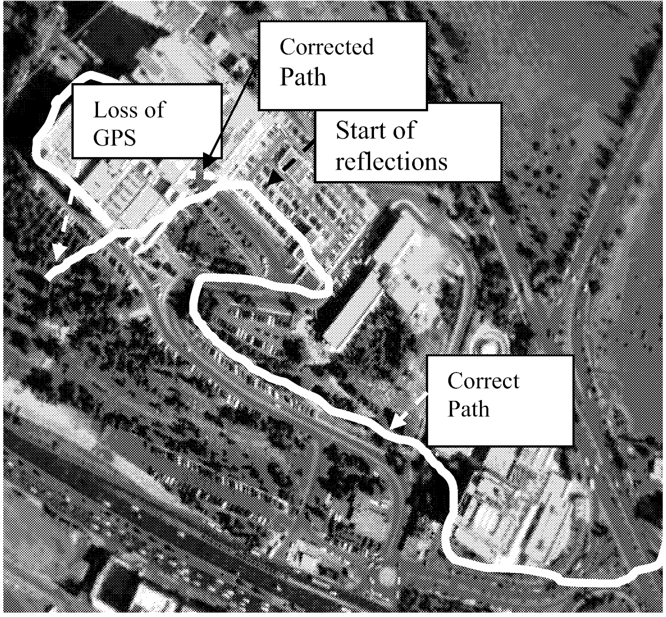


Figure 6: Aerial View of the Test Drive Street Showing the Effect of Reflection on GPS Reading.

In this example we assume the prediction length *I*=2 and solving equation(6) we find that:

$$\begin{aligned}\alpha_1 &= \lfloor \alpha_1^x \quad \alpha_1^y \rfloor = [0.589 \quad 0.328] \\ \alpha_2 &= \lfloor \alpha_2^x \quad \alpha_2^y \rfloor = [1.213 \quad -0.194]\end{aligned} \qquad \textbf{(11)}$$

Here we have several GSM base stations surrounding the path under consideration, we measure the base station locations and fed them to the algorithm. The output is then calculated for this example as shown in the dotted path.

References

El-Rabbany Ahmed, *Introduction to GPS The Global Positioning System*. Artech House, 2002.

Hofmann-Wellenhof B et al. *Global Positioning System: Theory and Practice.* Springer:New York, NY, U.S.A., 1997.

Guorong H, Weihong C. "Discussion on GPS GLONASS for vehicle navigation and Monitoring system." Engineering of Surveying and Mapping 2000; pp:19–21.

Retscher G, Mok E. "Integration of mobile phone location services into intelligent GPS vehicle navigation systems." Proceedings, The 3rd International Symposium on Mobile Mapping Technology, CD ROM S17-1, Cairo, Egypt, 3–5 January 2001.

Yilin Zhao, Vehicle Location and Navigation Systems, Norwood, MA: Artech House, 1997.

Gregory T French, *Understanding the GPS: An Introduction to The Global Positioning System*. GeoResearch, 1996.

Weill, L. R., "Conquering Multipath: The GPS Accuracy Battle," GPS World, Vol. 8, No. 4, April 1997, pp. 59-66.

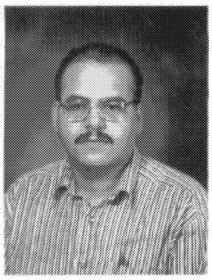
Rahhal Jamal, "Propagation Loss in Jordan for Cellular Communications," Mutah Journal for Research and Studies, 2001.

Sypniewski J. "The DSP algorithm for locally deployable RF tracking system." *Proceedings of the International Conference on Signal Processing with Applications*, Orlando, October 2000.

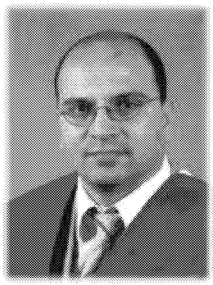
Nusser S, Miller L, Clarke K, Goodchild M. Geospatial IT for mobile field data collection. Communications of the ACM 2003; pp:63–64.

Briggs D, Gulliver J, Crookel A, Evans R, Walker M, RandolphW. "Integration of GPS and dead reckoning for real-time vehicle performance and emissions monitoring," GPS Solutions 2003; 6:229–241.

Bibliography



Jamal Rahhal received the B.S. degree from the University of Jordan in 1989, the M.S. and Ph.D. from IIT, Chicago IL, in 1993, 1996 respectively. He is currently an assistant professor at the University of Jordan. His current research areas including; cellular communication systems, quantum computing, array processing and GSM-GPS integrated systems.



D. I. Abu-Al-Nadi received BS from Yarmouk University- Jordan, MS from Oklahoma State University-USA, and PhD from the University of Bremen-Germany all in Electrical Engineering, in 1987, 1991, and 1999, respectively. He is currently an associate professor at the University of Jordan. His current research areas are array processing and GSM-GPS integrated systems.

ZONA: A FORWARD COLLISION AVOIDANCE SYSTEM

Lucy T. Gunawan
Augustinus H.J. Oomes
Delft University of Technology
2628 CD, Delft, The Netherlands
Email: {l.t.gunawan|a.h.j.oomes}@tudelft.nl

ABSTRACT

Human perception of distance is limited; moreover it gets worse when another factor such as speed is added. This is considered one cause of many rear-end accidents on the road. Apparently there is insufficient support for the driver in this regard.

Zona tries to solve this problem. It is designed to display safe inter-vehicle distances and to be projected on heads-up car displays, where it shows the zone of safe inter-vehicle distances. In this paper, the process and evaluation of the design are described. The paper concludes with a preliminary evaluation of the design and reflections on future research.

BACKGROUND

A Rear-end collision is a traffic accident caused by an impact to the rear of a vehicle. The common conditions often associated with rear-end collisions are a sudden braking/ deceleration by the leading vehicle followed by the rear-ending vehicle, which does not have enough time to brake because it is following too closely.

Several driver performance factors contribute to the cause of such collisions, including driver inattention and long perception/reaction times from limitations in the human visual system. Human perception is limited in judging distance correctly (Bishop, 2005). Perception gets worse when distance is combined with other factors such as speed in judging safe inter-vehicle distance.

Rear-end collisions are the number one cause of motor vehicle accidents in the USA (Traffic Safety Facts 2004). In 2004, there were approximately 1.8 Million motor vehicle rear-end crashes (30.5% of all accidents) resulting in approximately 2,083 deaths (5.4% of all traffic related deaths) and approximately 555,000 injuries. Each year in the Netherlands, tailgating and driving too fast are the cause of a great deal of irritation on the roads. Moreover, these violations often play a role in accidents and congestion (Belonitor project website, 2006). These show the potential benefit of collision avoidance systems.

Several projects have been developed in the domain of collision avoidance systems to prevent or decrease the severity of rear-end crashes. Collision avoidance systems should detect potential crash situations or close inter-vehicle distances, then provide support to the driver in the form of issuing a warning, automatic braking, or automatically steering away, where each or a combination of these actions has its own requirements and designs. The basic equipment for this system usually includes a camera, radar or lidar, a processing unit, and some form of output. In this project, our focus is collision avoidance systems that provide wa8.5rnings, as opposed to systems that take over some part of vehicle control.

The success of a collision warning system depends on how well the algorithm and driver interface are tailored to driver capabilities and preferences (Lee et. al., 2004). Some studies have examined various algorithms and motion and geometric parameters (Parasuraman et. al., 1997; Seiler et. al., 1998) for collision warning effectiveness. Information on motion and geometry was usually generated using 3D recognition software, based on imagery gathered from the camera and sensors implemented on the car. Algorithms calculate when to issue a warning, and the design of the driver interface influences the reaction time of the driver.

The effectiveness of collision avoidance warning systems depends on several critical factors. First, a warning system must promote a timely and appropriate driver response (Lee et al., 2004). If the driver responds to the situation incorrectly, inappropriate actions will occur such as unnecessary braking or braking sharply. These inappropriate actions can harm the driver's safety. Second, the system must be highly reliable to achieve driver acceptance. Annoyance associated with nuisance warnings must be minimized, and drivers have to trust the system in order to accept it (Kiefer et. al., 1999). Third, warnings given should result in a minimum load on driver attention (Seiler et. al., 1998). Frequent warnings may desensitize the driver and cause ignorance to future warnings. Rare warnings can distract the driver during critical situations.

There are some interface characteristics that could affect driver performance and acceptance, such as:

Graded and single-stage warning

A graded warning presents a signal proportional to the degree of threat, such as a louder auditory signal as the driver approaches a lead vehicle. A single-stage warning provides a signal only when the degree of threat exceeds a threshold. A graded warning might enhance a driver's response by priming the driver's response and enhancing their understanding and trust in the system. Graded warnings provide a greater safety margin and no increased annoyance associated with the greater number of alerts produced (Lee et. al., 2004).

Sensory modality

Torque-based kinesthetic stimuli reduce reaction times more than auditory cues (Gielen and Schnidt., 1983). Haptic stimuli were preferred to the auditory stimuli on several dimensions including trust, overall benefit to driving, and annoyance (Lee et. al., 2004). Haptic stimuli also improve driver reaction to collision situations (Janssen and Nilsson, 1993; Tijerina et. al., 2000).

Substantial studies also show that reciprocal effects between different sensory modalities speeds reaction time (Todd, 1913; Forster et. al., 2002). Vibrotactile stimuli enhanced reaction time to visual stimuli (Diederich, 1995 cited by Lee et. al., 2004). Observers responded faster to simultaneous visual and tactile stimuli than to single visual or tactile stimuli. Multimodal stimuli are also faster than dual visual or dual tactile stimuli (Forster et. al., 2002). Therefore, there is a possibility of enhancing haptic and visual stimuli for collision avoidance systems.

Current systems and similar research

Mobileye-AWS (Mobileye AWS system website, 2006) is a mobile system to alert and increase the awareness of drivers for forward-collision warnings and maintaining safe distances. It is based on a single camera located on the front windscreen which tracks vehicles on the road ahead providing distance, relative speed, and lane assignment of the vehicles ahead. The interface is displayed on a small display unit (Figure 1) and speakers inside the car to provide audio and visual warnings. We thought that the color-coded perspective lanes and the car icon do not correspond naturally with the reality. Firstly, the interface is not consistent. The car icon has the same size for safe, caution or dangerous zone while the lanes beside it use a perspective view. Secondly, the color-coding does not correspond naturally with the zone represented; The green zone is presented opposing the driver; it is seen from the leading car's perspective, so the driver's zone is always red.

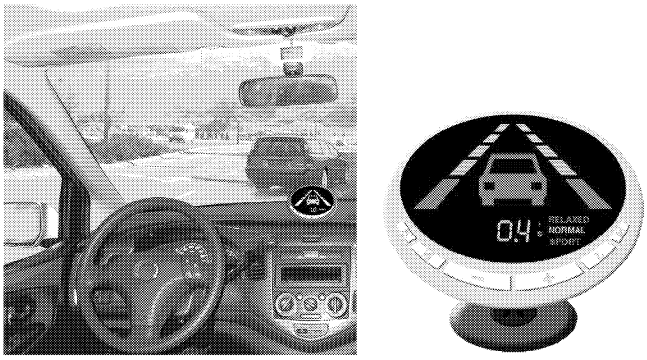


Figure 1: A display unit inside the car to provide visual warning

Another related research project in the Netherlands is Belonitor (Belonitor project website, 2006). The Belonitor trial was developed within the innovation programme 'Roads to the Future' of the Dutch Ministry. The idea of Belonitor is to give rewards instead of punishment in response to driving behaviour. The drivers were rewarded with points by showing safe driving behaviour such as keeping a sufficient distance from other vehicles and obedience to the speed limit. The points collected can be exchanged for gifts. The Belonitor technology consists of a display, a calculating unit, a digital speed map of the Netherlands, GPS, GPRS, and a radar sensor. Although this project is not focused on a collision avoidance system, there is an interface displayed for distance maintenance, as shown in Figure 2.

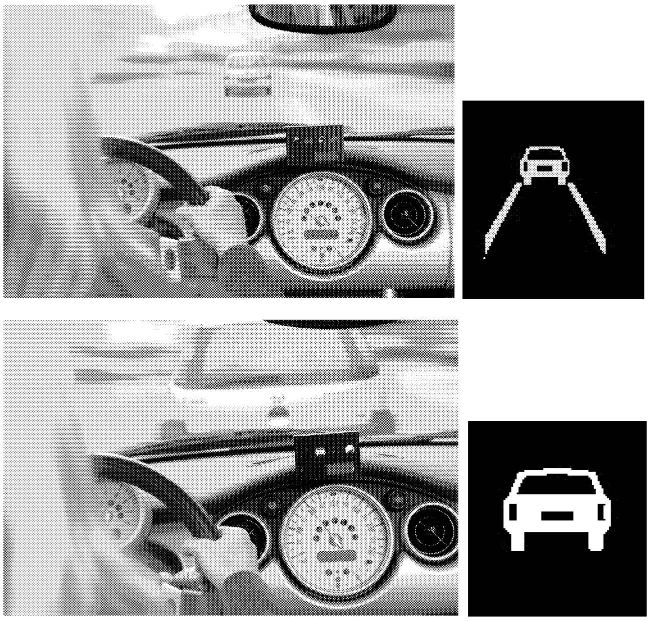


Figure 2: Interface of Belonitor, the first icon (car with lane in green) represents the distance to the leading car is more or equal than 1.3 seconds, while the yellow car icon represents distance is less than 1.3 seconds.

Sufficient distance in this project is considered to be the distance covered in 1.3 seconds. The interface in this project is very simple, based on three icons with different color. It does not give any distance information and the feedback is discreet.

GOAL

For the Zona project, our goal was to provide visual stimuli (complimented with haptic stimuli) to enhance a driver's eye on the road while providing an alert to the possibility of a rear-end collision, especially informing the driver of what a safe distance is from a vehicle ahead.

LIMITATION

We limited ourselves to collision avoidance systems that provide information of the safe distance ahead, and by assuming that the parameters and algorithms were given. We address only the interface design aspects.

REQUIREMENTS

The requirements were gathered from (Kiefer et. al., 1999; Lee et al., 2004; and Seiler et. al., 1998) with additional application of usability guidelines, formulated as follows:

- Zona should display the information clearly for effortless interpretation
- Zona should not be annoying to the user
- Zona should display graded information
- Zona should result in a minimum load on driver attention
- Zona should promote a timely and appropriate driver response
- Zona should be reliable

Our target users are drivers who have difficulties in judging inter-vehicle distances, usually novice to intermediate skill drivers.

To focus the domain for the experiment we want to conduct, we decided to only consider driving activity on the highway.

DESIGN

We brainstormed ideas for the design based on our requirements and came up with 5 possible designs. The first one is a trail bar from the car ahead or behind, as shown in Figure 3a. The trail shows three distance zones: green is a safe distance, yellow is beware and red is the danger zone. The second design (Figure 3b) is almost the same idea as the first design, but less area of the trail is displayed by using an arrow. The reason behind this idea is to make it less annoying to the user. The third design (Figure 3c.) doesn't display the area but instead displays only the border between the zones. The fourth design is the top view of the car positions on the road (car ahead, our car, and car behind) as shown in Figure 3d, displayed vertically. The fifth design is almost the same idea as the fourth design, but has the side view of the cars displayed horizontally (Figure 3e)

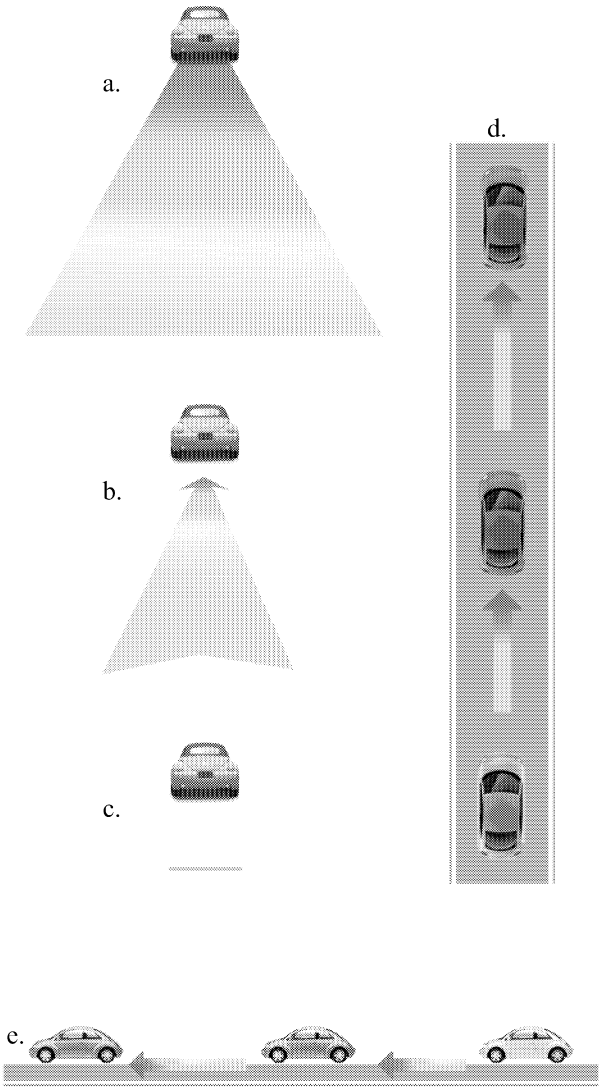


Figure 3: Design ideas

We also want to know what kind of information about distance should be displayed. Do people need this additional information to complement the display that we choose? Distance information is displayed in time (seconds), merely because it is used broadly to measure inter-vehicle distance and is easier for drivers to understand than distance in miles or kilometers. This information can be represented either by using an analog gauge or digital numbers (Figure 4), or is not presented at all. These parameters will be tested to elicit user's preferences.

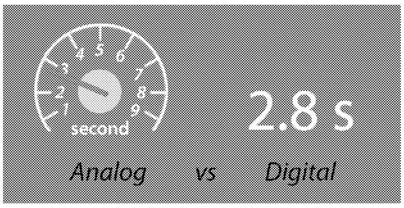


Figure 4: Distance information by using an analog gauge or digital numbers

EVALUATION

The goal of our evaluation was to check the clarity and lack of annoyance of our preliminary designs (all combinations of the designs and choice of available information). The test was conducted mostly in offices at the MMI group, TU Delft. Subjects were chosen who could drive a car and have a driver's license. The interfaces were printed in color on paper, showing the road ahead with a car in front and showing the road behind with virtual rear mirror on the right top of the interface as shown in Figure 5.

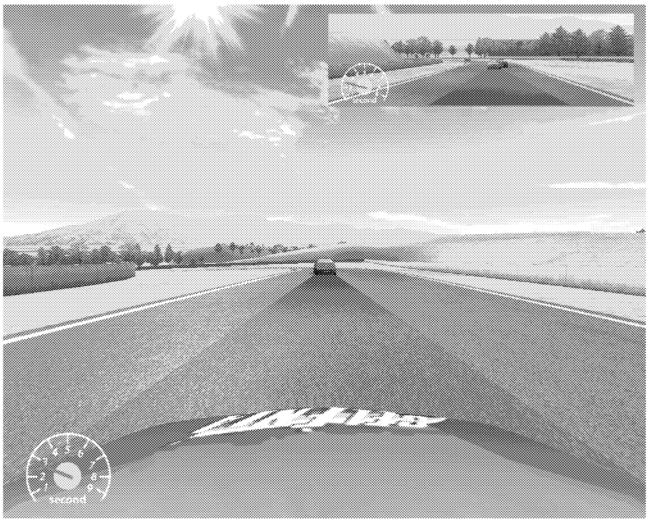


Figure 5: Interface prototype

Figure 6. Experimental setting

In total, 15 versions of the interface were shown to each of ten participants, who were then asked to rate the clarity ("clearness") and annoyance of all the interfaces. The responses were on a continuous scale between "not clear" to "clear" and "not annoying to annoying". The results indicated that users in clarity preferred the projected design on the road (Figure 3a, 3b, and 3c) instead of iconic information of the car positions (Figure 3d and 3e). This might be because iconic presentations require additional step of processing in the brain for interpretation. The trail bar from the vehicle in front (Figure 3a) is most preferred. Among three different ways of presenting the distance (in seconds of driving time) to the vehicle ahead, digital numbers were preferred. The level of annoyance on the design and information is the same. The only indication that we got was from the mean: the trail bar design (first design, Figure 3a) is most annoying and the border zones (third design, Figure 3c) are least annoying. Some comments from participants were gathered after the test, including opinions and suggestion for the design. Iconic presentations were considered bad, because of the time it takes to interpret them. The fourth design (horizontal top view, Figure 3d) was considered dangerous for pedestrians who want to cross the street, because it obstructs the view of the driver to objects beside the road. Some people mentioned that maybe information about distance displayed as numerals is not really necessary because the colored zone has shown it clearly.

This experimental setting was described to be relaxing and gave an appropriate amount of time to grasp the presented information, as seen in Figure 6. One of the limitations of this experiment is the static presentation of the picture, and we couldn't present a real or near real-life situation, only a snapshot of driving experience. These results will be implemented into a video prototype and taken over to the next phase to be implemented dynamically.

Video Prototype

To clarify our design, we've made a video prototype to demonstrate how it will be implemented. We chose the bar design (the first design because of its clarity), and included the digital distance meter (in seconds). We revised the bar design by eliminating the green color and blending it with the road color to make it less annoying for the drivers.

The video shows the animation of the Zona design when a car gets closer/farther to the car ahead.

CONCLUSION

The purpose of the Zona project was to design a system for giving drivers information about safe inter-vehicle distances. Comparing the outcomes of the preliminary evaluation against the requirements, we conclude that:

- Zona provides a clear and effortless interpretation of safe inter-vehicle distances
- All the tested designs were equally annoying, so further work is needed to reduce annoyance
- Zona with having three zones of safe distance displayed graded information.

The last three requirements (minimum load on driver attention, promote timely and appropriate driver response, and reliability) need to be tested further.

Future work should evaluate the design demonstrated as the video prototype. Furthermore, implementation in a real or near-real driving situation needs to be considered. As this project considered only the visual stimuli, enhancement to include haptic stimuli need to be researched further as well.

The system should later be enhanced to take into account the entire traffic situation, not only distance from car ahead.

ACKNOWLEDGMENT

The author would like to acknowledge the support and facilities given by the group of Man-Machine Interaction, Delft University Technology. I would also like to thank Steve Haryono, Tom Brinck, Jan Misker, John Flach, Matthew Smith, Trent Victor, David Woods, Michel Valstar, Kelvin Geerlings, Jason Siesser and all experiment participants for their help and comments.

REFERENCES

Bishop, R. Intelligent Vehicle Technology and Trends. Artech House, MA. 2005

Diederich, A. Intersensory facilitation of reaction-time: Evaluation of counter and diffusion coactivation models. Journal of Mathematical Psychology, 39 (2). 197-215. 1995.

Forster, B., Cavina-Pratesi, C., Aglioti, S.M., Berlucchi, G. Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. Experimental Brain Research 143 (4). 480 – 487. 2002

Gielen, S.C.A.M. and Schnidt, R.A. On the nature of intersensory facilitation of reaction time. Perception and Psychophysics, 34. 161-168. 1983.

Janssen, W. and Nilsson, L. Behavioural effects of driver support. in Parkes, A.M. and Franzen, S. eds. Driving Future Vehicles, Taylor & Francis, Washington, D.C., 1993, 147-155.

Kiefer, R., LeBlanc, D., Palmer, M., Salinger, J., Deering, R. and Shulman, M. Development and validation of functional definitions and evaluation procedures for collision warning/avoidance systems, Crash Avoidance Metrics Partnership, Washington DC, 1999.

Lee, J.D., Hoffman, J.D., and Hayes, E. Collision Warning Design to Mitigate Driver Distraction. Proceedings of the SIGCHI conference on Human factors in computing systems 2004. 65 – 72.

Parasuraman, R. Hancock, P.A. and Olofinboba, O. Alarm effectiveness in driver-centered collision warning system. Ergonomics, 40 (3). 390-399. 1997.

Seiler, P., Song, B., and Hendrick, J.K. Development of a Collision Avoidance System. Society of Automotive Engineers, Inc. 1998.

Tijerina, L., Johnston, S., Parmer, E., Pham, H.A., Winterbottom, M.D. and Barickman, F.S. Preliminary Studies in Haptic Displays for Rear-end Collision Avoidance System and Adaptive Cruise Control Applications, National Highway Transportation Safety Administration, Washington, DC, 2000.

Traffic Safety Facts 2004. A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System. National Highway Traffic Safety Administration, Washington, DC 20590

Todd, J.W. Reaction time to multiple stimuli. Archives of Psychology, 3. 1-65. 1913.

Mobileye AWS system. (n.d.). Retrieved January 23, 2006, from http://www.mobileye.com/aws.shtml

Safe speed and save following, PReVENT European union project. (n.d.). Retrieved January 23, 2006, from http://www.prevent-ip.org/en/prevent_subprojects/safe_speed_and_safe_following

Belonitor project. (n.d.). Retrieved January 23, 2006, from http://www.wegennaardetoekomst.nl

WORKSHOP ON DIGITAL TELEVISION & DIGITAL SPECIAL INTEREST CHANNELS

WEB SERVICES AND TOOLS FOR THEIR COMPOSITION CONSIDERING ASPECTS OF DIGITAL TV WORKFLOW

Hristina Daskalova, Tatiana Atanasova
Institute of Information Technologies -BAS, Acad. G. Bonchev 2
1113 Sofia, Bulgaria
E-mail: daskalovahg@abv.bg t.atanasova@iit.bas.bg

KEYWORDS: *Web Services, Service Composition, Semantic Services, Software Tools*

ABSTRACT

In this paper some research directions and tools in the area of web service composition are discussed. The service composition is defined as a process that combines web services into new service aiming at constructing a solution to the given business problem.

INTRODUCTION

The term “service” became very popular last years. It is used in different context: for example, Software-as-a-Service (SaaS) gives to user an access to remote resources and develops an idea for accessing applications from service providers. Customer-facing services are services offered from various web-sites and a spectrum of such services varies widely – they cover e-commerce, e-mail, and providing real-time data. IT-services are services that information departments provide to the enterprises that are governed by IT Service Management (ITSM) and Business Service Management (BSM). Loosely coupled services are used to construct systems, and ensure data exchange and software components functionality. Such services in principle are web services (in spite of fact that loosely coupled services can use other technologies) and they are considered as a base for services-oriented architectures. Using of Service Oriented Architecture (SOA) provide: Independent and loosly coupled services that can communicate; Services with high granularity which are acceptable through Service Bus.

SOA promises more flexibility with lower tehnological expenses and better suits business needs by

- Unification based on simplification of the integration process;
- Reusing of components and their combination;
- Process automatization and support of changing business requirements;
- Standartization;
- Modernization, merger and compatibility of IT systems after merging;
- Migration from existing systems.

In the paper the web services and possibilities for their compositions are discussed in this sense.

WEB SERVICES

Web services can be accepted as a concept for universal data integration between different web and software applications independently from their vendors, methods, operation systems and principles of their realization. This is a module of application software that can be published, discovered and invoked in web by XML-based standards.

Communication between distributed applications is not a new problem, but the development of open standards for web services provides an interaction that is unrestricted by company’s private rights and gives an opportunity for achieving more interactive and intelligence behaviour. Web services are functioning in the web, they are using the web as transport and an environment, but they have their own protocols and standards. Tools for web service construction assist automatic generation of everything that is necessary for their functioning (at present, on the base of SOAP, WSDL and UDDI) with help of wizards. For example, CapeStudio generates Java from WSDL by using WSDL Assistant. Sun ONE Studio 4 supports web service generation and makes web services attainable for every experienced Java programmer. IBM Web Service Toolkit (WSTK) is a set of tools and tutorials for using Java with SOAP, WSDL and UDDI. WSTK works with WebSphere Application Server. With Microsoft Visual Studio .NET, Enterprise Edition the constructing of web services is more simple then in Java.

Technologies and concepts for web services are still in the continuing process of standardization and discussion of services nature itself. A lot of research projects are conducting aiming at design of program environment allowing web service composition for universal integration of various applications. For realizing this ambitious goal it is necessary to identify web services that can be combined for satisfying user’s functional needs and quality criteria. Clear definition of service composition does not exist yet. Beside the term “service composition” often a “behaviour of web services” or “properties of the interactions”, or “high level description of interactions” are considered. It is reasonable to accept that composition of services is a process by which services are connected and which gives new service using appropriate operators to find a solution for some business task. Service composition has to satisfy functional and non-functional requirements and has to provide a correct result. A service or composed service can be an element of the service composition.

Every service is characterized by its input and output parameters. The input parameters are necessary for introducing the functionality of the service, and output parameters give a result from service execution. For service invocation a set of conditions is needed. Service execution can change the world state and it determines effect that the service produces. Service composition is defined by control flow of the process in which actions in the composition are ordered (sequential, parallel, alternative, etc.) and is tightly connected to workflow. But traditional workflow is only partially compatible with service-oriented engineering of resources. Workflow composition is based on the idea that a workflow as sequence of tasks makes a plan implicitly defined by the pre-conditions and post-conditions of all tasks in the application domain. The initial and final states of such a plan are determined by the business goal requirements. The tasks composition can transform an initial world state into a final required world state.

There are some industrial existing methods for service composition as BPEL4WS, WS-CDL and some abstract methods as Petri Nets, Pi-calculus and Web Component, the methods are accompanied with a lot of discussions. Industrial methods do not provide correct verification of created service composition but they are connected with real applications. Abstract methods have precise theoretical mathematical foundation and correct verification; they can be used for development of other languages and methods.

Methods for service composition can be divided into manual, semi-automatic an automatic, everyone with its advantages and shortness. But general steps for every method are service discovery; service matchmaking; data- and control-flow linkages.

MANUAL SERVICE COMPOSITION

Until now the composition of services usually is made manually. Starting form the process description the composer tries to combine services; first of all, needed services are found, as the text description of the service given by service provider is used for defining its capabilities and non-functional properties. After that the services are connected in the desired order.

During searching for the correspondence between input and output parameters in participated services it can be detected that they have different data structures and the mechanism for their transformation has to be found manually. So step by step the chain of services that can reach the desired goal of the process is constructed.

Modeling of these aspects with languages like BPEL4WS does not allow foreseeing all possible situations, which can arise during execution of the process. Reasoning about possible failures is not enabled too. The composer itself allows mistakes or inadequatenesses during service composition. These factors do not guarantee that manually made service composition is correct. Beside that the composition process is not flexible enough at the time of execution. One of services in the composition can be not valid at the stage when the prescriptions for the composition are made and this can lead to breakdown of the process. That is why investigations are conducted to find adaptive semi- or fully automatic methods for modeling of the composition of services

SEMI-AUTOMATIC COMPOSITION OF SERVICES

The main idea in the semi-automatic service composition is to provide users during composition with filtering, selection and decision support. The composer has to analyze the current composition made by the user and to suggest next possible steps.

In this process the user can follow one of these strategies:

- Select components without description of the desired result, the desired result is specified starting from some abstract model;
- Select components on the base of desired result, the user describes the desired result and wants to simulate a situation that leads to this result;
- Select component on the base of their initial description, and the user wants to receive a simulation model that describes possible result.

Within semi-automatic composition a problem for discovery and matchmaking is automatically resolved, but difficult tasks for determination of control- and data-flow as interconnections between them are leaved to the user. Service discovery can be implemented using repositories like UDDI, which give means for registering and publishing services and allow their detection according to the request. The amount of potential services can be large so a special mechanism for finding appropriate service is needed. One possible way in this direction is to use semantic descriptions of web services - Semantic Web Services (SWS) like OWL-based Web service ontology (OWLS), Semantic Web Service Language (SWSL), Web Service Semantics - WSDL-S and Web Service Modeling Ontology (WSMO). Semantic description in form of shared ontologies diminishes the semantic heterogeneities of service interfaces.

Some tools for semi-automatic service composition are considered further.

In (Atanasova et al 2006) the semantic service composer is considered for composition of semantic services. SWS composition deals with combination of different already created semantic services to obtain new service. The SWS Composer provides ontology-oriented service-based workflow of WSMO-described services. This approach is characterized by decision supported selection of appropriate services and querieng Organisational Memory for service description.

The similarity-based Organisational Memory helps searching for compatible services by filtering and suggesting services that use lexical and ontological similar context. The Organisational Memory provides support for making decision on the base of similarity inference when the composer tries to combine services from different application domains that usually use different set of ontologies. But composition of these services is still possible because they are based on some general set of ontologies that are domain independent and can ensure semantic consistency of the composition. The services that can be combined are supposed to use similar semantic annotations and their annotations match with some measure of similarity. One of the main advantages of using the similarity-based memory is that the result of similarity-based searching can be used in terms of the application domain terminology. The retrieved services are similar in a declarative way. But for the satisfaction of desired properties described by logical expressions with given set of concepts and relations the logical discovery should be used.

Logical discovery is necessary for searching of ontologically interconnected services; these are services with common ontologies set. Discovery of matching services is realized by quering the Service Agent Middleware for searching semanticaly suitable service. The query is constructed as WSMO goal from the description of selected service in the composition.

Data integration in this tool is realized by

- referencing shared variables;
- importing service ontologies;
- matching ontological attributes from postcondition to precondition of participated services;
- constructing additional logical relations and logical constraints for the composed semantic service.

Visualization of service composition is organized by graphical tools for importing/exporting services with associated ontologies and graphical tools for editing and modifying data and control flows.

SWS Composer (figure 1) works in the development environment to compose SWS components in accordance with WSMO specification. It is naturally that the composed service may have additional constraints defining interaction of services. The Composer gives means for defining such constraints that are logically expressed in precondition of the whole composed service

The dataflow between participating services in the SWS Composer is designed by the user and is based on the idea of semantic bridge. The concept of semantic bridges has been used in several approaches for SWS integration, for example - translator services in Mindswap composer [www.mindswap.org/2005/composer] and concept of mediator services in WSMO. Instead of using data mediators (that are underspecified in WSMO yet) this Composer performs mapping between ontological terms using pre-defined axioms in form of WSML logical expression to be included in the capability description of the composed service.

The composed service pays attention to transferring needed data between component services. The choreography of the composed service is formed from rules of participating services. It is possible to add new rules, modes and ontologies into choreography of the new composed service. The choreography describes the interoperation of the service with its client. When the client of the service is another service then we have an orchestration. This gives a solution to construct an orchestration of composed service from the choreography descriptions of the component services.

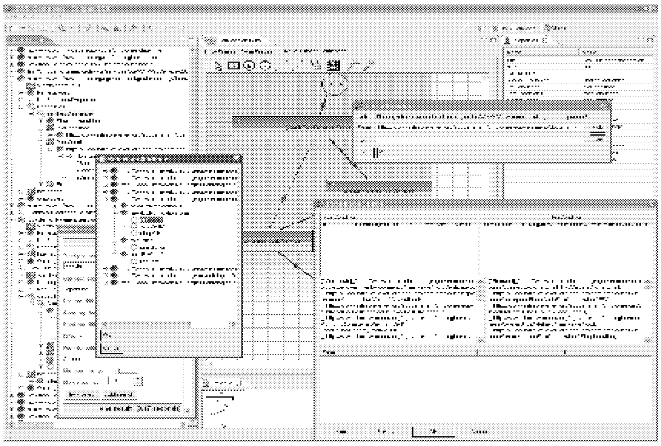


Figure 1. SWS Composer

In (Sirin et al 2004) Web Service Composer is presented, that allows executive composition of Web services, which are semantically specified with OWL-S (Martin et al 2003). Composed Web services can be remembered and stored as OWL-S “process models”. These models are parts of OWL-S ontologies and usually are used for choreography definition of the composed service. In OWL-S the process models can be used as control structures in the area of Workflow Management.

During creation of the composed service the user takes an advantage of the feedback connection and selects the constructed desired service as a result of composition by using a list of all available services. Further the user interface gives an additional list connected to every OWL input type of the service as a result of completed process. This list does not contain all potential services but only those services that are generated on the output from particular input type to which it is connected. One output from service *A* associates to the input of service *B* if their types are absolutely the same or the output of *A* includes input of *B* (that is the input of *B* is defined as output of *A*). If the selected service belongs to the list of appropriate services then the inputs of these services again have to construct appropriate output through selection of services. The process is repeated until the user decides which inputs are not connected and not included to the service as input variables. As a result the plan built with this tool is not always optimal. When one service invocation produces two outputs that satisfy different inputs of the constructed service, then this invoked service is repeated twice in the composition.

In (Kim et al 2004) Composition Analysis Tool (CAT) is described that illustrates the author’s method for interactive workflow composition. Even the proposed tool is not connected directly to service composition; the computational workflow can be used for this purpose. Activities in the workflow are represented as services which realize data transformation. Authors have developed their own format of knowledge base used for semantic description of the components in the workflow and their input and output parameters. “Component ontologies” describe hierarchy of components from abstract level to the operations implementations. As abstract components a set of type representative properties is considered. “Domain ontologies” semantically describe data that attend to inputs and outputs of the components described in “Component ontologies”.

In CAT the user can add components to the composition at any time. Instead of list of services that can participate in the composition CAT gives a recommendation for next action. These recommendations may lead to wrong decisions that are notified with visual warnings. Authors introduce a set of properties to receive suitable composition format. These properties have to provide successful implementation of all necessary operations. The proposed tool does not ensure the desired functionality or in part plans creation.

(Myers et al 2002) introduces Plan-Authoring System based on Sketches, Advice, and Templates (PASSAT) as interactive mean for plan construction. PASSAT as well as CAT is not connected directly to the service composition. Representation of plans in PASSAT is based on HTN (hierarchical task networks) model which can be adapted to service composition. With HTN planning (Tate 1977) the planning system formulates a plan by decomposing tasks into smaller subtasks until primitive tasks are reached. Variables describe states before and after the task execution. Constraints on these variables can be used for expression of preliminary conditions. HTN-based method imposes plan and the user is included in the planned strategy. The user can start planning by adding tasks to the plan. The tasks that are suitable for the plan can be incorporated into the HTN templates. The template consists of a set of subtasks which substitute existing task as prerequisites and submits individual task as completed template. The user can reject incompatible requirements during template creation. This is especially useful when it is impossible to provide a full range of template collection. The process is repeated until the plan can not be more extended. The system helps user to select appropriate templates for the given stage of the composition according to previous experience, the statistics results are provided about how frequently the template is incorporated into the plan.

IRS-III (Hakimpour et al 2005) also includes a tool that supports a user-guided interactive composition approach by recommending component Web services according to the composition context. Their approach uses the Web Services Modeling Ontology (WSMO) as the language to describe the functionality of the Web services. The available services are filtered at each composition step.

In (Schaffner and Meyer 2006) an environment for semi-automatic modeling of Web service during their composition is presented. At every stage of the composition process the environment suggests some amount of appropriate Web services and ensures a possibility of their invocation. During construction of composed service from large number of single services there are difficulties for modeling of appropriate operations originated from their discovery for particular goals. The idea of using mixed initiatives aims at easier realization of service composition. The researches propose systems that ensure automatic planning for every particular case of composition in real time which is opposite to the idea of service composition that covers as much as possible cases. Plans are created for fully automated execution using knowledge base and semantic description of services. This allows decreasing the complexity and increasing flexibility but possibility of eventual errors in composite service is increasing too. Difficulties originate from incomplete formal representation of the knowledge base. The problem for formal specification of domain knowledge with sufficient exactness is an enormous challengeable. Especially for complex domains it is reasonably to assume that full ontologies will not be available in the near future. Imperfect presentation of the domain knowledge does not allow construction of the composition plan. Wrong description provokes wrong planning of the desired composition. Human expert experience in the specific area should be included. Such experience can compensate the lack or wrongly used ontologies. Still more, methods for full automatic composition of services do not recognize exclusion of human operator from the composition process as an organizational or legal hindrance. However, in real business cases particular people are responsible for correct process passing. And this is a practical advantage of semi-automatic methods against automatic. This is why fully automatic planning technique is not been accepting from industry side yet and the transfer of academic investigation to practical applications is going on slowly. But this technology can be used to make it easier construction of business processes manually. Unification of technologies from automatic planning with tools for semi-automatic modeling for business process composition gives some advantages. Problems with manual service composition can be relaxed or removed through adding of “mixed initiative features”. In (Schaffner and Meyer 2006) some investigations expand the last results on semi-automatic composition and show that modeling tools are completed in sense of developed functionality. Existing methods use different ontological formats for representing specific domain knowledge. This remains a problem and their integration is getting more complicated. With discussed methods the user is not provided with an access to the expert area during service composition because this requires special training for the user. This also restricts industrial application of semi-automatic composition methods.

The scenario that realized “Leave Request” uses software Duet (http://www.duet.com) from SAP and Microsoft.

Duet expands achievements of SAP in business process description by involving of large number of Microsoft's Office users. In (Schaffner et al 2006) featurees of three mixed initiative features for semi-automatic composition of services are realized. Tools for service orchestration are proved both for training and for industrial service repositories. For the present the correctness of semantic in contrast to syntactic description remains unsolved problem.

Run-time service discovery and late-binding are those advantages of web service technology that have the biggest value in real-time solutions. Run time discovery allows automatic recovery of services by coordination of semantic request with service description. This is getting in accordance with the query for the service (as a result of offering an abstract service), or for different services (as a result of offering concrete service) found as a result of similar parts in described functionality. The selection between these services can be dictated from functional properties (as a hotel reservation accompanied with tour reservation) or non-functional properties - Quality of Service (QoS) - as attribute quality. For example, selection can be made between quick service and cheap service or as a compromise between these two variants. Given possibilities for selection of different services to be included in the composed service is necessary to define a set that connects abstract and concrete services not only according functional requirements but those of QoS, defined by Service Level Agreement (SLA) with participation of the user. Finding solution for this problem needs applying some optimization technique. Beside that during execution of the composed service it may be necessary to change service connection that was defined preliminary. Thus in some cases it is more appropriate do not connect services before execution, this depends on the context of the composition and may be not clear before the execution.

The framework that supports a pre-execution binding (able to satisfy some global composition constraints), run-time binding of single service invocations and run-time re-binding of the whole workflow is developed within SeCSE (Service-Centric System Engineering) European project (http://secse.eng.it).

Tools for semi-automatic service composition help user during composition by suggesting appropriate services and checking for errors at every step of creating of composite service. The user makes a decision and fits together control and data-flows. This approach does not provide general solution for the dynamic composition. The changes that may come during execution have to be taken into account by user himself and possibility to interrupt the process should be leaved to the user.

In semi-automatic modeling of the composition tools on every step of process construction a set of suitable Web services is invoked. It is reasonably to harmonize proposed research tools with business applications from leading industrial providers.

AUTOMATIC COMPOSITION OF SERVICES

Automatic service composition suggests to automatically implementing discovery, matchmaking, data- and control flow and interconnections between them. The human factor is ignored here and the composition is accomplished according to the request by automatic adapting to the general current state. Arising errors are removed through re-planning realized again as composition of services prescribed before. One of techniques suggested for solving such complicated task is Artificial Intelligence (AI) planning. (Russel and Norvig, 1995) interpreted planning as: "a kind of problem solving where an agent uses its beliefs about available actions and their consequences, in order to identify a solution over an abstract set of possible plans". The planning contains: description of possible actions that can be described by some formal languages; descriptions of initial general state; description of the desired goals.

Formalization of the area of physical operations recognizes abstract operations available or connected with some agent. Operations can be precisely defined by preliminary conditions and effects of their realization. During planning participated agents have to report initial state that gives a plan leading to the accomplishing of specific goal. Difficulties in realizing of this task are in the specification of all relevant knowledge during task planning. The planner has to identify the plan that will be implemented strating from initial conditions and leading to satisfying the goal. The goal usually specifies properties of the final state or gives description of operations in already executed task. The plan consists of sequence of operations. It is possible to define unfavorable situations which can arise during task execution so the plan branch prepared preliminary can be selected.

In (Sirin et al 2003) an attemp to automatically compose services using Hierarchical Task Network (HTN) (Erol et al 1994) planning is made. In SHOP2 the method for transforming of simple (elementary) tasks as composed in OWL-S in hierarchical net of tasks.

The user specifies the initial state and the composed services which have to be decomposed into elementary services by using given optimization rule. The advantage of this method is in convenient execution when large number of application domain is present. However it is not easy and not always possible to prepare complete description of the task in the dynamical environment. Some intentions for dealing with the difficulties in such directions can be found in (Peer 2005). The automatic service composition can be achieved under condition that there is a complete and correct semantic annotation to the service description.

SERVICES IN DIGITAL TV

New information technologies are coming into all areas of human life but new requirements to the telecommunication services are arising too. The i2010 initiative cares about the possibility for European citizens to "watch and listen to the audiovisual content anywhere, anytime and from any technical platforms – TV, computer, PDA, mobile phones, etc. Really most of the ambient information is in digital form. This information is heterogeneous, multimedia and more and more multilanguage. Development of methods and tools for processing and structuring of this information has a key significance. Progress in processing of certain types of information facilitates development of methods and tools for utilizing of digital TV and interactive feedback services.

Software services in digital TV can be considered by analogue with web services. Methods of their combination to construct new value added services for realizing business models in digital TV can be used.

BUSINESS MODELS RELATED TO DIGITAL TV

Digital TV combines two specific types of communication channels: channels for TV broadcasting and interactive feedback channels. The technical platform can ensure new business possibilities providing an infrastructure for creating new value-added services and facilitating e-business.

Interactive feedback channel allows the user to be an active participant in:

- Receiving additional information
- Playing games
- Video on Demand

The new functionalities in digital TV require different degree of feedback. Different user interactivity defines different services and payment schema.

Providers of TV content can construct packages combining their own services with services provided from different partners. With existing of independent services and active role of the user new mix of value added services can arize. The combination of TV as basic services can be made with adding news or shopping services. They are free or paid services with some payment schemes. New services in an interactive information and communication society can only be successfuly created in a collaborating environment by composing services from different content providers.

CONCLUSION

The information technologies is an area of investigation with so quick and constant changes that before some achievements become proved and utilized they are replaced by new results and their use in business processes often does not justify their expectations. Business-oriented researches show that new instruments for industrial process control are requested that are able to integrate strategic processes for decision making with operative planning and project realization to made decisions together with providing feedback connection from their realization. The composition of services has to ensure functional and non-functional user requirements and to guarantee correct execution of the business process. Future investigations may propose methods for the composition and decomposition of services, to revise languages for semantic service specifications and descriptions of the process of composition itself. The investigations of the problems around service composition are challengeable and possible results are expected to be useful for the system development and integration of information resources.

REFERENCES

Kim, J., Spraragen, M., Gil, Y. An Intelligent Assistant for Interactive Workflow Composition. *In: Proc. of the Int. Conf. on Intelligent User Interfaces (IUI-2004)* Madeira, Portugal, (2004), pp. 125–131

Martin D. et al. OWL-S: Semantic Markup for Web Services. Technical report, Nov (2003).

Roman, D. H. Lausen, U. Keller, J. de Bruijn, Ch. Bussler, J. Domingue, D. Fensel, M. Kifer, J. Kopecky, R. Lara, E. Oren, A. Polleres, M. Stollberg. Web Service Modeling Ontology (WSMO), 2005 http://www.wsmo.org

Sirin, E., Parsia, B., Hendler, J. Filtering and selecting semantic Web services with interactive composition techniques. *In IEEE Intelligent Systems*, Vol. 19, Issue 4, 2004, 42-49

Atanasova T., H. Daskalova, N. Bakanova, Design Aspects of Applications Integration in Distributed Information Systems, *Proc. of the International Workshop DCCN 2006 - Distributed Computer and Communication Networks,* Sofia, Bulgaria, October 30 - November 3, 2006, pp.27-36.

Myers K. L. et al. PASSAT: A User-centric Planning Framework. *In Proc. of the 3rd Int. NASA Workshop on Planning and Scheduling for Space*, Houston, TX, USA, AAAI, (2002).

Tate A., Generating Project Networks. *In Proc. of the Fifth Joint Conf. on Artificial Intelligence, Cambridge, MA, USA*, pages 888–893, (1977).

Hakimpour, F., Sell, D., Cabral, L., Domingue, J., Motta, E.: SemanticWeb Service Composition in IRS-III: The Structured Approach. In: *7th IEEE Int. Conf. on E-Commerce Technology (CEC 2005)*, München, Germany, IEEE Computer Society (2005) 484–487

Schaffner, J., Meyer, H.: Mixed Initiative Use Cases For Semi-Automated Service Composition: A Survey. In: *Proc. of the Int. Workshop on Service Oriented Software Engineering (IW-SOSE'06)*, 27–28 May, 2006, Shanghai, China, ACM Press, NY, USA (2006)

Schaffner J., Meyer H., Tosun C., A Semi-automated Orchestration Tool for Service-based Business Processes, *In: Proc. of the 2nd Int. Workshop on Engineering Service-Oriented Applications: Design and Composition,* Chicago, USA (2006)

Russel, S. and Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice-Hall Inc., (1995)

Erol, K., Hendler, J., and Nau, D. S.: Semantics for HTN planning. *Technical Report CS-TR-3239*, (1994)

Sirin E., B. Parsia, D. Wu, J. Hendler, D. Nau: HTN Planning for Web Service Composition Using SHOP2. In *"Web Services: Modeling, Architecture and Infrastructure: Workshop in conjunction with ICEIS,* (2003).

Peer J.: Web Service Composition as AI Planning – a Survey. Second, revised version, *Technical Report,* University of St.Gallen, (2005)

FRAMEWORK APPROACH FOR SEARCH AND META-DATA HANDLING OF AV OBJECTS IN DIGITAL TV CYCLES

Tatiana Atanasova

Institute of Information Technologies -BAS
Acad. G. Bonchev 2, Sofia,
Bulgaria
Email: t.atanasova@iit.bas.bg

H Joachim Nern

Aspasia Knowledge Systems
Postcode: 200710, Duesseldorf
Germany
Email: nern@aspasia-systems.de

Andrei Dziech

University of Wuppertal
Rainer-Gruenter-Str. 2, Wuppertal
Germany
Email: dziech@uni-wuppertal.de

Nikitas M. Sgouros

University of Piraeus
Karaoli Dimitriou 80, 18534,Piraeus
Greece
E-mail: sgouros@unipi.gr

KEYWORDS: *Audio-visual objects, WEB TV, Digital Content, Digital TV Workflow, Meta Data, Semantic Services*

ABSTRACT

The paper objective is to describe an approach for a software framework consisting of an application oriented tool set for enhanced search, retrieval and processing of audio-visual (AV) content objects. Using semantic web service technologies and watermarking techniques the framework is designed to cover several aspects of object and workflow handling. It allows optimized system integration to support workflow environments realized as P2P as well as open network and mobile service approaches.

INTRODUCTION

Existing approaches related to AV object search and retrieval is reduced to the application of conventional search engines, like Google, yahoo, Lycos, etc. (Henten Anders and Reza Tadayoni, 2005). Accordingly the degree of integration and encompassing features and characteristics is low. In the following an integration approach that combines existing tools and models as well as innovative attempts out of the area of resource discovering, watermarking and semantic web service based interoperability is discussed (Nern H Joachim, and T Atanasova, 2005)

The state of the Art in the Semantic Web area is mainly characterized by two streams:

The WSMO initiative aims to meet three goals: to describe the domain of Semantic Web Services, to define a conceptual model for a formal language design and to define an execution environment, which can enable the complete definition and execution of Web Services' interactions (Stollberg M., D. Roman, 2005, Fensel D, 2000)

OWL-S: According to the OWL-S ontology specification, a service can be defined through the use of core components the Service class is made of: the Service Profile, the Service Model, the Service Grounding and the Resource Ontology. Previous releases of OWL-S were referred to as DAML-S and were built upon the DARPA Agent Markup Language and Ontology Inference Language, DAML+OIL (Ruben Lara et al, 2004).

The framework proposed in this paper is focused on the important aspect of integration. The described framework allows the integration of existing digital repositories so that the content can be made available to a semantically enriched digital environment. This encompasses both interactive content services and potentials for targeting advertisements at specific customers or customer target groups. Two main technologies are essential for the realisation of such framework:

- encryption based watermarking for unique and unambiguous identification as well as encrypted object description via embedded metadata;
- semantic annotation and knowledge formalization for optimized object descriptions as well as for ensuring a high degree of interoperability due to the application of semantic web services.

DIGITAL CONTENT DESCRIPTION

Business models on the basis of digital content require adequate representation of that content. Architecture, oriented to services, should allow a processing that content in heterogeneous digital application environments.

Description of AV objects has to provide expressive semantic descriptions of digital content, based on ontologies of information objects. Agents should be embedded into digital infrastructures implemented on advanced knowledge content carrier architecture.

SOFTWARE AND TECHNOLOGY FRAMEWORK

The proposed framework is intended to ensure searching, interpreting, navigating and retrieving audio-visual content objects. It consists of an application oriented software tool set for allowing optimized system integration to support workflow environments realized as P2P as well as open network approaches.

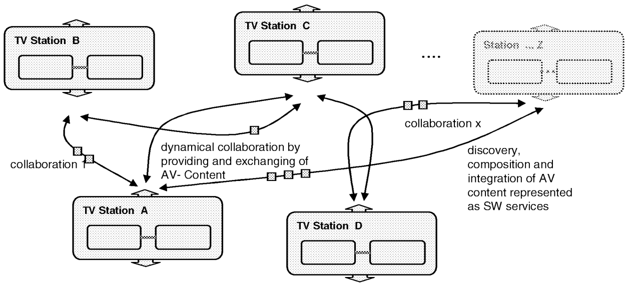

TV Station B
TV Station C
....
Station ... Z
collaboration x
dynamical collaboration by providing and exchanging of AV- Content
collaboration 1
discovery, composition and integration of AV content represented as SW services
TV Station A
TV Station D

Figure 1: Network of TV stations – Digital TV workflow

Fast Generation of New Added Value Services Based on Accurately Retrieved AV Content

Applied to professional TV market relevant structures complex search engines enable and enhance the ability of the service providers to generate new services and to react fast and flexible to market changes. Reflecting customers behaviour (regarding the access and retrieval behaviour of AV content in distributed registries) deliver additional feedback information, which can be applied to complex strategy and marketing planning. Content scanning based on features for service specific AV yields in stimulation with respect to the creation and generation of new products along the B2B-TV added value chain.

Fast and deterministic search – based on contextual semantically enrichment of meta data increases the capabilities for generating new added value services in collaborating workflow cycles (Figure 1).

The discussed framework approach is structured in three combined layers: adequate description of resources; search and retrieval and workflow-based interoperability. This enables AV-object users as well as AV-object providers to access audio visual content in an optimal and flexible way.

The use of such integrated AV-object search and maintaining environment ensures optimal conditions for open horizontal markets. Furthermore it embeds the user as well as the provider in an encompassing manner. The positive effects will be related to optimal selection and choosing capabilities.

Three layers are combined in the framework:

- Search and feature extraction – Firstly, "known" and "owner provided" AV objects are retrieved and transformed for subsequently processing.
- Watermarking for unique identification and description - In the second step the AV objects undergo a watermarking process, whereby the ID information as well as the object description are encrypted in the AV object itself. Furthermore this metadata description are formalized using semantic annotation procedures, which enable an unambiguously knowledge access.
- Effecting workflow enabling interoperability – The semantic annotation is used in the third step for formalization of semantic web and distributed service repositories. The description and identification of the AV objects are embedded in semantic web services. Full semantics are applied to closed P2P networks, while light semantics are used within the open network (WEB).

The software framework is divided into a triple step approach (Figure 2), whereby in each step the degree of object annotation capability (semantic annotation & watermarking), interoperability and search resp retrieval characteristics is increased.

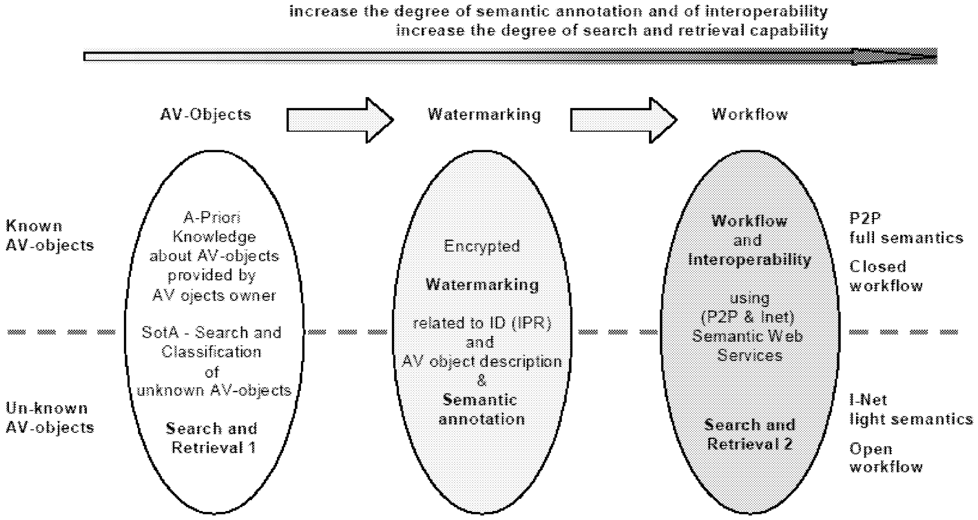

increase the degree of semantic annotation and of interoperability
increase the degree of search and retrieval capability
AV-Objects
Watermarking
Workflow
Known AV-objects
Un-known AV-objects
A-Priori Knowledge about AV-objects provided by AV ojects owner
SotA - Search and Classification of unknown AV-objects
Search and Retrieval 1
Encrypted
Watermarking
related to ID (IPR) and AV object description
&
Semantic annotation
Workflow and Interoperability
using (P2P & Inet) Semantic Web Services
Search and Retrieval 2
P2P full semantics
Closed workflow
I-Net light semantics
Open workflow

Figure 2: Three step approach

SEMANTIC DESCRIPTION OF AV OBJECTS

The description and identification of the AV objects are embedded in semantic web services (SWS), which ensure a high degree of interoperability. The semantic annotation ensures optimized search and deterministic indexing and accordingly stable retrieval characteristics. The semantic annotation is used in the third step for formalization of semantic web services (ontologies formalizing object specific knowledge) and distributed service repositories, which are applied in a P2P as well as open network oriented workflow environments. Full semantics are applied to closed P2P networks, while light semantics are used within the open network (WEB) also accessible by mobile devices.

Enhanced search and workflow enabling interoperability

Due to the machine processable feature of Semantic Web Services (SWS) a great impact to the area of Digital TV e-business can be achieved. Dynamic and scalable collaboration between entities of different nature can be facilitated by accessing and retrieving well formalized metadata of AV-objects in tuned workflow environments. Semantic Web Services enables organisations and heterogeneous entities to build faster and more effective partnerships with respect to the service generation, exchange and distribution process.

Representation of visual and audio objects with certain semantic meaning is opposite to traditional representation, but it is essential element for the content-based multimedia services such as editing and manipulating of image sequences, image and video indexing, and retrieval applications.

The interaction with the semantic content is still far from being satisfactory and is still under investigation and development. Now it is not possible to search visual and audio objects by their semantic meaning. This is due to the lack of methods for the automatic extraction of the semantic content of visual and audio objects (semantic annotation). In the semantic-enabled systems and services knowledge/semantics has to be explicitly represented and, therefore, the visual and audio objects can be effectively retrieved, shared, and exploited.

One of the technical challenges is to create a new generation of self-adjusting service environments building on design and control paradigms consisting of intelligent knowledge management and dynamical networking and platform structures. In this context the integration of closed loop features would ensure comprehensive consistency and traceability effects

Seamless Content Provision & Broadcast Resource Discovery

Based on the semantic annotation features of the proposed technique a seamless content provision to users, customers as well as providers irrespective of the technology is ensured. The complexity of given structures is hiding by this way. It is the basis for build up new and innovative approaches – for example, interactive TV.

Interplay and Mutual Influences of Professional and User Created Media.

The growing and huge amount of AV-objects can only be handled and structured by search and annotation environments like the proposed system. The system delivers both – a fast and precise search and retrieval – and an accurate and precise structuring of data.

Exploitation of the Interoperability features of Semantic Web Technology

Semantic Web technologies have to be explored more for:

- Semantic object extraction from image sequences;
- Designing of converter for user's semantic input into a form that can be conveniently integrated with low-level video processing
- Integrating of high-level semantic information and low-level video features

Fast Reaction and Interchange in Virtual Enterprises and Entities

The vision of "fast reaction and interchange in virtual enterprises and entities" can be pushed to reality using semantically based search environments. Virtual enterprises could be established consisting of several building blocks, modules, products, and services that are interchanged, modified and optimised within one service-based collaboration-space.

CONCLUSION

The proposed framework tryes to overcome the lacks with unsufficient development of the interaction with the semantic content via encompassing and comprehending three main methods: security and metadata related watermarking, enhanced state-of-the-art search considering recognition and classification as well as semantic web service technology providing a high degree of semantic annotation and interoperability.

REFERENCES

Nern H Joachim, A Dziech, T Atanasova, Applying Clustering and Classification Methods to distributed Decision Making in Semantic Web Services Maintaining and Designing Cycles, *EUROMEDIA 2005, Workshop for "Semantic Web Applications"*, April 11-13, 2005, IRIT, Université Paul Sabatier, Toulouse, France

Henten Anders & Reza Tadayoni, Articulation of Traditional and Internet TV, 2nd International Conference of the COST Action A20, University of Navarra (Pamplona, Spain), on 27-28th of June 2003.

Stollberg M., D. Roman, I. Toma, U. Keller, R. Herzog, P. Zugmann, and D. Fensel, Semantic Web Fred - Automated Goal Resolution on the Semantic Web. In Proceedings of the 38th Hawaii International Conference on System Science, January 2005.

Fensel D., "The Semantic Web and its Languages." IEEE Intelligent Systems, Trends and Controversies, pp. 1, November/December, 2000.

Ruben Lara, Dumitru Roman, Axel Polleres, Dieter Fensel: A Conceptual Comparison of WSMO and OWL-S. European Conference on Web Services (ECOWS 2004), Erfurt, Germany, September 27-30, 2004, pages 254-269.

TOWARDS DYNAMIC, USER-DRIVEN CONTENT CREATION AND DELIVERY IN IPTV ENVIRONMENTS

Nikitas M. Sgouros
Dept. of Technology Education & Digital Systems
University of Piraeus
Karaoli & Dimitriou 80, 18534, Piraeus
Greece
E-mail: sgouros@unipi.gr

KEYWORDS
IPTV, content synthesis, peer-to-peer networks.

ABSTRACT

Novel IPTV platforms offer new creative possibilities for dynamic, user-driven development and delivery of multimedia content. This paper describes an approach for building a new generation of authoring and presentation environments for IPTV applications capable of allowing an unlimited number of users to modify existing programming content and post their changes on the net. In addition, these environments should allow users to tailor the presentation of IPTV programming content to their taste by selecting a desired subset of changes implemented by other users in it and dynamically viewing and/or broadcasting a version of the material implementing all the desired changes.

INTRODUCTION

The rise of IPTV platforms, massive multimedia publishing sites, peer-to-peer repositories and technologies along with various e-commerce sites has changed dramatically the way content is being distributed, managed and reviewed by its audience. More specifically, content consumption and review is now immediate and massive, leading to the creation of a large amount of content meta-information including audience reviews that may suggest useful ways to improve its quality. On the other hand, IPTV platforms have an insatiable appetite for new content as they need to offer a large number of viewing alternatives to their customers in order to remain competitive. Given that professional IPTV content development is in general very costly, this puts a significant financial strain on novel IPTV efforts and discourages the development of serious alternatives to existing old-fashioned TV broadcasters. Consequently, new creative possibilities for dynamic, user-driven development and personalization of multimedia material on a large scale open up, as IPTV broadcasters can explore new sources of content by tapping into the large number of media professionals or talented amateurs willing to publish or modify original multimedia content in order to improve its impact. These processes can be compared with the incremental creative processes by which folk cultural material and 'open source' software are developed.

On the other hand, geographically distributed, creative teams involved in IPTV content development need to have affordable and efficient ways of implementing their work given the large computational and financial demands of multimedia production. Unfortunately, content creation and management shaped by all these and possibly other types of social interaction is hampered by the scarcity of appropriate web-based environments that can effectively manage the fusion of original multimedia content with the modifications proposed and implemented by creative groups. This unfortunate situation is further exacerbated by the lack of appropriate personalization tools that can allow users to dynamically compose and view desired versions of available multimedia material.

RESEARCH APPROACH

We seek to respond to these trends through the development of a new generation of participative authoring and presentation systems for dynamic, user-driven development and dynamic personalization of multimedia content in IPTV platforms. In particular, our approach aims to develop novel media management methods for:

- Describing adaptation/enrichment of existing content through the formalization of appropriate annotation meta-data and the provision of
  - low-cost tools for meta-data composition and inclusion in the
  - collaborative workflow using encryption-based watermarking techniques.

  The meta-data will effectively capture the types and relations between the editing actions performed on the content by its users along with issues related to the identification of the contribution of each author in each modified version and the access rights for the content.

- Personalizing content presentation through the use of dynamic content composition techniques allowing users to select, view or broadcast modified versions of existing multimedia material. These versions will automatically incorporate user-selected subsets of adaptation actions implemented by other users. Content personalization will be facilitated with the use of 3D visualization techniques enabling users to view the existing annotation meta-data associated with a multimedia asset and freely select a subset of them in order to be presented with a version of the material implementing all the selected actions.

In order to realize a scalable solution for content development we employ a hybrid p2p architecture consisting of a central server (CS) for organizing the communication between the peers and an expandable number of peers that will act as content providers. Each time a user wants to post new multimedia content in the system, the peer will send a message to CS informing it of the content location along with semantic annotations for it that will be automatically generated by the system. These annotations will facilitate content indexing and retrieval by the CS. Furthermore, the system will provide the user with tools for embodying in the content automatically generated id information and access rights using encryption-based watermarking techniques.

Each peer will be able to request from CS a catalog with all the content files satisfying user-selected criteria consistent with the semantic content annotation used by the system. He will then be able to download from the relevant peer the desired content in order to view or edit it provided he has suitable access rights for it.

Each time a user edits a content file he will inform CS of all the changes he has implemented in it by posting to CS a list of meta-data describing the editing actions he has performed. The meta-data for such actions will include, among others, the location of the file implementing the described action so that interested peers can download the content and additional semantic annotations for CS indexing and retrieval purposes. Id information and access rights for the edited content will be watermarked in the file as well.

A user will be able to ask CS for a list of all the editing actions posted for a particular content file and select a subset of these actions. During content playback, the user will be able to visualize in 3D space the meta-data for all the editing actions that have been implemented in the current content file as these are sent by the server. Based on these meta-data the user will be able to select a subset of these actions and ask from the system to automatically provide him with a dynamically generated content file implementing all the selected editing actions.

APPLICATION SCENARIO

The application scenario for our approach will focus on the creation of an IPTV content publishing environment implementing all the techniques described above for ‘open source’ multimedia development and personalized content consumption through IPTV platforms. The proposed environment will foster the development of participative authoring processes and increase the productivity of creative teams by allowing geographically dispersed designers to collaborate easily and effectively. In addition it will allow IPTV broadcasters to constantly tap into the material under development and select promising content for inclusion in their programming alternatives using criteria such as the popularity of particular versions of edited content or the quality of the development teams involved in each edited version.

RELATED WORK

This approach emerged from our interest in providing appropriate technological infrastructures to the social shaping of multimedia content (Sgouros 2003). Although there has been an increased research interest for the use of p2p technologies in multimedia systems this has focused mainly on the support of various forms of content access and distribution (Chan 205; Liu 2006). Our approach will seek to complement all these efforts focusing on the creation of efficient multimedia development tools in p2p networks.

Similarly, there exists a large amount of work in multimedia authoring environments (for a recent overview see (Bulterman 2005)) a significant part of which focuses on supporting various forms of collaborative processes (e.g. (Wittenburg 1994; Candan 1998)) in client-server environments and highly structured collaboration scenarios (e.g. video conferences). This research approach focuses instead on dynamic, user-driven modifications and delivery of existing content by large numbers of self-motivated users in p2p topologies, a situation clearly different from structured collaboration.

REFERENCES

Chan, G.S.H. et al (eds),” *Proceedings of the ACM workshop on Advances in peer-to-peer multimedia streaming*, ACM Press 2005.

Liu, Z., Yu, H., Kundur, D., Merabti, M., On Peer-to-Peer Multimedia Content Access and Distribution, *2006 IEEE International Conference on Multimedia and Expo*, pp.557-560.

Bulterman, D., C., A., Hardman, L., Structured Multimedia Authoring, *ACM Transactions of Multimedia Computing, Communications and Applications*, vol. 1, no. 1, ACM Press, 2005.

Wittenburg, T. M., Little, D. C., An Adaptive Document Management System for Shared Multimedia Data, *Proc. IEEE Intl. Conf. Multimedia Computing and Systems*, 1994.

Candan, K. S., Rangan, P. V., Subrahmanian, V. S., Collaborative Multimedia Systems: Synthesis of Media Objects, *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 3, 1998.

Sgouros, N. M., Analysis, Management and Indexing of Distributed Multimedia Performances based on Audience Feedback, *Multimedia Systems*, vol. 8, no. 6, Springer, 2003.

MODULES FOR AN INTEGRATED SYSTEM APPROACH FOR ADVANCED PROCESSING OF AV-OBJECTS IN DIGITAL TV WORKFLOW

H Joachim Nern
Aspasia Knowledge Systems
Postcode: 200710, Duesseldorf
Germany
Email: nern@aspasia-systems.de

Andrzei Dziech
Katedra Telekomunikacji, AGH
al. Mickiewicza 30, 30-059 Kraków
Poland
Email: dziech@kt.agh.edu.lpl

Victor Dimtchev
Travel Television Ltd
Ivan Ivanov Bulv. 70, Sofia
Bulgaria
Email: dimtchev@ttvi.biz

Georg Jesdinsky
Big7 GmbH
Liebigstr. 13, Duesseldorf
Germany
Email: gj@big7.net

KEYWORDS: *Audio-visual objects, Web Services, Digital TV, Service Composition, Workflow, Semantic Services*

ABSTRACT

The scope of this paper is to describe the integrated system approach for advanced processing of AV-Objects. This system approach is oriented on the design of a software framework consisting of several modules and services, whereas the service designing and composing process is oriented on the WSMO specification. The proposed integrated system approach consists mainly of a threefold approach: meta-search, recognition and classification of objects, application of watermarking methods and the integration of interoperability features of semantic web services. The integrated modules concerning the retrieval, the watermarking, the semantic annotation and the workflow aspects are overviewed.

INTRODUCTION

The digitalization of almost all ambient information has resulted in developing of new infrastructures for better integration of different resources. The digitalization in TV provides several advantages – among others it allows user to chose *what* to watch, *how* to watch and interact with it, and *when* to watch it. Furthermore, video services can be bundled in new ways and there are potentials for customized pricing (Henten and Tadayoni, 2005). It allows for an extended interactivity in the sense that different aspects can be explored.

The right business models for exploiting the potentials of digital TV are still under development, but a necessity for designing frameworks of software and technological tools for delivering services and contents to customers in new ways has been come into being. Such kind of frameworks has to be based on the enhanced searching, retrieval and processing of audio-visual objects (AV-objects).

In the following a system integration approach is described, which combines technologies and methods from mainly three different areas. Accordingly in the following sections an overview about the activities in the European community and a state of the art in the areas of networked audiovisual systems is given.

The TIRAMISU - (The Innovative Rights and Access Management Inter-platform Solution) project addresses the problem of creation, delivery and consumption of audio-visual media across a wide range of hybrid networks and platforms, where security issues and access rights are of major concern.

AGAVE aims at solving the problem of end-to-end service provisioning over IP networks by studying, developing, and validating a new inter-domain architecture based on Network Planes that will allow multiple partners to provide parallel internets tailored to service requirements. Uni-Verse – The goal of this project is to create an open source Internet platform for multi-user, interactive, distributed, high-quality 3D graphics and audio for home, public and personal use. ASTRALS is a project focused on personalised, scalable A/V encoding, transcoding / transrating, storage and distribution in existing households via streaming-optimised wireless links. ASTRALS motivation is to implement scalable solutions, for new products and services including personalised, network-aware video adaptation and distribution in multiple heterogeneous terminals (from low-cost PDAs to high-end home cinemas).

VISNET is a European Network of Excellence with the objectives: to integrate world-leading institutions to create a world force, to make significant contributions to the sustainable advance of research in this field, to implement an efficient dissemination plan aiming at improving knowledge economy and enhancing the socio-economic life standards of EU citizens. Home networking is the collection of elements that process, manage, transport, and store information, enabling the connection and integration within a broadband wide area network of multiple computing, control, monitoring, and communication devices in the home (a.o. mobile phones and PDAs).

The project BOEMIE pave the way towards automation of the process of knowledge acquisition from multimedia content, by introducing and implementing the concept of evolving multimedia ontologies which will be used for the extraction of information from multimedia content in networked sources, both public and proprietary.

K-SPACE integrates leading European research teams to create a Network of Excellence in semantic inference for semi-automatic annotation and retrieval of multimedia content. The aim is to narrow the gap between content descriptors, that can be computed automatically by current machines and algorithms, and the richness and subjectivity of semantics in high-level human interpretations of audiovisual media. The project LUISA addresses the development of a reference semantic architecture for the major challenges in the search, interchange and delivery of learning objects in a service-oriented context.

IP TV AND WEB TV

IPTV is one of the biggest challenges of the broadcasting market. IPTV denotes delivering of TV over the IP protocol. Internet as a platform for distributing TV services implies the possibility of customized transmission and facilitates new forms of interactivity and personalisation of services. The concept WEB TV is used both when transmitting TV over the WEB and WEB services over TV networks. Because of interactivity on the Internet, it is possible to add other values to these services. The most successful Internet TV business models are likely to involve syndication to or from other media. On the user side (Video on demand, for example), all services will be provided through one integrated network.

But there is a problem due to the existence of different standards for content and metadata description. At present there is a lack of suitable frameworks enabling organizations to manage knowledge alongside content and in a coherent manner.

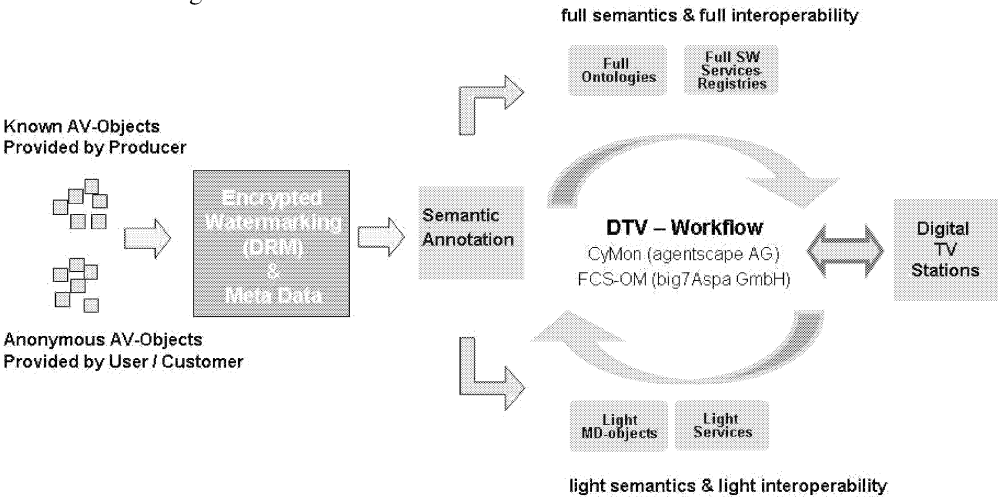

full semantics & full interoperability
Full Ontologies
Full SW Services Registries
Known AV-Objects Provided by Producer
Encrypted Watermarking (DRM) & Meta Data
Semantic Annotation
DTV – Workflow
CyMon (agentscape AG)
FCS-OM (big7Aspa GmbH)
Digital TV Stations
Anonymous AV-Objects Provided by User / Customer
Light MD-objects
Light Services
light semantics & light interoperability

Figure 1 Integrated System Approach

INTEGRATED SYSTEM APPROACH

The main benefit of the proposed framework is to search / retrieve and to convert given AV-objects. The inputs of the system are AV objects, which are partially known as well as partially identified. They are not watermark encrypted and not textually described – therefore an optimised structure for effective retireval and handling is missing. The proposed system delivers as output watermarked and unambiguously identified AV-objects, which are described by textual meta data as well as precisely semantically annotation and due to this fact optimised for retrieval in open as well as closed workflow and resource discovery environments.

The overall methodology (Atanasova T. et al, 2007) is characterised by integrating three main groups of methods into the system (Figure 1):

- meta-search, recognition and classification of AV objects,
- watermarking methods (concerning ID and textual object description),
- interoperability features of semantic web services to a closed framework and system for retrieval, identification and workflow oriented handling and processing of AV objects.

One of the main characteristics of the system is the two step approach of object retrieval and discovery. A rough first step retrieval provides a first selection of objects. In a subsequent step the retrieval is enhanced and optimised by applying semantically annotation. Whereas the watermarking process itself enhances the retrieval process due to the immanent application of textual description in the object itself. It has to be considered that the interaction with the semantic content is still far from being satisfactory and still has to be investigated and developed. Now it is not possible to retrieve visual and audio objects by their semantic meaning. This is due to the lack of methods for the automatic extraction of the semantic content of visual and audio objects (semantic annotation). Also the merging of security DRM methods and the semantic web oriented retrieval is still missing.

Combining Meta-Search, Watermarking and SWS Technologies for application in the AV networked area

The proposed integrated system framework comprises several modules with specifically well defined functions, input and output descriptions.

Known AV-objects provided by the producers (AV-object owner, provider) are retrieved via a GUI interface and conveyed to a process for transformation of the AV-objects to the spectral domain for enabling the encrypted watermarking process (Wassermann et al, 2007).

The AV-object identification (AV-object ID) as well as textual description of the object itself is marked into the object in an encrypted manner.

The aspect related to the "reduced resolutions" AV Objects demands a modified approach of watermarking creation. Mobile front ends like PDA, Mobile Phone or Wireless Notebooks include an appropriate scaling before the broadcasting so the related AV Objects has to be adapted to the reduced resolutions.

Merging of Professional and User Created Media for Generation of new Types of Services

The merging ability of AV-objects and content delivered by professional as well as consumer oriented entities causes and entails the creative generation of new types of services. Complex search engines specified for the retrieval of arbitrary AV content are the basis for such merging and expansion dynamics.

Structuring the Peripherals and the Environment between unstructured versus structured Media, Single versus Multiple Users

AV content search engines bring together and consolidate the structured and unstructured data pools – with the result of effective handling and creation abilities. Applied into workflow environments AV search engines optimises the exchange and collateral processing of commonly used and maintained AV content.

OVERVIEW OF THE SERVICE DESIGNING AND COMPOSING PROCESS

Broadband service development should switch from a technology-oriented approach to the development of sector-specific services, either services for individual specific sectors or the design of context-bound service clusters.

The broadband web services can provide immediate advanatage in:

- automatic selection of the most suitable AV resources based on current use case;
- interoperability between different providers;
- improving flexibility.

The demand for broadband services will be increased, which in turn will enable a broader range of commercial services to be provided.

INTEGRATED MODULES OF THE FRAMEWORK

The proposed integrated system framework - illustrated in Figure 1 – comprises several modules with specifically well defined functions resp. input and output descriptions.

The whole system framework consists of 5 main modules. In the following these modules are explicated and described as well as illustrated in detail in building block diagrams.

Module 1: Known AV-objects provided by the producers (AV-object owner, provider) are retrieved via a GUI interface and conveyed to a process for transformation of the AV-objects to the spectral domain for enabling the encrypted watermarking process (Figure 2). Within this module the watermarking is processed based on selected fast algorithms of "wavelet piecewise linear" transforms and "pseudorandom pulse generators" for encryption purposes. The AV-object identification (AV-object ID) as well as textual description of the object itself is marked into the object in an encrypted manner.

On the one hand the ID / object description is provided by the owner via a GUI-Interface. On the other hand extracted synthetically IDs and object descriptions - generated by the recognition part in module 3 - are used for encrypted watermarking. The processed final objects are stored in a repository for subsequent semantic annotation.

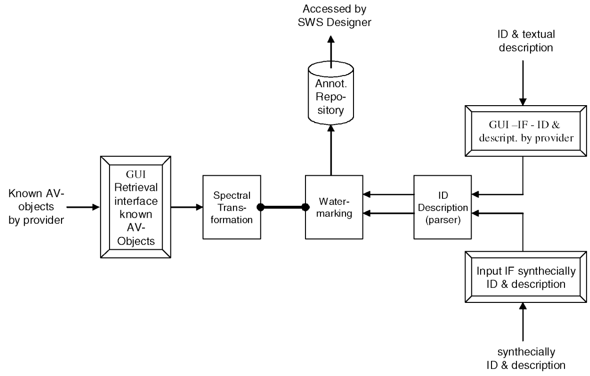

Accessed by SWS Designer
ID & textual description
Annot. Repo-sitory
GUI –IF - ID & descript. by provider
Known AV-objects by provider
GUI Retrieval interface known AV-Objects
Spectral Trans-formation
Water-marking
ID Description (parser)
Input IF synthecially ID & description
synthecially ID & description

Figure 2: Block diagram of the watermarking module

An important issue will be the distribution of AV Objects to the mobile devices. This demands an adoption of the related AV Objects to the resolutions of the mobile front ends like PDA, Mobile Phone or Wireless Notebooks including a appropriate scaling before the broadcasting. The aspect related to the "reduced resolutions" AV Objects demands a modified approached of watermarking creation, which will be investigated in this workpackage.

Module 2: A further stage of AV-object retrieval and resource discovery is given in module 2 (Figure 3). A search component realised as a meta search engine using conventional plug-ins and APIs of existing search engines (Google, Yahoo, MSN, Webcrawler, Lycos etc) executes the first step search (SEARCH 1). In this manner anonym (unknown) AV-objects as well as already watermarked objects are retrieved and subsequently indexed in an indexer (see Figure 3).

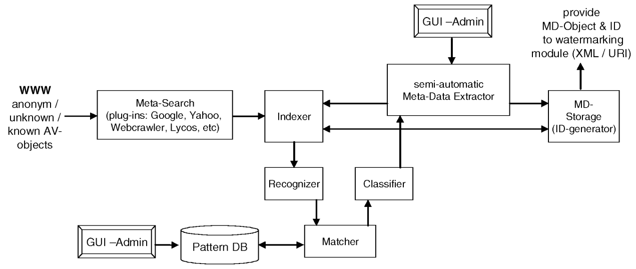

GUI –Admin
provide MD-Object & ID to watermarking module (XML / URI)
WWW
anonym / unknown / known AV-objects
Meta-Search (plug-ins: Google, Yahoo, Webcrawler, Lycos, etc)
Indexer
semi-automatic Meta-Data Extractor
MD-Storage (ID-generator)
Recognizer
Classifier
GUI –Admin
Pattern DB
Matcher

Figure 3: Block diagram of the Search 1 and recognizer module

Within a recognizer the indexed anonym objects are transformed to a appropriate representation and pattern recognition layer (pattern features) and matched (by fuzzy matching) with existing object patterns stored in a pattern storage (Andonova et al, 2006). The matching process is done by estimating similarity measures between currently retrieved and known (existing) patterns. This yields in an object ranking by using centroid distance algorithms. Based on this object ranking a classification is processed (novel multilayered fuzzy methods), which finally yields in a semi-automated (supervised) feature extraction (Admin-GUI).

The extracted features are used for generating a synthetically textual description of the object using reference tables consisting of relationship functions and measures between textual and feature relevant patterns. (It will be furthermore investigated if rule based approaches are applicable.). The extracted textual descriptions are formalized as Meta Data (modified DC) and stored in a MD storage, which are provided to the watermarking module 2 as already mentioned in the previous section.

.

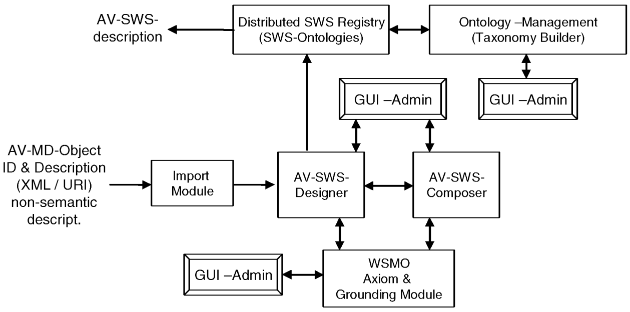

AV-SWS-description
Distributed SWS Registry (SWS-Ontologies)
Ontology –Management (Taxonomy Builder)
GUI –Admin
GUI –Admin
AV-MD-Object ID & Description (XML / URI) non-semantic descript.
Import Module
AV-SWS-Designer
AV-SWS-Composer
GUI –Admin
WSMO Axiom & Grounding Module

Figure 4: Block diagram of the semantic annotation and SWS module

Module 3: The output of module 1 (the watermarking module in conjunction with module 2) are watermarked AV-objects, whereas the ID as well as the object description information are immanently given in the object itself. This textual meta data objects are semantically annotated in module 3, whereas the annotation process is strictly formalised and oriented on given semantic web specifications (Nern et al, 2005). The main reason for the semantically annotation is to obtain a high degree of machine readability (machine processability) to ensure fast and precise search capabilities as well as interoperability characteristics (for the optimised search 2)

The AV-Meta Data objects are imported and processed by the SWS (Semantic Web Service) designer, which generates machine readable (understandable) service descriptions out of the textual meta data. The designer is coupled to an axiom and grounding editor, for fulfilling WSMO compliance. In this way the designed and generated SW services represent thematically and conceptionally the AV-objects and implicitly thematically related domains. Furthermore they ensure a high degree of interoperability demands. A SWS composer enables the composition - the combination – of concepts, representing AV-objects and / or specific domains. The generated SWS are marked up and formalised in SWS ontologies within a distributed registry, which are maintained by an adequate ontology management tool. This kind of semantic formalisation ensures a high degree of retrieval precision and fast object access rate.

Module 4: The module 4 consists mainly out of a service access middleware for enabling user requests, access and discovery of AV-objects represented by service descriptions.

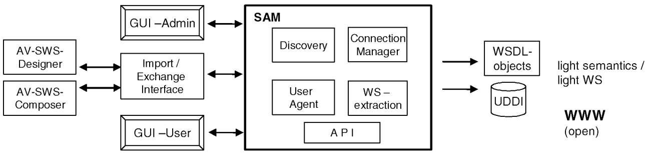


Figure 5: Block diagram of the service access middleware

Furthermore the SAM (Service Access Middleware) supports the service designing and composing process with respect to the discovery of AV-objects (services discovery). The outputs of the SAM module (Fülop et al, 2005) are “static” service descriptions, endowed with “light semantics”. The service descriptions itself are formalised as WSDL files and stored in an UDDI – the registry for static web services. This kind of formalisation is useful and feasible for open networks – like the WWW.

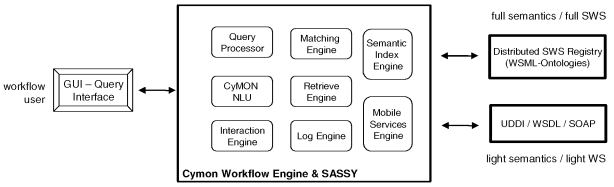


Figure 6: Block diagram of the workflow environment for fast and precise retrieval of AV-objects

Module 5: The module 5 enables the next step of search – the precise and fast retrieval of AV-objects represented as SW services. Furthermore it comprises a workflow environment for handling and processing of AV-objects using existing tools and engines (CyMON Engine, Agantscape). The existing components will be customised for specific project purposes regarding the TV media workflow, which demands a fast and precise resource discovery.

Within the workflow engine the enhanced search and retrieval capabilities are used in several processes like: content production and aggregation for interactive broadcast media, for broadcast resource discovery, for automated generation of meta data content according to given taxonomies and targeted advertisement services. The semantic index engine (sub-module on the right) is responsible for semantically pre-processing as well as interoperation with the SWS registry of module 3.

CONCLUSION

The rapid technological development of broadband laid the foundation for the innovation and development of frameworks for new services. A software and technology framework for searching, interpreting, navigating and retrieving audio-visual (AV) content objects is presented in the paper. The proposed framework sets to support the development of semantic-based and context-aware technologies for automated solutions which combines knowledge, multimedia and Web technologies. In this area the first impulses are given to follow an encompassing approach by combining semantic and multimedia techniques.

REFERENCES

Henten Anders & Reza Tadayoni, Articulation of Traditional and Internet TV, 2nd International Conference of the COST Action A20, University of Navarra (Pamplona, Spain), on 27-28th of June 2003

Atanasova T., Dziech A., Nern Joachim, Sgouros Nikitas; “Framework Approach for Search and Meta-Data Handling of AV-Objects in Digital TV Cycles”, Digital TV Workshop, Euromedia 2007, Delft, The Netherlands

Wassermann J, Nern Joachim, Dziech Andrzei, “Aspects of Watermarking Technologies Applied to Digital TV Broadcast Objects”, Digital TV Workshop, Euromedia 2007, Delft, The Netherlands

G. Andonova, G. Agre, H.-J. Nern, A. Boyanov . "Fuzzy Concept Set Based Organizational Memory as a Quasi Non-Semantic Component within the INFRAWEBS Framework." IPMU2006 IPMU 2006 proceedings. (2006): pp. 2268-2275.

H.-Joachim Nern, G. Agre, T. Atanasova, A. Micsik, L. Kovacs, T. Westkaemper, J. Saarela, Z. Marinova, A. Kyriakov. "Intelligent Framework for Generating Open (Adaptable) Development Platforms for Web-Service Enabled Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent-Systems." W3C Workshop on Frameworks for Semantics in Web Services, Digital Enterprise Research Institute (DERI), Innsbruck, Austria . (2005): pp. 161-168. June 9-10, 2005

C. Fülöp, L. Kovács, A. Micsik. "The SUA-Architecture Within the Semantic Web Service Discovery And Selection Cycle." EUROMEDIA 2005, Toulouse, France . ISBN 90-77381-17-1 (2005): . April 11-13, 2005

ASPECTS OF WATERMARKING TECHNOLOGIES APPLIED TO DIGITAL TV BROADCAST OBJECTS

Jakob Wassermann
Univesity of Applied Sciences
Herzogenburger Straße 68, St. Poelten
Austria
jakob.wassermann@fh-stpoelten.ac.at

H Joachim Nern
Aspasia Knowledge Systems
Postcode: 200710, Duesseldorf
Germany
Email: nern@aspasia-systems.de

Andrzei Dziech
University of Wuppertal
Rainer-Gruenter-Str. 2, Wuppertal
Germany
Email: dziech@uni-wuppertal.de

KEYWORDS: *Watermarking Technologies, Audio-visual objects, Digital TV*

ABSTRACT

The objective of the presented paper is to give a short overview about watermarking approaches for audio visual objects (AV-objects). Since the market of handling digital media within digital TV stations is strong developing in this paper aspects of innovation-related activities of watermarking applications as well as current research and investigation streams are pointed out. Furthermore an approach for watermarking of AV-objects considering DCT, Wavelet and PLT transforms is discussed.

INTRODUCTION

Watermarking is the process by which additional data (WM) are embedded into AV-objects such as images, movies or audio objects. The WM itself is not noticeable and detectable by “human eyes and ears”, but special software detectors can easily recognize this hidden information. These so called digital watermarks are currently used for a variety of different applications. The widespread application of these watermarks is as a unique content identifier that remains constant throughout a variety of manipulations like editing, compression, encryption, and broadcast without affecting the quality of the content. But this is not solely the type of application of watermarking technology. Especially in this discussed approach it can also be used for monitoring and tracking of content in broadcast and internet distribution furthermore for transmitting metadata information.

Existing standard methods are quite good for applications like copyright and content protection with respect to DRM (Digital Rights Management). But for such applications like metadata tagging, where the content information of the image or video is embedded, the current methods are not suitable due to the reduced amount of the included and embedded additional data.

Therefore in this paper the investigation of an approach based on the piecewise linear transform is presented.

ASPECTS OF INNOVATION-RELATED ACTIVITIES OF WATERMARKING APPLICATIONS

The monitoring of the exploitation of Broadcast or Distributed Images concerns the evaluation of broadcast audience of a program in a legal context, as well as the tracking of illegal exploitation (piracy). To overcome this problem several approaches are provided like “Fingerprinting”. In order to complete the tracing of AV content exploitation on a distribution media, a different mark is inserted in each distributed copy of an AV-object before delivery by its legal distributor. This identifies a transaction or a sold item.

A further aspect is “Document Integrity Checking” a type of "Fragile" Watermarking: the mark permits to detect eventual changes in an AV-object due to some attacks and their locations. Then, it is expected to be partly alterable.

Authentication and Identification of an AV-object: a user which receives an AV-object may need to identify the source of a document or the document itself with a high degree of certainty, in order to validate this document for a specific use.

Usage Control: the reception of some distributed AV-objects by some digital equipment in a distribution network, may be controlled in using watermarks inserted in these AV-objects, and only enabled on equipment, whose owner has paid for some access rights. A first example of it is provided by Copy Protection on DVD & CDROM for the Consumer Market: a mark is embedded in DVD video disks in order to prevent copy of DVD, in co-operation with playback and recording devices manufacturers.

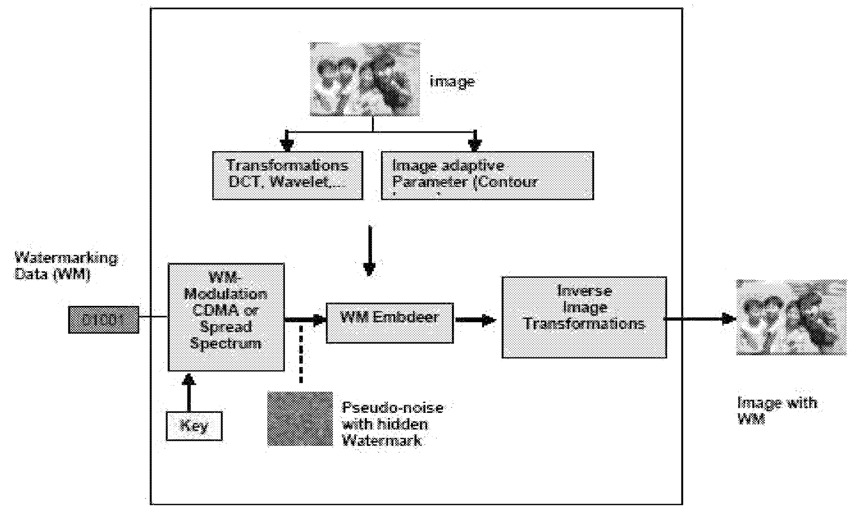


image
Transformations DCT, Wavelet,...
Image adaptive Parameter (Contour
Watermarking Data (WM)
01001
WM-Modulation CDMA or Spread Spectrum
WM Embdeer
Inverse Image Transformations
Key
Pseudo-noise with hidden Watermark
Image with WM

Fig. 1: The principle function of encoding

APPROACH FOR WATER-MARKING FOR APPLICATION IN DIGITAL TV AREA

An original digital message is encrypted by a key that maps it into a special pattern (for example a pseudo noise pattern), which is embedded as a watermark in an AV-object. On the decoder side the enabled knowledge of the key decodes the watermark.

The principle function of the proposed watermarking technology is shown in Fig. 1. The watermarking data are modulated by a spread spectrum or CDMA Modulator. Necessary for this purpose is a pseudo random noise (PN), which is generated by a key. With this key the same PN can be generated on the detection side. As a result a pseudo noise image pattern is created, which contains the watermarking information hidden by the spread spectrum modulation. Before merging the image and the watermark, the image is transformed into other representational forms, like frequency domain or contour image. This should improve the robustness properties by unintentional image operations like compression, filtering etc. For this purpose mostly DCT or Wavelet transformations are used.

The pseudo noise pattern image is embedded into the transform image by the WM embedded unity, which mostly works by simple multiplication or addition or by substituting some coefficients of the spectrum. Afterwards an inverse image transformation is done and an image with embedded watermark information is created.

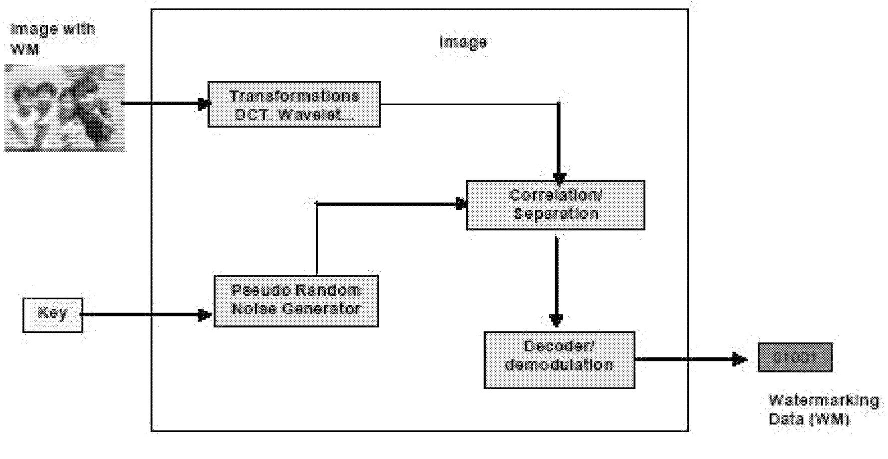


Fig. 2: The principle function of the decoding process

The detection procedure on the decoder side is depicted in Fig. 2. The image containing the watermark information is transformed into the corresponding representation domain (DCT or wavelet). The knowledge of the key is necessary, because it generates the pseudo random noise that is used in the correlation unit to separate spread spectrum modulated pattern from the spectral domain of the image. In the decoder unit the demodulation of the spread spectrum is done and the watermarking data can be extracted. The watermarking procedure is resistant against an attacker, who tries to destroy, to remove or to alter the embedded watermark.

The image representation procedure used on the encoder side is essential for the success of the watermarking technique. There are two different approaches for embedding the encrypted data into the image; methods that are based on spatial domain techniques, like MIT Patchwork or Digimarc algorithm (Bender et al 1994, Rhoads 1998) or frequency domain based techniques that are the most widespread like DCT (Cox et al 1997) or Wavelet based linear transforms (Podilchuk and Zeng 1998). Also image adaptive methods like contour based image representations are in use.

NEW ASPECTS AND ADVANTAGES

Compared to "traditional" watermarking technology, based on DCT transform, an approach is provided, where the proposed method has significant advantages: a method for which very fast algorithms for computation based on piecewise linear transforms are used. To improve the efficiency of searching and retrieval the special method of distribution of watermarking bits generated by a pseudorandom pulse generator is applied. To provide the necessary security of watermarks the Gold pseudorandom sequences will also be tested. The proposed watermarking approach is invariant to compression, rotation and shifting of AV objects.

A further aspect of the ongoing research activity is related to the question: how many bits can be distributed in AV objects by the proposed watermarking technology without significant distortion of the analyzed objects? The idea here is to find the threshold of efficiency for watermarking description of the AV objects. Additionally it is envisaged to improve this efficiency of the distribution of watermarking bits along the contours of analyzed objects.

The mentioned research and investigation targets in detail are:

- Investigation of the most suitable transform with respect to fast computation algorithm (among the family of piecewise linear transforms) for watermarking technology
- Investigation and development of the methods for optimal bit allocation of watermarks using different pseudorandom sequences
- Determining a method of combining of pseudorandom noise and Gold sequences with the QAM modulation method to increase the encryption properties and the security of watermarking.
- Investigation of the relationship between the capacity of watermarking description and the quality of AV objects

Regarding existing watermarking techniques and methods it has to be stated that "traditional" watermarking technology is mainly based on DCT transformation with the main advantage of robustness and stability.

The proposed approach of watermarking is attempted by the use of a special transform based on fast algorithms of piecewise linear transform (PLT) and pseudorandom pulse generators in transform domain. Compared to "traditional" state-of-the-art watermarking technology, which is mainly based on DCT transform, the proposed method has the advantages of:

- high computation speed and reduced algorithmic calculation time
- a larger amount of meta data can be coded into the AV-objects

CONCLUSION

These advantages of the proposed watermarking technique are especially relevant for the AV area, where the objects are generally "heavy weighted" – amounts 100 MB up to 10GB. However the robustness of the method is slightly low. This aspect will be a further major research topic. Especially for mobile devices it is very important to have very fast decoding algorithms due to power consumption issues. For this reason "very fast" algorithms for computation based on PLT will be investigated resp adopted.

REFERENCES

Chan C-K, L.M. Cheng, K-C Leung and S-L Li, "Image hiding based on block difference", *in 8th Int. Conf. On Control, Automation, Robotics and Vision*, 2004

Bender W., D. Gruhl and N. Morimoto, „Techniques for data hiding" *Technical Report, Massachusetts Institute of Technology Media Lab*, 1994

Rhoads G.B., "Method and apparatus responsive to a code conveyed through a graphic image" *US Patent* 5710834, Issued January 20, 1998

Cox I., J.Kilian, F.T. Leighton and T. Shamaoon, "Secure spread spectrum watermarking for multimedia", *IEEE Transactions on Image Processing*, vol. 6 no. 12, pp. 1673-1687, December 1997

Podilchuk C. and W. Zeng, "Image-adaptive Watermarking Using Visual Models" *IEEE JSAC*, vol. 16, no 4, pp. 525-539, May 1998

"SOUVENIRS FROM THE EARTH" - AN INNOVATIVE SPECIAL INTEREST CHANNEL FOR VIDEO ART

Marcus Kreiss

Souvenirs from the earth
10 Rue Perree, 75003 Paris
France
marcus@souvenirsfromtheearth.com

KEYWORDS

Flat screen, Digital Television, Video Art, Science Fiction, Subconscious TV, IPTV

ABSTRACT

"Souvenirs from the earth" is the first TV channel with a 24h art-video program. It is actually available on a German cable network. The films are custom-made for big flat screens transforming them into futuristic paintings: video-paintings. Souvenirs from the earth developed a specific language based on color fields and the mechanics of human memory for this unique new TV format. The program is made of 20 minutes long videos from different artists. It can be shown with electronic music or no sound at all. Based on sponsors, the format also offers a new concept of TV advertisement. "Souvenirs from the earth" is typically the kind of program that will benefit from the fusion of Internet and TV, making an international distribution possible in the future.

CURRENT STATE

The broadcast started in September 2006 on the Kabel BW network (DVB-C) in south West Germany. In 2007 it is planned to join other German and French cable networks and setting up an Internet TV station.

CONCEPT

We assume that every owner of a bigger flat screen will feel the need for a specific content to fill the 'black hole" left by an inactive screen. Just like a painting, the video painting, is there for you when you need it but does not intrude on other activities. Just like a painting it ads a vibration to a space, illuminates it.

The proposed and realized high-end ambient channel is the same for TV what screensavers are for computers; it transforms the screen into an art terminal, a futuristic painting.

This is not only a completely new TV format but in a way a new media since the viewing experience is something between watching an artwork and a video clip. The place this program will take in people's lives is very different from what TV is for us today. The business behind this channel is also different from other traditional programs. Mainly financed by sponsors, "Souvenirs from the earth" has to reinvent the TV ad to make sure it fits in to the smooth continuous stream of rich images the program was defined to be.

TV ads will be seen as original art works and the brand benefits from the prestige supporting a truly innovative format and will be part of modern interiors with its films.

CONTENT

The program is a lineup of 20 minutes art films that do not have any story and exist only for their visual attraction, for the universe they reveal, for the emotions they can generate. Sound is not important. In the future the program could eventually be shown with an independent stream of electronic music as an option, the actual program we broadcast does not have any sound.

PRODUCTION TECHNIQUES USED TO GENERATE CONTENT

On the very opposite from normal TV programs that are eager to attract the attention, often in a very aggressive way, our program is slow, very contemplative often even hypnotic. The productions use few cuts. Long traveling, sequence shots and slow motion effects are frequently used to get a smooth, slowly changing image. As far as we know we are the first to conceive a program using the screen as a canvas, readapting classical concepts of composition and completely respecting the 16:9 formats. Big HD screens and the light they are emitting can act directly, "physically to the viewer", bypassing the brain that was actually doing a lot of work in interpreting the TV images of the old TV sets. This possibility is mainly used.

In our videos we use the power of color, the tensions and emotions we can create with it, and invent a visual language close to the human subconscious.

Dreams and human memory transform reality, color it, slow it down, and switch actors, replay different possibilities.

In our new format, we can do this kind of features. We are free from storytelling and predefined durations. We can actually make our public dream with eyes wide open, connecting directly with their subconscious.

HISTORY

The vision of a TV channel for video-paintings comes probably from 60s and 70s science fiction films. In these movies, the offices and living rooms had always huge moving pictures placed at the wall. Historically **Nam June Paik** in the 60s and **Brian Eno** in the 70 were dreaming about a TV channel completely designed to video art. The concept for our video art channel was presented for the first time in 1998 in the institute for contemporary arts in London with students from different art-schools. After London it was presented in Paris and at the 2000 Biennale of Venice. The company itself was created in 2001 to explore the commercial applications of video-paintings after various associative structures.

DISTRIBUTION

Current state is that the closing of analog capacities, increasing internet connection speeds, projects like Joost (www.joost.com) and systems like apple TV that link internet to the TV-set provide spaces for this kind of special interest channels.

Television, as a mass media, is not really the place where you'd expect art or better that you expect to be art, but things are changing. Everybody can't sell football and the competition between the different networks, the cable, the satellite the digital terrestrial, the triple play and the IPTV lets original special interest content appear to be competitive today. Not using speech to support a story and based on a generic language addressing the subconscious, the program of souvenirs from the earth is easy to export to any country

CONCLUSION

Bigger Flat screens and new distribution possibilities make it possible today to redeem the promise made in science fiction films, to decorate interiors with moving images displayed in screens. Thanks to New distribution technologies, television is not a mass media anymore and a special interest channel can be viable addressing a relatively a small but worldwide audience

FURTHER RESEARCH

The efforts of the next years will mainly be focused on upgrading the catalogue, to find more sources and artists to guarantee our high standard of innovation and novelty.

AUTHOR BIOGRAPHY

Marcus Kreiss was born 1961 in Hamburg, Germany. After finishing Film-School in Rome and Art-School in south of France he was active as painter and conceptual artists presenting his works in international collections. After that he started developing the concept of video-painting and finally founding “Souvenirs from the Earth” in Paris and Cologne.

Figure 1: Colored Knees

Figure 2: Cantina

RELEVANT BUSINESS ASPECTS OF AMBIENT MEDIA IN FUTURE DIGITAL TELEVISION AREA

Lars Riedel
SFTE GmbH
Neusserstr 476 50733 Cologne, Germany
riedel@sfte.de

KEYWORDS

Special interest, innovative advertising formats, TV Channel, Ambient revolution, digitalization

ABSTRACT

Ambient advertising started to appear in media jargon about four years ago, but now seems to be firmly established as a standard term within the advertising industry. It refers to almost any kind of advertising that occurs in some non-standard medium outside the home.

This kind of Advertising as a category has been developed to harness niche media opportunities as gross audience fragments. The capture of customers while they are in an active mindset and have the time to pay attention as part of their weekly routine is more effective and long-term based than the traditional way of advertising.

INNTRODUCTION

The famous "agenda setting approach" is proving that the mass-media controlled the view from viewer's perception. Further taken the media also influence above the effect of perception. In this case it certain the Awareness of the viewer their existent properly meaning and the Salience suppose to be the priority of any individual being. The effects along any circle life can be first cognitively then affectively and also emotionally. However the success of any medially influences is found by their connotative direct response. How effective the agenda setting approach will be depends of their obtrusiveness: In the case of direct reachable information e.g. weather the effect used to be less intensive than in the issue of unreachable experience e.g. overseas wars. Varieties are also found in the different kind of the medium.

- TV-Media coverage used to have a more shortly "spotlight-effect", while the Print-media coverage leads to a longer agenda-setting approach.
- Exponents of the "priming effect" also believe that Media-coverage could provoke a value based charging about issues, those have an impact of the viewer's perception.

Additionally the accessibility of advertising is depending on the advertising-dose, thus the pressure of advertising. Thereby is the contact-dose of a defined target group responsible for the recall performance, further the after deduction coverage express calculative fundamental contact-prospects.

SYMPTON OF FATIGUE ACROSS SHOCKING COMMON TV SPOTS

Luxton und Drummond define Ambient Advertising as "the placement of advertising in unusual and unexpected places often with unconventional methods and being first or only ad execution to do so"(in: ANZMAC 2000, Visionary Marketing for the 21st Century, S. 735; Original) In reference to this issue ambient advertising ought to have in the future digital television an separate position. Above-mentioned the pressure of advertising is less in the sense of direct response than common TV-Media advertising but obviously more intensive as well. In terms of ambient-advertising pressure supposed to be less, thus the medium and the transferring message reach the viewers subconsciously.

This kind of background influence runs slowly towards the viewers mind and become manifest in the formation of opinion. The following consumer behavior will affected by these art oft new advertising.

AMBIENT ADVERTISING REVOLUTION

On the one hand common TV-Spots are loud, shrill and even overwhelmed, but on the other hand seems TV-Advertising the most successful way to reach consumer and selling products or services worldwide. The development of becoming a niche community, the demand for individual content and also the emotional blunting compared to traditional advertising forms leads to this new kind of content advertising: TV Ambient Advertising.

VIDEO PAINTINGS ARE „SOUVENIRS FROM THE EARTH“

“Souvenirs from the earth” (SFTE) was founded in 2001 in Cologne and Paris to develop specific content for bigger LCD and Plasma screens transforming them into "video-paintings". As pioneers SFTE also have to create an economical base for this sci-fi advertising vision to become reality.

They becoming successful producing advertising ambient media spots for restaurants, bars, flagship stores and commercial presentations.

Apparently become this Video Channel part of advertising revolutions. The mentioned ambient media spots (AMS) are based on visual power and less real content: „Our films are 100% narrative free with strong visual aesthetics“.(Marcus Kreiss, Creative film producer).

Furthermore offer SFTE their customers the opportunity to integrate individual advertising message in high value „video paintings“. This content, rather the message is not bound to any regional boarders or supposed to fail at language barriers. This advantage allows an international direction of advertising for every company and simplified the communication activities. Conditional on the proceeding digitalization enabled ambient media working against the common way of avoiding TV advertising with the aid of hard disc recorder, video on demand or IP-TV. Additional data is available from: http://www.sfte.de

AUTHOR BIOGRAPHY

Lars Riedel was born 1981 in Cologne, Germany. He is related to the department of media and marketing economics at the University of Applied Science in Cologne. He is working on his final dissertation about financial services marketing in London. After his graduation he plans to work in the marketing field or applied research of media impacts.

AUTHOR LISTING

AUTHOR LISTING